

# Une approche *a contrario* pour la détection de changements dans des images IRM multimodales 3D

F. ROUSSEAU<sup>1</sup>, S. FAISAN<sup>1</sup>, F. HEITZ<sup>1</sup>, J.-P. ARMSPACH<sup>2</sup>, Y. CHEVALIER<sup>3</sup>, F. BLANC<sup>2</sup>, J. DE SEZE<sup>2</sup>, L. RUMBACH<sup>2</sup>

<sup>1</sup>LSIIT, UMR ULP-CNRS 7005, pôle API, Bd S. Brant, BP 10413, 67412, Illkirch, France  
tel: 03 90 24 44 89, fax: 03 90 24 44 55, rousseau@lsiit.u-strasbg.fr

<sup>2</sup>LINC, UMR ULP-CNRS 7191, Faculté de médecine, 4 rue Kirschleger, 67085, Strasbourg, France

<sup>3</sup>CH Montbéliard, 25200 Montbéliard, France  
rousseau@lsiit.u-strasbg.fr

**Résumé** – La détection de changements significatifs entre deux images demeure un problème délicat. Dans ce contexte, une méthodologie récemment proposée dans [DMM03] émerge : l’approche *a contrario*. Il s’agit d’une approche non paramétrique présentant l’avantage de prendre en compte dans le processus de décision l’information contextuelle et différentes valeurs de seuil de détection. Nous étendons ici cette approche de manière à traiter des images multimodales desquelles sont extraites différentes images de mesure. Pour cela, deux règles de fusion sont développées de manière à combiner l’information provenant des images de mesure et celle provenant des différents seuils de détection. De plus, une nouvelle règle de décision, basée sur des tests de permutation, est proposée. La méthodologie *a contrario* est décrite dans la Section 1. Nous proposerons ensuite un nouveau cadre statistique dans la section 2. Enfin, la section 3 illustre l’application de la méthode pour de la détection de changements dans des images IRM dans le contexte de la sclérose en plaques.

**Abstract** – Estimating significant changes between two images remains a challenging problem in medical image processing. This paper proposes a non-parametric region based method to detect significant changes in 3D multimodal Magnetic Resonance (MR) sequences. The proposed approach relies on an *a contrario* model [DMM03] which defines significant changes as events with very low probability. We adapt the *a contrario* framework to deal with multimodal images from which are extracted measures related to intensity and volume changes. Two fusion rules are carefully designed to handle a set of decision thresholds and a set of image measures. The final decision is taken using a permutation framework to deal with multiple tests. The *a contrario* approach is described in Section 1 and a new statistical framework is presented in Section 2. Finally, Section 3 presents results in the context of multiple sclerosis (MS) lesion analysis over time in multimodal MR sequences.

## 1 Approche *a contrario* et relation avec les tests d’hypothèses

La détection de changements significatifs entre deux images d’une même scène demeure un problème délicat en imagerie médicale. Un état de l’art en détection de changements a récemment été proposé par Radke *et coll.* [RAAKR05] et décompose la problématique de la façon suivante : 1) ajustement géométrique (ou recalage des images), 2) ajustement en intensité, 3) comparaison des images pour identifier les changements. Dans cet article, nous nous focalisons sur l’identification des changements en utilisant la méthodologie *a contrario* récemment introduite par Desolneux *et coll.* [DMM03] et appliquée en détection d’alignements [DMM03], en reconnaissance de formes [CDDMS07] ou encore en détection de mouvement [VCB06].

Le modèle d’observations en l’absence de changements est appelé modèle *a contrario*. Dans ce modèle, les changements significatifs sont définis comme étant des événements de faible probabilité d’occurrence : un événement est dit  $\epsilon$ -significatif si l’espérance du nombre d’occurrences de cet événement est inférieur à  $\epsilon$  dans le modèle *a contrario* [DMM03].

On note  $X$  la variable aléatoire reliée à une mesure de

changement. Tout au long de l’article, on suppose que cette mesure est grande en cas de changement. Soient  $\mu_i$ ,  $i = 1, \dots, N_\mu$  un ensemble de seuils et  $R$  une région de  $n$  voxels indépendants. On considère l’évènement  $E_{\mu_i, R} = \{ \text{au moins } k_i \text{ points de } R \text{ ont une mesure supérieure à } \mu_i \}$ , la probabilité de  $E_{\mu_i, R}$  est

$$P(E_{\mu_i, R}) = \sum_{j=k_i}^n C_j^n P(X > \mu_i)^j (1 - P(X > \mu_i))^{n-j}$$

(i.e. la queue de la binomiale). Considérant un ensemble de  $N_R$  régions, une région  $R$  est dite  $\epsilon$ -significative si :

$$N_R \cdot N_\mu \cdot \min_i P(E_{\mu_i, R}) \leq \epsilon \quad (1)$$

La quantité  $N_R \cdot N_\mu \cdot \min_i P(E_{\mu_i, R})$  correspond à l’espérance du nombre de fausses alarmes sous l’hypothèse nulle. Il est important de remarquer certaines caractéristiques de la méthodologie *a contrario*. Tout d’abord, l’évènement  $E_{\mu_i, R}$  est défini par rapport à une région ce qui signifie que la procédure de décision prend en compte le contexte spatial. De plus, cette approche permet d’utiliser un ensemble de seuils  $\mu_i$ . Enfin, dans [VCB06], l’ensemble des  $P(E_{\mu_i, R})$  est combiné en prenant le minimum de cet ensemble, ce qui peut être vu comme un opérateur OU puisque l’on consi-

dère uniquement le seuil qui permet de rejeter au mieux l'hypothèse  $\mathcal{H}_0$ .

L'approche *a contrario* consiste à rejeter l'hypothèse  $\mathcal{H}_0$  pour la région  $R$  si la région est  $\epsilon$ -significative. La méthodologie *a contrario* est en fait reliée au cadre classique des tests d'hypothèses. La détection de changements appliquée à  $N_R$  régions nécessite l'utilisation de  $N_R$  tests. Ainsi, considérant une erreur de type I (notée  $\alpha$ ) pour chaque test, l'application simultanée de  $N_R$  tests conduit à un nombre de fausses alarmes égal à  $\alpha N_R$  sous  $\mathcal{H}_0$ . Ainsi pour  $\alpha$  égal à 5% et une image contenant 100 000 régions, il faut s'attendre à détecter 5000 régions pour des données respectant  $\mathcal{H}_0$ .

Un moyen simple de contrôler le nombre de faux positifs consiste à choisir un seuil tel que le nombre de fausses alarmes est contrôlé pour toutes les régions de l'image. Une approche possible consiste à utiliser la correction de Bonferroni : pour chaque test un nouveau seuil égal à  $\alpha$  divisé par le nombre de tests. La quantité  $N_R \cdot N_\mu \cdot \min_i P(E_{\mu_i, R})$  définie dans l'approche *a contrario* (Eq. 1) se révèle ainsi être équivalente à la méthode de correction de Bonferroni dans la mesure où il y a implicitement  $N_R \cdot N_\mu$  tests. Cependant, la correction de Bonferroni est conservative dès lors que les tests sont corrélés, ce qui peut conduire à un risque élevé de non-détection. En mettant en lumière la relation qui existe entre la méthodologie *a contrario* et le cadre classique des tests d'hypothèses, nous proposons d'étendre le cadre *a contrario* en proposant des règles de fusion capables de prendre en compte des données multimodales et en utilisant des procédures de comparaisons multiples moins conservatives.

## 2 L'approche proposée

La chaîne de traitement complète est résumée dans la figure 1. Le principe de la méthode est le suivant : l'information contenue dans l'ensemble d'images de mesure segmentées est traduite statistiquement à l'aide d'un ensemble de  $N_M \times N_\mu$  p-valeurs. La décision finale est prise par rapport à une p-valeur (par région), qui est obtenue en combinant l'ensemble des  $N_M \times N_\mu$  p-valeurs.

Considérons  $N_M$  images de mesure  $\{I_m\}_{m=1, \dots, N_M}$  et l'hypothèse nulle suivante : les observations  $(X_i)$  sont indépendantes et identiquement distribuées (iid). L'approche que nous proposons est divisée en sept étapes (notées  $S_1$  à  $S_7$  dans la figure 1).

**Etape 1 :** Cette étape consiste tout d'abord à partitionner l'image en  $N_R$  régions :  $\{R_i\}_{i=1, \dots, N_R}$ . Aucune hypothèse n'est nécessaire concernant la forme de ces régions (l'étape de segmentation est décrite dans la section 3).

**Etape 2 :** La fonction de densité de probabilité (fdp) de la mesure  $x_m$  associée à l'image  $I_m$  est estimée empiriquement sous  $\mathcal{H}_0$  par des techniques non paramétriques. En pratique, il n'est pas possible d'estimer ces densités sous  $\mathcal{H}_0$  directement puisque certains voxels correspondent à des changements. Pour circonvenir ce problème, nous utilisons un schéma itératif : à chaque itération, les régions identifiées comme changements ne sont pas considé-

rées pour estimer la densité sous  $\mathcal{H}_0$  [NH01].

**Etape 3 :** Dans l'équation 1, la définition de  $E_{\mu_i}$  dépend d'un certain seuil  $\mu_i$  qui peut être relié à la notion de *p-valeur*. Etant donné un ensemble de  $N_\mu$  p-valeurs  $p'_j, j = 1, \dots, N_\mu$ , la troisième étape  $S_3$  consiste à calculer un ensemble de seuils  $\mu_j^m$  pour chaque mesure  $m$  à partir de l'équation suivante :

$$\int_{x=\mu_j^m}^{+\infty} f_{|\mathcal{H}_0}^m(x) dx = p'_j$$

**Etape 4 :** On calcule pour chaque région  $R_i$  le nombre de points  $k_{i,j}^m$  pour lesquels l'observation  $x_m$  est supérieure à  $\mu_j^m$ . La probabilité  $p_{i,j}^m$  qu'au moins  $k_{i,j}^m$  points de la région  $R_i$  de taille  $n_i$  aient une mesure observée  $x_m$  supérieure au seuil  $\mu_j^m$  est calculée de la façon suivante :

$$p_{i,j}^m = \sum_{l=k_{i,j}^m}^{n_i} C_l^{n_i} p_j'^l (1-p_j')^{n_i-l}$$

L'approche *a contrario* consiste à calculer pour chaque région la quantité  $N_\mu N_R N_M \min_{j,m} p_{i,j}^m$ . A la place, nous proposons deux règles de fusion qui sont appliquées dans les étapes  $S_5$  et  $S_6$ . Le but de ces deux étapes est de combiner pour chaque région l'ensemble de p-valeurs  $p_{i,j}^m$  afin d'obtenir une p-valeur  $p_i$  caractérisant chaque région  $R_i$ .

**Etape 5 :** La première règle de fusion que nous proposons repose sur le fait qu'un changement est significatif s'il est présent dans au moins un canal (opérateur OU). Soit  $X^\cup = \min_k(Y_k)$  une variable aléatoire où  $\{Y_k\}_{k=1 \dots n}$  est un ensemble de variables aléatoires iid suivant une loi uniforme sur  $[0, 1]$ . Soit  $x^\cup = \min_k(y_k)$  une réalisation de  $X^\cup$ . On peut montrer que la p-valeur associée à  $x^\cup$  est :

$$P(X^\cup < x^\cup) = 1 - (1 - x^\cup)^n$$

Cette règle de fusion est utilisée pour combiner les p-valeurs  $p_{i,j}^m$  (informations obtenues à partir de l'ensemble des seuils). L'étape  $S_5$  consiste donc à calculer les  $p_i^m$  de la façon suivante :  $p_i^m = 1 - (1 - \min_j(p_{i,j}^m))^{N_\mu}$  (on rappelle qu'une p-valeur est uniformément distribuée sous  $\mathcal{H}_0$  sur  $[0, 1]$ ). L'opérateur OU permet ici de détecter différents profils de liaison.

**Etape 6 :** La seconde règle de fusion repose sur le fait qu'un changement est significatif s'il est présent dans tous les canaux (opérateur ET). Soit  $X^\cap = \max_k(Y_k)$  une variable aléatoire. La p-valeur associée à  $x^\cap$  est

$$P(X^\cap < x^\cap) = (x^\cap)^n$$

Cette règle de fusion est utilisée pour combiner les p-valeurs  $p_i^m$  (informations obtenues à partir des différentes images). L'étape  $S_6$  consiste alors à calculer les  $p_i$  de la façon suivante :  $p_i = (\max_m p_i^m)^{N_M}$ . Dans le contexte de détection de changements dans des images IRM, l'opérateur ET permet de combiner des informations provenant des différentes modalités en se fondant sur l'hypothèse que les changements doivent être détectés dans tous les canaux.

**Etape 7 :** La dernière étape consiste à calculer un masque binaire des changements détectés à partir de l'ensemble des p-valeurs non corrigées  $\{p_i\}_{i=1, \dots, N_R}$ . Pour cela, nous proposons d'utiliser deux méthodes : le *false discovery rate* (FDR) [BH95] qui permet de contrôler la

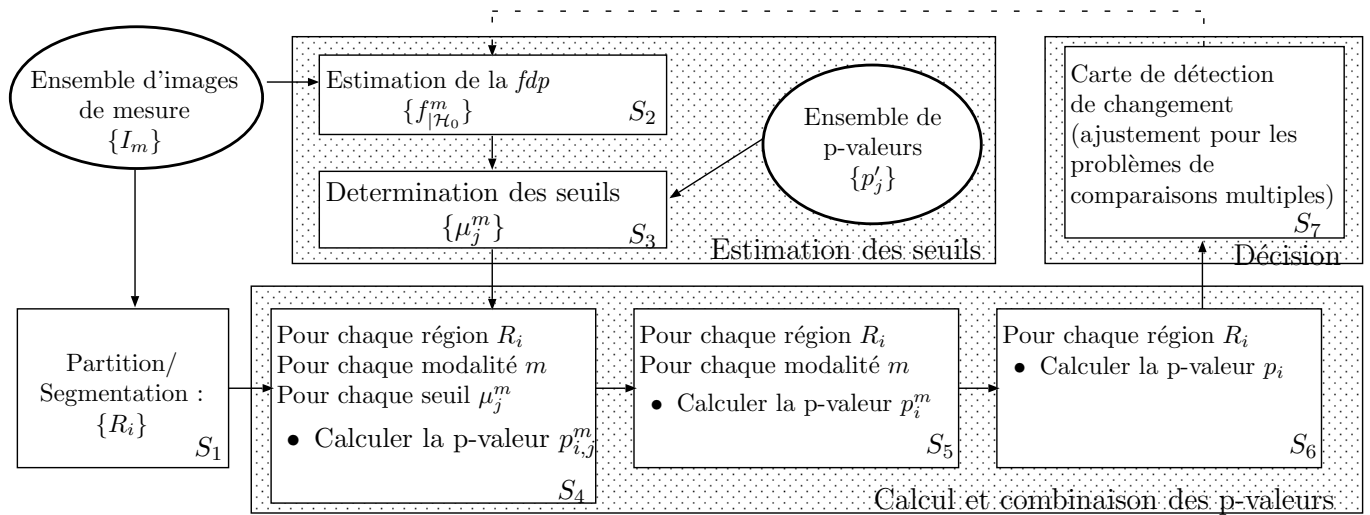


FIG. 1 – Synoptique de l’approche proposée.

proportion de faux positifs parmi les détections et une méthode de permutations [NH01] dont l’objectif est d’estimer sous  $\mathcal{H}_0$  la distribution de  $\max_i p_i$ .

### 3 Application à l’analyse de lésions en sclérose en plaques

L’IRM est un outil fondamental pour l’analyse temporelle quantitative et qualitative de la sclérose en plaques. Dans ce cadre, l’utilisation de la multimodalité est essentielle afin d’extraire et de suivre l’évolution des lésions dans le temps. La détection de changements est alors un outil indispensable pour le suivi des patients.

Il est tout d’abord nécessaire de choisir des caractéristiques de l’image pertinentes pour la problématique de détection de changements. La première caractéristique extraite est le rapport de vraisemblance généralisé (GLR). La seconde caractéristique apporte des informations concernant le changement local de volume en reposant sur l’utilisation du jacobien d’un champ dense estimé à l’aide d’un algorithme de recalage non rigide appliqué entre deux images successives.

L’un des éléments importants de l’approche *a contrario* réside dans le fait que c’est une approche région. Il est donc nécessaire de fournir une carte de segmentation en entrée de l’algorithme (étape  $S_1$ ). Dans l’approche *a contrario*, il n’est pas possible d’utiliser n’importe quelle carte de segmentation. En effet, pour pouvoir appliquer la correction de Bonferroni ou le FDR, il est nécessaire que la segmentation soit indépendante des données. Dans le cas contraire, les p-valeurs doivent être corrigées en considérant le nombre total de régions possibles [DMM03]. Ce nombre, même connu, serait trop grand et l’algorithme ne détecterait alors plus rien. Ainsi, pour ces deux types de corrections, nous utiliserons des régions cubiques de taille  $3 \times 3 \times 3$ . En revanche, les méthodes de permutations permettent l’utilisation d’une segmentation quelconque. Ainsi, pour chaque permutation effectuée, l’algorithme de segmentation est relancé et la p-valeur  $p_i$  est calculée pour

chaque région. Ce processus permet d’obtenir une estimation de la densité de  $\max(p_i)$  sous l’hypothèse nulle. En pratique, nous appliquons un algorithme de segmentation utilisant le principe du *mean shift* [FH75].

La validation d’un algorithme de détection de changements en imagerie médicale est délicat de par l’absence de vérité terrain. Ainsi, afin de pouvoir évaluer les performances de la méthode, nous avons tout d’abord utilisé des images simulées pondérées en T1 et en T2 provenant de BrainWeb (3% de bruit et 1mm d’épaisseur de coupe). La première image ne contient pas de lésion et la seconde contient des lésions modérées de type péri-ventriculaires pour la plupart. Nous utilisons l’index kappa  $KI$  pour calculer le recouvrement entre la vérité terrain et la carte de détection :

$$KI \triangleq \frac{2\#(CD \cap VT)}{\#CD + \#VT}$$

(CD : carte de détection, VT : vérité terrain). Les paramètres de notre méthode utilisés pour tous les résultats sont :  $\{p'_j\} = \{0.01, 0.001\}$ ,  $\alpha = 0.01$ . Le tableau 1 regroupe les résultats obtenus. On peut noter que, grâce à l’utilisation d’une meilleure segmentation, les résultats obtenus avec la méthode de permutation ( $M_3$ ) sont significativement meilleurs que ceux obtenus avec le critère *a contrario* original de Desolneux *et coll.* ( $M_1$ ) ou avec le FDR ( $M_2$ ). En utilisant des approches par segmentation pour les mêmes données synthétiques, des méthodes décrites dans la littérature obtiennent des valeurs de  $KI$  proches de 0.7 et de 0.75 ( $KI \geq 0.7$  représente un résultat très satisfaisant [Bar91]).

L’algorithme a également été testé sur des données cliniques réelles (la figure 2 présente ces résultats pour un patient). Les données IRM longitudinales ont été acquises avec un scanner Philips (image pondérée en T2 : temps de répétition (TR)  $TR = 8693ms$ , echo time (TE)  $TE = 90ms$ , 2mm d’épaisseur de coupe ; FLAIR : TR = 11909ms, TE = 100ms, temps d’inversion (TI)  $TI = 2000ms$ , 2mm d’épaisseur de coupe) et l’intervalle entre deux acquisitions était de trois mois. La méthode a été appliquée sur des données FLAIR et des données pondérées en T2. Les images de mesures ont été calculées avec le GLR et les

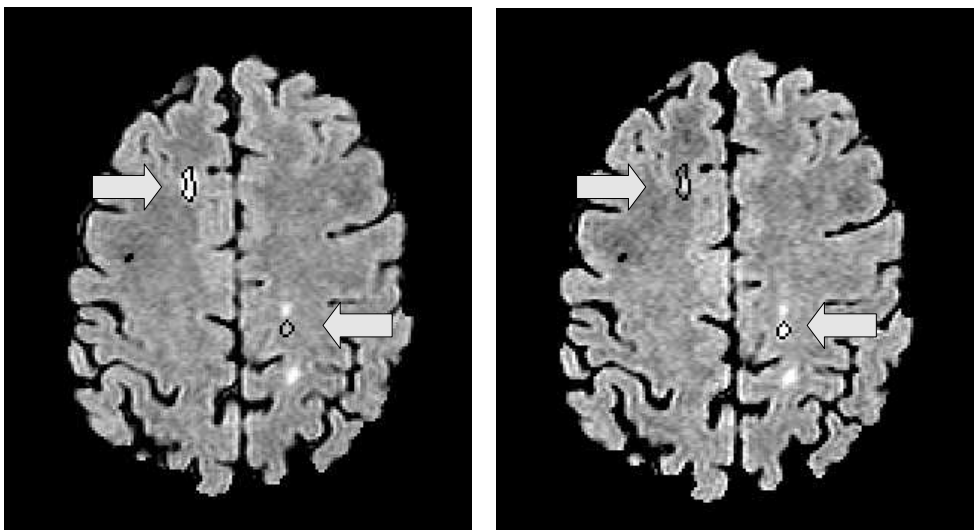


FIG. 2 – Données cliniques (images FLAIR), à gauche : premier examen, à droite : deuxième examen. Deux changements signalés par les flèches ont été détectés pour cette coupe (les contours des lésions sont en noir). Une lésion présente à l'examen 1 a disparu à l'examen 2 (en haut à gauche) et une lésion apparaît au deuxième examen (en bas à droite).

	$O_1$	$O_2$	$O_3$
$M_1$	0.60	0.60	0.54
$M_2$	0.60	0.60	0.55
$M_3$	0.70	0.78	0.71

TAB. 1 – Résultats pour l'index kappa ( $M$  représente l'algorithme et  $O$  les images de mesures utilisées)  $M_1$  : a contrario classique,  $M_2$  et  $M_3$  : méthodes proposées avec FDR ( $M_2$ ) et permutations ( $M_3$ ),  $O_1$  : GLR calculée sur les images pondérées en  $T_2$  (1 mesure),  $O_2$  : GLR et mesure liée au jacobien calculés sur les images pondérées en  $T_2$  (2 mesures),  $O_3$  : GLR et mesure liée au jacobien calculés sur les images pondérées en  $T_2$  et en  $T_1$  (4 mesures).

jacobiens extraits des champs de déformation. La figure 2 montre des résultats de détection de changements pour de petites lésions. Les premiers résultats sur données réelles ont été très encourageants et une procédure de validation sur plusieurs sites cliniques est actuellement en cours.

Ainsi, les résultats obtenus montrent que notre approche permet de détecter de manière efficace des changements significatifs dans des séquences d'images IRM (notamment grâce à l'utilisation rigoureuse de carte de segmentation et à l'introduction de règles de fusion adaptées). De plus, il est à noter que le cadre proposé améliore nettement les résultats obtenus par rapport à l'approche *a contrario* originelle. Cependant, les résultats obtenus avec quatre images de mesure ( $O_3$ ) révèlent que le comportement de l'opérateur de fusion ET n'est peut être pas assez flexible. Aussi, l'utilisation de nouvelles règles de combinaison des p-valeurs sont à l'étude.

## Références

[Bar91] J. J. Bartko. Measurement and reliability : statistical thinking considerations. *Schizophr Bull*, 17(3) :483–489, 1991.

[BH95] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate : A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, 57(1) :289–300, 1995.

[CDDMS07] F. Cao, J. Delon, A. Desolneux, P. Musé, and F. Sur. A unified framework for detecting groups and application to shape recognition. *Journal of Mathematical Imaging and Vision*, 27(2) :91–119, 2007.

[DMM03] A. Desolneux, L. Moisan, and J.-M. Morel. A grouping principle and four applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(4) :508–513, April 2003.

[FH75] K. Fukunaga and L. Hostetler. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Trans. on Information Theory*, 21(1) :32–40, Jan 1975.

[NH01] T. E. Nichols and A. P. Holmes. Nonparametric permutation tests for functional neuroimaging : a primer with examples. *Hum Brain Mapp*, 15(1) :1–25, Jan 2001.

[RAAKR05] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam. Image change detection algorithms : a systematic survey. *IEEE Trans Image Process*, 14(3) :294–307, Mar 2005.

[VCB06] T. Veit, F. Cao, and P. Bouthemy. An a contrario decision framework for region-based motion detection. *International Journal on Computer Vision*, 68(2) :163–178, 2006.