

Evaluation de méthodes de segmentation bayésiennes pour l'imagerie TEP en oncologie

M. Hatt¹, C. Roux^{1,2}, D. Visvikis¹

¹Inserm, U650, Brest, F-29200 France ; ²ENST Bretagne, GET-ENST, Brest, F-29200 France

hatt@univ-brest.fr ; christian.roux@enst-bretagne.fr ; dimitris@univ-brest.fr

Résumé – Ce travail se concentre sur l'étude de méthodes de segmentation statistiques pour la détermination des volumes fonctionnels dans le cadre de l'imagerie TEP avec des applications en oncologie, en particulier la radiothérapie. Nous présentons ici une comparaison des performances de différentes approches (locale et globale, « dure » et « floue ») pour des images réelles d'acquisitions TEP du fantôme IEC. Les résultats obtenus sur images simulées sont ici confirmés : la méthode locale est la mieux adaptée, et fonctionne mieux que l'approche par chaînes de Markov cachées et la méthode de référence par seuillage.

Abstract – This study deals with statistical segmentation methodologies for functional volume determination in PET imaging for oncology applications like radiotherapy. We present here a performance study of various Bayesian approaches (local and global, "hard" and "fuzzy") that were validated using simulated data and that are here confirmed using real acquisitions of the IEC phantom. Results show that the local fuzzy approach is the best one with respect to the reference thresholding method and other Bayesian approaches.

1. Introduction

L'imagerie de Tomographie par Emission de Positons (TEP) s'est imposée comme un outil puissant en oncologie, en particulier concernant le diagnostic. Il existe toutefois d'autres applications pouvant bénéficier de l'utilisation de la TEP, comme la radiothérapie [1] ou le suivi des patients pendant et après traitement [2]. Bien que la quantification soit cruciale pour le diagnostic et l'évaluation de la réponse thérapeutique, des applications comme la radiothérapie nécessitent également une détermination précise de la forme et du volume d'une lésion. Différentes méthodes manuelles ou automatiques ont été proposées pour déterminer un volume d'intérêt en TEP. Les méthodes manuelles présentent le défaut d'être hautement subjectives et très laborieuses [2], en particulier en TEP où les images sont bruitées et floues, à cause d'une faible résolution spatiale (environ 5 mm suivant les trois dimensions) et les effets de volumes partiels (EVP) résultants, ainsi que des paramètres d'acquisition (dose de produit injecté au patient et durée d'acquisition limitées, faibles contrastes...) et de reconstruction des images (choix de l'algorithme, nombre d'itérations...). Les méthodes semi-automatiques les plus utilisées sont des seuillages dont les valeurs sont fixées par des études sur fantômes ou à partir d'*a priori* de scanner X. Comme démontré dans des études cliniques récentes, ces méthodes mènent à des volumes d'intérêt très variables [3]. D'autres approches ont été étudiées pour la détection d'objets d'intérêt en TEP, comme la détection de contours, le fuzzy C-Mean ou les watersheds. Les performances de ces algorithmes sont souvent sensibles aux variations de l'intensité du bruit ou aux changements de contraste dans les images. Elles sont de plus généralement dépendantes d'initialisations et de pre- ou post-traitements gérés par l'utilisateur, ainsi que d'informations *a priori* supplémentaires. Enfin, la majorité de ces travaux ont étudié la capacité de ces méthodes à détecter des objets dans des images de corps entier, mais ne se sont pas intéressés à leurs performances concernant la détermination précise du volume d'une lésion. Les méthodes bayésiennes ont prouvé leur capacité à résister à de fortes variations

d'intensité de bruit ou de contraste grâce à leur modélisation statistique, et ont donné de bons résultats sur de très nombreux types d'images différentes (radar, sonar, astronomique, satellite, médicale), mais ont rarement été appliquées en TEP [4]. Nos travaux portent sur l'évaluation des méthodes bayésiennes pour la détermination des volumes fonctionnels en TEP, et nous comparons ici l'utilisation de différentes approches de segmentation, adaptées pour traiter les images TEP. L'analyse est effectuée sur des images d'un fantôme couramment utilisé en TEP, reconstruites avec des algorithmes itératifs qui sont aujourd'hui largement utilisés en contexte clinique. Différentes conditions d'acquisition (contraste et qualité statistique) d'image ont été considérées dans le cadre de cette étude, pour laquelle des images réelles ont été acquises, pour confirmer les résultats précédemment validés sur des images simulées.

2. Approches considérées

Dans la grande famille de méthodes bayésiennes en traitement d'image [5], nous avons choisi de mettre l'accent sur deux approches différentes : une méthode globale et une locale. La première fait appel à une modélisation par chaînes de Markov cachées [6]. Cette approche est moins intuitive que les champs de Markov cachés pour une application en traitement d'image du à la modélisation spatiale utilisée, toutefois les chaînes peuvent présenter des résultats similaires aux champs dans de nombreux cas, avec à la clé des temps de calcul nettement moins longs et des facilités d'implémentations non négligeables [5]. La méthode locale est dépourvue d'hypothèse de Markov et travaille « pixel par pixel », ce qui est *a priori* moins efficace. Pourtant dans certains contextes, les méthodes locales se sont montrées compétitives, en particulier grâce à l'estimation adaptative des paramètres qui prend en compte le voisinage spatial de chaque voxel [7]. Ces deux méthodes peuvent également être modélisées de façon classique (« dure ») ou alors « floue ». Les modèles flous sont d'un intérêt tout

particulier en TEP, car les EVP sont majeurs, surtout pour les objets d'intérêt de taille inférieure à 2cm. Dans le cadre bayésien, appelons T l'ensemble correspondant aux voxels de l'image (en TEP une image est un volume 3D). On considère deux processus aléatoires $Y = (y_t)_{t \in T}$ et $X = (x_t)_{t \in T}$. Y représente l'image observée et X représente la vérité cachée. La problématique de la segmentation statistique est d'estimer ce X caché à partir de l'observation Y que l'on suppose être une version bruitée de X . X prend ses valeurs dans $\Omega = \{1, \dots, K\}$ avec K le nombre de classes c , et Y prend les siennes dans \mathbb{R} . On modélise ensuite la relation entre X et Y grâce à la distribution jointe $P(X, Y)$. Cette distribution peut être obtenue avec la loi de Bayes :

$$P(X|Y) = \frac{P(X, Y)}{P(Y)} = \frac{P(Y|X)P(X)}{P(Y)} \quad (1)$$

où $P(Y|X)$ est la vraisemblance de l'observation Y conditionnellement à la vérité cachée X et $P(X)$ est la probabilité *a priori* sur X . On peut ensuite définir $P(X, Y)$ avec ou sans l'aide d'une modélisation markovienne.

L'idée derrière l'implémentation de la mesure floue a été introduite par [8] et est basée sur l'incorporation d'un nombre fini de niveaux de flou F_i en combinaison avec deux classes « dures » homogènes, contrairement au modèle standard où seul un nombre fini de classes « dures » est considéré. L'outil statistique modélise l'incertitude : on suppose que le voxel est clairement identifié mais que l'observation attachée est bruitée. La partie « floue » modélise quant à elle l'imprécision de l'appartenance du voxel : on suppose en plus qu'un voxel peut appartenir à deux classes à la fois. Le fait que des voxels puissent contenir différents tissus avec des proportions variables du fait de la résolution spatiale du scanner (ce qui est particulièrement important en TEP) est donc modélisé par la mesure floue utilisée. Une façon de réaliser cette extension est de considérer simultanément les mesures de Dirac et de Lebesgue : on considère donc que X prend ses valeurs dans $[0, 1]$ au lieu de $\Omega = \{1, \dots, K\}$. Soient δ_0 et δ_1 les mesures de Dirac en 0 et 1 et ζ la mesure de Lebesgue sur $]0, 1[$. On définit alors la mesure $\nu = \delta_0 + \delta_1 + \zeta$ sur $[0, 1]$. Il est important de noter qu'utiliser 2 classes « dures » et deux niveaux de flou n'est pas équivalent à utiliser 4 classes « dures » dans le modèle classique où $\nu = \delta_1 + \delta_2 + \dots + \delta_k$. L'extension au modèle flou est une généralisation du modèle standard dans le sens où on retrouve le modèle dur si la composante floue est nulle, et le modèle ne confond pas bruit et flou [5, 8]. Cette nouvelle mesure implique une redéfinition de tous les outils de l'approche de segmentation.

2.1 Modèle *a priori* : chaînes de Markov

L'approche par chaînes de Markov consiste à faire une hypothèse « globale » sur le champ de la vérité cachée X . On suppose en effet qu'un processus markovien (ici une chaîne) peut modéliser ce champ *a priori* X . X est une chaîne de Markov signifie :

$$P(x_t | x_1, \dots, x_{t-1}) = P(x_t | x_{t-1}) \text{ for } 1 < t \leq T \quad (2)$$

La distribution de X est alors entièrement définie par les probabilités initiales qui définissent les probabilités du premier élément d'appartenir à une des classes, et les probabilités de transitions, c.a.d. les probabilités de changer ou non de classe entre deux éléments successifs de la chaîne. Il est nécessaire de transformer l'image 3D en chaîne (une dimension donc). Pour cela, on utilise le parcours fractal d'Hilbert-Peano [6]. Pour définir une chaîne floue, on utilise une densité conjointe g pour (x_t, x_{t+1}) sur $[0, 1] \times [0, 1]$ qui permet de définir les probabilités initiales et celles de transitions [6].

2.2 Modèle *a priori* : approche locale

Dans l'approche locale, aucune hypothèse de Markov n'est utilisée. Le modèle *a priori* est donc simplement constitué par les probabilités *a priori* pour chaque x_t d'appartenir à une classe donnée. En conservant cette modélisation, aucune prise en compte de la position spatiale d'un élément x_t dans le volume n'est garantie, et la méthode est beaucoup moins performante face au bruit où à des structures complexes qu'avec une modélisation markovienne, qui introduit intrinsèquement cette prise en compte de la corrélation spatiale des voxels. Il est possible d'utiliser un modèle contextuel [9] pour inclure le voisinage spatial de chaque élément dans la modélisation des probabilités *a priori*, toutefois, inclure l'information spatiale de cette façon conduit à une explosion du nombre de paramètres, et en pratique cette modélisation ne permet de prendre en compte qu'un nombre limité de voisins, en général un ou deux dans le cadre d'images 2D [9]. Or notre application doit nécessairement faire appel à un voisinage 3D complet pour tirer au mieux parti de l'information disponible dans le volume. Une solution est l'utilisation du contexte dit « adaptatif » [9] dans lequel l'information de voisinage spatial est introduite dans la phase d'estimation des paramètres. Dans cette modélisation, la probabilité *a priori* d'appartenir à une classe donnée est ré estimée non pas en utilisant toute l'image, mais spécifiquement pour chaque voxel x_t en utilisant uniquement son voisinage. Dans notre application, le voisinage 3D de chaque élément est considéré, soit 26 voisins (la 8-connexité étendu en 3D). La densité h dans le cas local s'écrit :

$$\begin{aligned} h(0) &= P[X_t = 0] = p_0 \\ h(1) &= P[X_t = 1] = p_1 \\ h(\varepsilon) &= 1 - p_0 - p_1 \text{ for } \varepsilon \in]0, 1[\end{aligned} \quad (8)$$

où h vérifie la condition: $h(0) + h(1) + \int_0^1 h(\varepsilon) d\varepsilon = 1$.

Dans le cas de l'estimation adaptative, p_0 et p_1 sont estimés pour chaque voxel et non sur tout le volume.

2.3 Modèle d'observation: bruit gaussien

Dans le contexte de la segmentation statistique d'images, le terme bruit désigne la variation des valeurs des voxels autour d'une moyenne avec une certaine variance au sein de chaque classe. Il est possible de modéliser ces distributions par des lois normales ou d'autres lois, telles la loi beta ou gamma. Dans toutes les

approches considérées dans cette étude, les lois d'observation $P(Y|X)$ sont supposées gaussiennes, comme c'est le cas dans la grande majorité des applications des méthodes bayésiennes à différents types d'imagerie (astronomique, satellite, médicale...). Nous sommes conscients qu'il s'agit d'une hypothèse simplificatrice et nos futurs travaux s'attacheront à tester d'autres lois d'observation, notamment grâce au système de Pearson [10] qui en contient sept autres différentes et dont l'intérêt à été révélé notamment en imagerie radar et sonar.

2.4 Estimation des paramètres

Pour toutes les méthodes bayésiennes, l'estimation des paramètres (distributions initiales et de transitions dans le cas des chaînes, probabilités *a priori* pour chaque voxel dans le cas local, et moyennes et variances de chaque classe pour les deux) est une étape indispensable. Elle permet en effet de rendre la segmentation non supervisée (aucun des paramètres n'est connu *a priori*). Cette estimation est réalisée grâce à l'algorithme itératif Stochastic Expectation Maximization (SEM) [11]. Ce dernier consiste à simuler une réalisation aléatoire de X suivant sa distribution *a posteriori* de l'étape en cours, puis à estimer tous les paramètres d'intérêt sur cette réalisation, et ce de façon itérative jusqu'à stabilisation. L'algorithme SEM est une version stochastique du classique Expectation Maximization (EM) qui a l'avantage d'être beaucoup moins dépendante de l'initialisation. Cette initialisation est faite avec l'algorithme K-Means pour les paramètres du modèle d'observation. Les paramètres du modèle *a priori* sont initialisés avec des valeurs équiprobables pour chaque classe. Changer cette initialisation n'a que très peu d'influence sur le résultat final de l'estimation grâce à l'aspect stochastique de cette dernière.

2.5 Segmentation

Dans le cas des chaînes de Markov, le critère du *marginal posterior mode* (MPM) a été retenu car il semble le plus adapté (contrairement au MAP) à la segmentation d'image car il pénalise une configuration proportionnellement au nombre de sites mal classés [5]. L'objectif du MPM est de minimiser l'espérance $E\{L(x_i, \hat{x}_i) | Y\}$ où L est une fonction de perte:

$$L(x_i, \hat{x}_i) = |x_i - \hat{x}_i| \quad (8)$$

avec x_i la classe réelle et $\hat{x}_i = \hat{s}(y_i)$ celle affectée par le processus de segmentation \hat{s} . Pour le modèle local, le critère du *maximum posterior likelihood* [9] a été choisi.

3. Images TEP

Les images utilisées dans cette étude sont celles du fantôme IEC [12]. Il s'agit d'un cylindre qu'il est possible de remplir d'eau, contenant six sphères de diamètre 10, 13, 17, 22, 28 et 37 mm simulant les lésions, que l'on peut également remplir d'eau. La solution de ^{18}F , l'isotope émetteur de positons, est injectée en quantité variable dans le cylindre et les sphères, avec un ratio de 8:1 et 4:1 pour étudier différents contrastes. Des acquisitions du fantôme ont été effectuées en format *list mode* avec un scanner

Philips GEMINI PET/CT, en injectant 7.4 kBq/cm³ dans le fond et 29.6 ou 59.2 kBq/cm³ dans les sphères. Différents temps d'acquisition ont été considérés pour étudier l'influence de la qualité statistique des images sur la performance des algorithmes (plus l'acquisition est longue, plus le rapport signal sur bruit est réduit), et ce en reconstruisant 1, 2 ou 5 minutes de fenêtre temporelle d'acquisition. L'algorithme de reconstruction utilisé est le RAMLA 3D dont les paramètres ont été précédemment optimisés [13]. La taille des voxels de ces images reconstruites est de $2 \times 2 \times 2$ mm³. Deux images ainsi qu'un schéma du fantôme sont présentées dans la figure 1.

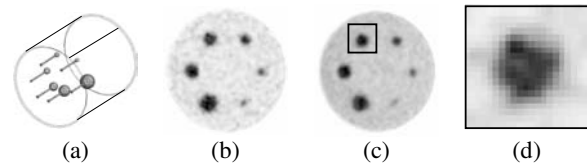


FIG. 1 : le fantôme IEC (a), tranche centrale, ratio 8:1 (b) et 4:1 (c) et sélection de la sphère de 22mm (d).

La sphère de 37mm semble plus petite sur ces images que ce qu'elle devrait, car elle ne se trouve plus dans le même plan que les autres suite à une casse du tube d'injection. De plus, la sphère de 28mm a été ignorée dans cette étude car elle avait été remplacée par une sphère dont le diamètre n'était pas connu avec suffisamment de précision. On s'intéresse à la capacité de chaque algorithme à définir avec précision le volume de l'objet d'intérêt ; on ne se situe donc pas dans le cadre de la détection de l'objet dans l'image et on fait alors l'hypothèse que le praticien a déjà repéré l'objet, et l'a isolé dans une « boîte » l'englobant complètement (tel qu'illustré en 2D dans la figure 1-(c) et 1-(d)).

4. Résultats et perspectives

La figure 2 résume les performances de chaque approche en terme de détermination des volumes. Les résultats sont présentés de la façon suivante : la partie gauche comprend les résultats pour un ratio de 4:1 et la partie droite pour un ratio de 8:1. Dans chaque partie, les résultats pour les 5 sphères (37, 22, 17, 13 et 10 mm de diamètre) sont présentés en % d'erreur par rapport au volume réel de la sphère. Chaque résultat présente une barre de variation correspondant aux 3 différents niveaux de bruit (1, 2 ou 5 minutes d'acquisition). LAB correspond à « local adaptive bayesian » (la méthode locale), HMC à « Hidden Markov Chain » (la méthode globale), FLAB et FHMC à leurs versions floues, et T50 au seuillage à 50% de la valeur maximale de la lésion, proposée par [2]. La figure permet de comparer les approches locales et globales mais également l'approche dure et floue (LAB contre FLAB et HMC contre FHMC). Enfin, on peut comparer ces approches avec la méthode de référence T50.

L'introduction de la mesure floue entraîne une amélioration significative de la segmentation des objets d'intérêt en particulier pour les petites sphères (<2 cm) où les méthodes dures entraînent des erreurs très importantes. Les erreurs sont largement inférieures pour les approches floues (en moyenne 7% et 11% pour FLAB et FHMC contre 38% et 46% pour LAB et HMC, en excluant la

sphère de 10 mm), et on observe également une réduction de la dépendance par rapport au niveau de bruit de l'image (6% et 9% pour FLAB et FHMC contre 13% pour LAB et HMC). Ceci démontre l'intérêt de la mesure floue dans les modèles pour l'application à l'imagerie TEP.

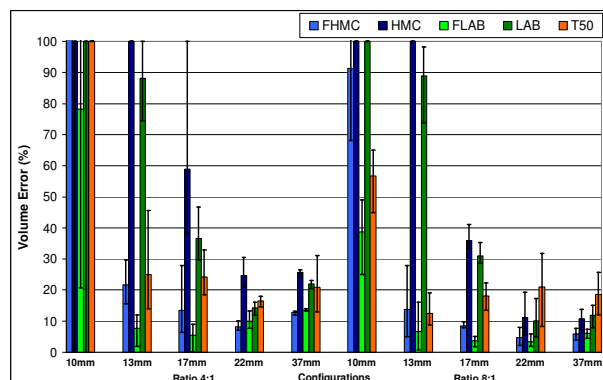


FIG. 2 : Comparaison des performances des méthodes.

La méthode locale semble mieux adaptée que la méthode par chaînes de Markov cachées, en particulier sur les petits objets (<2 cm), tandis que ses performances restent équivalentes sur les plus grandes sphères. Ceci s'explique par le fait que dans le cas des chaînes, la transformation du volume 3D en une chaîne à une dimension entraîne une perte de corrélation spatiale entre les voxels (très peu nombreux) définissant les plus petites sphères. En effet le parcours d'Hilbert-Peano utilisé, bien qu'étant celui qui garantit le mieux la conservation de la corrélation spatiale des voxels dans l'image, ne garantit pas que des voxels voisins dans l'image le soient également dans la chaîne (l'inverse est par contre vérifié, deux éléments consécutifs de la chaîne étant voisins dans l'image). Ainsi séparés sur la chaîne, certains de ces voxels ne sont plus assez significatifs pour constituer une classe pertinente et sont donc confondus avec du bruit au sein d'une même classe, particulièrement dans le cas de faibles contrastes. L'approche locale, bien qu'en général moins performante dans la plupart des applications que les approches globales, semble ici remédier à ce problème, du moins en partie. Noter que seuls les résultats de la modélisation adaptative sont présentés ici ; l'approche locale classique (sans estimation adaptative) donne en effet des résultats nettement moins bons que ceux de l'approche par chaîne de Markov. L'utilisation de la modélisation adaptative pour introduire l'information spatiale locale au niveau de l'estimation des paramètres améliore sensiblement les performances. L'approche locale a également l'avantage par rapport aux chaînes de Markov d'être plus simple à implémenter et plus rapide. Elle n'impose de plus aucune restriction sur les dimensions de l'image à segmenter (donc de la boîte de sélection par le praticien sur l'image originale), ce qui est le cas du parcours d'Hilbert-Peano [6].

Concernant la comparaison avec la méthode de référence (utilisant un seuillage), les deux approches bayésiennes floues se montrent bien plus performantes que cette dernière (qui est toutefois meilleure que les approches

bayésiennes dures), avec une erreur moyenne de 7% (variation moyenne 6%) pour FLAB contre 20% (variation moyenne 15%) pour T50, en excluant la plus petite sphère. Cette dernière n'est correctement segmentée par aucune méthode, bien que FLAB donne des résultats encourageants dans les meilleures configurations (moins de 25% d'erreur pour les images les moins bruitées et pour les deux contrastes de 4:1 et 8:1).

Concernant le temps d'exécution, FHMC est plus lent que FLAB, qui segmente les plus grandes sphères en moins d'une minute malgré une implémentation temporaire sous Matlab. Une fois implémenté en C/C++ on peut s'attendre à une exécution bien plus rapide.

Nos travaux en cours s'attachent à étendre cette approche à trois classes dures et trois transitions floues pour traiter les objets d'intérêt dont la distribution d'activité n'est pas homogène et dont la forme n'est pas strictement sphérique. D'autres hypothèses de bruit (lois beta en particulier) sont également étudiées pour valider ou au contraire remplacer l'hypothèse de loi normale pour le modèle d'observation.

Références

- [1] H. Jarritt et al., The role of PET/CT scanning in radiotherapy planning, *The British Journal of Radiology*, 79, S27-S35, 2006.
- [2] N.C. Krak et al., Effects of ROI definition and reconstruction method on quantitative outcome and applicability in a response monitoring trial. *Eur. J. Nucl. Med. Mol. Im.*, 32, 294-301, 2005.
- [3] U. Nestle et al., Comparison of Different Methods for Delineation of 18F-FDG PET-Positive Tissue for Target Volume Definition in Radiotherapy of Patients with Non-Small Cell Lung Cancer, *J. Nucl. Med.*, 2005.
- [4] J.L. Chen et al., Markov Random Field Model for segmentation of PET images, *Lecture Notes on Comp. Sc.*, Vol. 2082, pp. 468-474, 2001.
- [5] W. Pieczynski, Modèles de Markov en traitement d'images, *Traitement du Signal*, Vol. 20, No. 3, pp. 255-277, 2003.
- [6] M. Hatt et al., Fuzzy hidden Markov chains segmentation for volume determination and quantitation in PET, *Phys. Med. Biol.*, Vol. 52, pp. 3467-3491, 2007.
- [7] M. Hatt et al., 3D Fuzzy Adaptive Unsupervised Bayesian Segmentation for Volume Determination in PET, 4th IEEE ISBI, 2007.
- [8] F. Salzenstein, W. Pieczynski, Parameter Estimation in hidden fuzzy Markov random fields and image segmentation, *CVGIP: Graphical Models and Image Processing*, Vol. 59(4), pp. 205-220, 1997.
- [9] H. Caillol, et al., Estimation of Fuzzy Gaussian Mixture and Unsupervised Statistical Image Segmentation, *IEEE Trans. Im. Process.*, Vol. 6, No. 3, 1997.
- [10] Y. Delignon et al., Estimation of Generalized Mixtures and Its Application in Image Segmentation, *IEEE Trans. Im. Proc.*, Vol. 6, No. 10, 1997.
- [11] G. Celeux, J. Diebolt, L'algorithme SEM : un algorithme d'apprentissage probabiliste pour la reconnaissance de mélanges de densités, *Revue de statistique appliquée*, Vol. 34, No. 2, 1986.
- [12] K. Jordan, IEC emission phantom Appendix Performance evaluation of positron emission tomographs, *Medical and Public Health Research Programme of the European Community*, 1990.
- [13] D. Visvikis et al., Characterisation of SUV accuracy in FDG PET using 3D RAMLA and the Philips Allegro PET scanner, *J. Nucl. Med.* Vol. 45(5):103P, 2004.