



Université François Rabelais
Tours

Ecole Doctorale : Santé, Sciences et Technologies

Année Universitaire : 2002-2003

**THESE POUR OBTENIR LE GRADE DE
DOCTEUR DE L'UNIVERSITE DE TOURS**

Discipline : **Informatique**

Présentée et soutenue publiquement
par :

Emmanuel DELLANDREA

le 13 octobre 2003

ANALYSE DE SIGNAUX VIDEOS ET SONORES :
APPLICATION A L'ETUDE DE SIGNAUX MEDICAUX

Sous la direction de Nicole VINCENT (co-encadrant : Pascal MAKRIS)

Jury

André-Obrecht Régine	Professeur	Université de Toulouse III
Jolion Jean-Michel	Professeur	INSA Lyon
Makris Pascal	Maître de Conférences	Université de Tours
Vincent Nicole	Professeur	Université de Paris V
Metman Etienne-Henry	Professeur	Université de Tours

Remerciements

Travail difficile et de longue haleine mais immense source de satisfaction, cette thèse n'aurait pu être menée à bien sans le soutien d'une qualité exemplaire, aussi bien sur le plan scientifique qu'humain, de ma directrice de thèse Nicole Vincent et de mon co-encadrant Pascal Makris. Ils m'ont transmis le goût de la recherche et je tiens très sincèrement à les en remercier.

J'adresse également mes remerciements à Christian Proust pour m'avoir accueilli au Laboratoire d'Informatique de l'Université de Tours dont il assure la direction. J'ai trouvé au sein de ce lieu un environnement de travail de grande qualité grâce à une atmosphère amicale et studieuse entretenue par l'ensemble des membres du laboratoire ainsi que par les étudiants du Département Informatique de l'Ecole Polytechnique de l'Université de Tours. Je les en remercie chaleureusement.

Ce travail de recherche n'aurait pu exister sans l'active collaboration de Michèle Boiron, responsable de l'équipe Acoustique et Motricité Digestive du Laboratoire de Physiologie et Motricité Digestive de l'Université de Tours, d'Etienne-Henry Metman et de Zine Benchellal. Ils ont toute ma gratitude.

Mes remerciements vont également à Régine André-Obrecht et Jean-Michel Jolion pour avoir accepté de rapporter ma thèse ainsi que pour leurs remarques et conseils avisés. Plus généralement, je remercie Remy Mullot, Marc Richetin, Michèle Boiron et Etienne-Henry Metman pour leur participation à l'évaluation de mon travail.

Enfin, j'adresse un grand merci à toute ma famille, et plus particulièrement à mes parents et beaux parents pour leur formidable soutien. Merci Aurélie, sans toi je n'aurais pu y arriver.

Sommaire

Table des figures	7
Liste des tableaux	11
1 Introduction	13
1.1 Analyse simultanée de vidéos et de sons pour l'interprétation de séquences multimédia	13
1.2 Application au problème médical du reflux gastro-œsophagien . .	14
1.2.1 Analyse des images radiologiques	15
1.2.2 Analyse des données sonores	17
1.3 Organisation du mémoire	20
I Analyse de signaux vidéos	23
2 Suivi d'objets déformables	25
2.1 Introduction	25
2.2 Algorithmes de suivi	25
2.2.1 Classification des algorithmes de suivi	26
2.2.2 Structure générale des algorithmes de suivi	28
2.2.3 Modèles de mouvement	31
2.3 Suivi temporel d'un objet déformable	34
2.3.1 Aperçu de la méthode de suivi	34
2.3.2 Prédiction des paramètres de mouvement	35
2.3.3 Optimisation de la région prédite par un modèle de région déformable	40
2.3.4 Ajustement du contour de la région par contours actifs . .	47

2.3.5	Rééchantillonnage des points du contour	52
2.3.6	Estimation des paramètres de mouvement	53
2.4	Résultats	56
2.4.1	Initialisation de l'algorithme de suivi	56
2.4.2	Paramétrage de l'algorithme de suivi	59
2.4.3	Résultats du suivi	60
2.5	Conclusion	63
3	Suivi d'objets et ensembles de niveaux	65
3.1	Introduction	65
3.2	Les ensembles de niveaux	66
3.2.1	Présentation générale	67
3.2.2	Formulation mathématique	69
3.2.3	Les techniques d'évolution de front	74
3.2.4	Détection de contours	77
3.3	Suivi par contours actifs non régularisés	80
3.3.1	Les contours actifs non régularisés	81
3.3.2	Méthode de suivi	83
3.4	Résultats	89
3.5	Conclusion	92
4	Caractérisation des objets suivis	93
4.1	Introduction	93
4.2	Extraction de primitives	94
4.2.1	Extraction de paramètres de taille et d'orientation	94
4.2.2	Extraction de paramètres issus de la squelettisation de la région	94
4.2.3	Détermination de la trace des objets	97
4.2.4	Extraction de primitives de vitesse	98
4.3	Application à la caractérisation du SIO	98
4.3.1	Détection de la position du coude de l'œsophage	98
4.3.2	Détection et caractérisation du SIO	100
4.4	Conclusion	101

II	Analyse de signaux sonores	103
5	Identification de signaux sonores	105
5.1	Introduction	105
5.2	Etude de la structure bas-niveau des signaux	108
5.3	Etude des signaux par analyse cepstrale	110
5.4	Etude des signaux sonores par modélisation ARMA	112
5.4.1	Processus AR	112
5.4.2	Processus MA	113
5.4.3	Processus ARMA	114
5.5	Etude des signaux par analyse en ondelettes	115
5.5.1	Définition des ondelettes	115
5.5.2	Extraction des primitives	126
5.6	Etude des signaux par les lois de Zipf	129
5.6.1	Les lois de Zipf pour l'étude statistique de phénomènes	129
5.6.2	Codage des signaux sonores	131
5.6.3	Extraction des primitives	141
5.7	Classification des signaux	144
5.8	Résultats	145
5.8.1	Analyse bas-niveau des signaux	146
5.8.2	Analyse cepstrale des signaux	147
5.8.3	Modélisation ARMA des signaux	147
5.8.4	Analyse en ondelettes des signaux	148
5.8.5	Analyse des signaux par les lois de Zipf	149
5.9	Conclusion	154
6	Détection d'événements dans des signaux bruités	157
6.1	Introduction	157
6.2	Méthode de détection multirésolution	158
6.2.1	Extraction de l'information par analyse multirésolution	159
6.2.2	Fonction de décision multicritères	163
6.3	Résultats	164
6.4	Conclusion	166

Conclusion et perspectives	167
Annexe : Environnement matériel et logiciel	171
Publications	175
Bibliographie	179

Table des figures

1.1	Système d'acquisition	16
1.2	Image radiologique centrée sur la jonction gastro-œsophagienne extraite d'une séquence vidéo.	18
1.3	Signal de déglutition stéréo. La voie 1 (a) contient les bruits cricoïdiens, et la voie 2 (b) contient les bruits xiphoïdiens.	18
2.1	Illustration du rôle de chaque étape du suivi : segmentation de l'objet à l'instant t (a), prédiction de sa position à l'instant $t + 1$ par filtrage de Kalman (b), optimisation de la position par un modèle de région déformable (c), ajustement des contours par un modèle de contours actifs (d) et rééchantillonnage des points du contour (e).	35
2.2	Différentes configurations de la fenêtre de visualisation	57
2.3	Construction du superbloc	58
2.4	Etude des différences des niveaux de gris des zones du superbloc	58
2.5	Détection du bolus	59
2.6	Images extraites du suivi du bolus pour la première séquence	61
2.7	Images extraites du suivi du bolus pour la deuxième séquence	62
3.1	Evolution de l'interface entre de la glace et de l'eau.	68
3.2	Un contour (a) et la représentation en 3 dimensions de la fonction d'ensembles de niveaux associée (b).	69
3.3	Illustration de l'algorithme de <i>transformation de distance</i>	73
3.4	Masques utilisés par l'algorithme de <i>transformation de distance</i>	74
3.5	Principe de l'algorithme de la <i>bande étroite</i>	75
3.6	Illustration de la méthode d'évolution à <i>progression rapide</i>	76
3.7	Exemple d'initialisation des points calculés et estimés sur la première image de la séquence à considérer.	85

3.8	Exemple d'initialisation des points calculés et estimés sur une image de la séquence à l'instant $t + 1$, à partir de la segmentation et de la squelettisation de l'objet dans l'image à l'instant t	85
3.9	Exemple de squelettisation d'un objet, précédée d'une phase de lissage des contours par fermeture morphologique (l'élément structurant utilisé est un carré de dimension 3×3).	86
3.10	Exemple de représentation du temps de passage du front (b) à partir de l'image initiale (a).	87
3.11	Histogramme des temps de passage du front normalisés dans l'intervalle $[0;255]$, correspondant à l'image présentée figure 3.10.	88
3.12	Segmentation de l'objet et squelettisation permettant l'initialisation dans l'image suivante.	88
3.13	Images extraites du suivi du bolus pour la première séquence	90
3.14	Images extraites du suivi du bolus mettant en évidence la gestion du changement de topologie de l'objet suivi.	91
4.1	Segmentation de l'œsophage et du SIO	99
4.2	Squelettisation des régions	100
4.3	Evolution de la coordonnée horizontale des points le long des squelettes	100
4.4	Evolution du diamètre de la région le long des squelettes	101
5.1	Pavage temps-fréquence pour la transformée à fenêtre glissante	117
5.2	Pavage temps-fréquence pour la transformée en ondelettes discrète	120
5.3	Ondelette "chapeau mexicain"	121
5.4	Analyse multirésolution	126
5.5	Analyse en paquets d'ondelettes	127
5.6	Décomposition d'un signal de déglutition jusqu'au niveau 10. Le signal du haut correspond au signal d'origine, alors que la colonne de gauche présente les signaux d'approximation successifs, et la colonne de droite, les signaux de détail successifs.	128
5.7	Principe du codage TC1.	132
5.8	Principe du codage TC2.	134
5.9	Principe du codage TC3.	135
5.10	Principe du codage FC.	136
5.11	Principe du codage TSC1.	138
5.12	Exemple de courbes de Zipf dans le cas des codages TC3 (a) et TSC1 (b).	141

5.13	Exemple de courbes de Zipf rééchantillonnées dans le cas des codages TC3 (a) et TSC1 (b).	142
5.14	Exemple de courbes de Zipf Inverse dans le cas des codages TC3 (a) et TSC1 (b).	143
6.1	Décomposition jusqu'au niveau 10 d'un signal de déglutition.	160
6.2	Exemple de l'énergie d'un signal aux niveaux de détail 6 (a) et 8 (b).	160
6.3	Décomposition d'un signal (a) et évolution de l'entropie en fonction du niveau de décomposition pour les signaux d'approximation (b), de détail (c) et pour l'entropie totale (d).	161
6.4	Exemple de l'énergie des détails de niveaux 6 (a) et 8 (b), ainsi que leur fermeture morphologique.	162
6.5	Représentation des ondelettes utilisées pour la détection.	165
6	Contexte d'acquisition en salle radiologie de l'hôpital Trousseau (Tours).	171
7	Interface logicielle d'acquisition des signaux multimédia.	172
8	Interface logicielle de visualisation des signaux multimédia	173

Liste des tableaux

1.1	Description des ensembles de classes liées aux caractéristiques de la pathologie des patients, et de leur effectif. (PSIO signifie Pression du SIO)	19
5.1	Effectifs des différents ensembles constitués des classes c1 et c2, en fonction des types de primitives considérées : bas-niveau (1), cepstre (2), arma (3), ondelettes (4) et Zipf (5).	146
5.2	Résultat de la classification pour l'ensemble de primitives bas-niveau.	146
5.3	Résultat de la classification pour l'ensemble de primitives liées à une analyse cepstrale.	147
5.4	Résultat de la classification pour l'ensemble de primitives liées à une modélisation ARMA des signaux.	148
5.5	Résultat de la classification pour l'ensemble de primitives issues de l'analyse en ondelettes.	149
5.6	Résultat de la classification pour le codage TC1.	149
5.7	Résultat de la classification pour le codage TC2.	150
5.8	Résultat de la classification pour le codage TC3.	150
5.9	Résultat de la classification pour le codage FC.	150
5.10	Résultat de la classification pour le codage TSC1.	150
5.11	Résultat de la classification pour le codage TSC2.	151
5.12	Taux maximum de reconnaissance (Tmr) pour les différents ensembles de classes.	151
5.13	Discrimination <i>bruits de fond - bruits xiphoïdiens</i>	152
5.14	Résultat de la classification en utilisant un algorithme génétique pour la sélection de primitives.	154
6.1	Pourcentage de détection pour les ondelettes Daubechies 2 et 4, Coiflet 3 et Meyer.	166

Chapitre 1

Introduction

1.1 Analyse simultanée de vidéos et de sons pour l'interprétation de séquences multimédia

Les séquences multimédia font l'objet de nombreuses études puisqu'elles représentent un support de l'information riche et varié, pouvant contenir simultanément des données sonores, textuelles, et graphiques animées ou non. La diffusion de ces séquences est réalisée de manière de plus en plus aisée grâce à la démocratisation de l'outil informatique, par l'intermédiaire de CD-Roms, de DVD-Roms, et récemment d'Internet par des lignes à haut débit. La quantité d'information disponible est ainsi extrêmement importante. Cependant, cette richesse n'a d'intérêt que lorsque l'information peut être facilement et rapidement exploitée.

C'est pourquoi de nombreuses recherches sont réalisées concernant l'analyse de séquences multimédia dont l'objectif est l'extraction et l'interprétation de l'information contenue dans les séquences, afin de permettre la compréhension de leur contenu.

Dans le cadre de cette thèse de doctorat, nous nous intéressons à l'analyse du contenu de séquences multimédia constituées d'images et de données sonores. Ces deux types de données synchronisées dans le temps sont souvent corrélées. L'objectif que nous nous proposons est la compréhension du contenu global de la séquence par l'étude de la seule information sonore. Cependant, l'information relative à la séquence d'images est ici également extraite dans un premier temps afin de permettre la validation des résultats.

Les séquences multimédia, support de notre étude, correspondent à des séquences médicales, l'objectif étant d'extraire l'information relative aux caractéristiques de la pathologie du reflux gastro-œsophagien afin de fournir aux médecins experts un outil d'indexation et d'aide à l'interprétation et à la décision dans le cas de l'établissement d'un diagnostic. Le contexte médical est présenté dans la section suivante.

1.2 Application au problème médical du reflux gastro-œsophagien

Le sphincter inférieur de l'œsophage (SIO) joue un rôle primordial dans le phénomène de reflux gastro-œsophagien (RGO). Le diaphragme crural qui entoure le SIO participe également à la fonction anti-reflux. Les mécanismes physio-pathologiques du RGO sont liés au dysfonctionnement du SIO qui apparaît alors soit hypotonique soit soumis à des relaxations transitoires avec reflux importants. Il peut également être incapable de s'adapter à des élévations de pression abdominale.

Dans 90% des cas, les patients possédant un RGO souffrent également d'une hernie hiatale caractérisée par l'intrusion dans le thorax d'une portion plus ou moins importante de l'estomac proximal appelé fundus. L'influence de la hernie hiatale dans le RGO est encore discutée. Lorsque le traitement médical n'est pas satisfaisant (inhibiteurs de type histaminique et inhibiteurs de la pompe à protons), l'intervention chirurgicale est alors envisagée. Le but des montages chirurgicaux est de réduire la hernie hiatale et de renforcer le tonus du SIO par la fundoplicature de Nissen (effet de manchon) et la diminution du diamètre de l'orifice hiatal (resserrage par les piliers du diaphragme).

Les méthodes actuelles d'investigation permettant la détection de la pathologie reposent sur des explorations manométriques et radiologiques, qui présentent l'inconvénient d'être invasives.

C'est pourquoi une méthode d'enregistrement des bruits de déglutition a été créée et mise au point par Boiron et col [BOI97c, BOI97b, BOI97a, BOI98] au sein de l'équipe « Acoustique et Motricité Digestive » du Laboratoire de Physiologie et Motricité Digestive de la Faculté de Médecine de Tours, en collaboration avec les Services de Radiologie, Gastroentérologie et Chirurgie Viscérale des CHU de cette même ville. Grâce à cette technique d'enregistrement, une méthode d'investigation acoustique peut être envisagée, et aurait l'avantage d'être non invasive, atraumatique

et simple d'utilisation. En effet, elle reposerait uniquement sur l'enregistrement et l'analyse des signaux de déglutition contenant les bruits xiphoïdiens produits lors du passage d'un bolus alimentaire à travers l'œsophage et le sphincter inférieur de l'œsophage.

Afin de réaliser le développement informatique de la méthode acoustique, il est nécessaire d'étudier de manière parallèle les signaux radiologiques et sonores de manière à en extraire des propriétés caractérisant le comportement du SIO, et d'en étudier ensuite les corrélations pour permettre, avec l'aide des experts en ce domaine, une bonne interprétation des phénomènes de déglutition et de reflux gastro-œsophagien. Afin d'aider les experts dans cette démarche nous avons développé un logiciel permettant de visualiser simultanément les images radiologiques et les signaux sonores, offrant ainsi la possibilité de déterminer des corrélations entre les deux sources.

Pour réaliser cette étude, 111 séquences acoustico-radiologiques ont été acquises en salle de radiologie du Service de Radiologie du CHU Trousseau de Tours. Ces séquences sont centrées sur la jonction œso-gastrique, et ont été enregistrées chez 11 patients par l'intermédiaire d'un logiciel d'acquisition que nous avons développé afin de permettre la gestion d'une carte d'acquisition vidéo Imasys exploitant le signal vidéo composite mis à notre disposition, constitué d'images de dimension 1112x1024 pixels cadencées à la vitesse de 25 images par seconde.

Les bruits ont été recueillis au moyen de deux microphones cravates placés en regard du cartilage cricoïde et de l'appendice xiphoïde. Les enregistrements ont été réalisés sur l'entrée Line-In d'une carte sonore Marian échantillonnant le signal stéréo à 44,1 KHz.

Nous présentons figure 1.1 le schéma global du système d'acquisition.

1.2.1 Analyse des images radiologiques

L'objectif de l'analyse des données radiologiques est de déterminer les caractéristiques du SIO pouvant être liées à la pathologie. Pour cela, le suivi du bolus baryté dans les séquences d'images radiologiques, lors de sa descente à l'intérieur de l'œsophage et de son passage à travers le SIO doit être réalisé. Il sera alors possible de mesurer, de manière indirecte, certains paramètres tels que le diamètre maximal du SIO, sa dynamique, la longueur du manchon réalisé après fundoplicature de Nis-

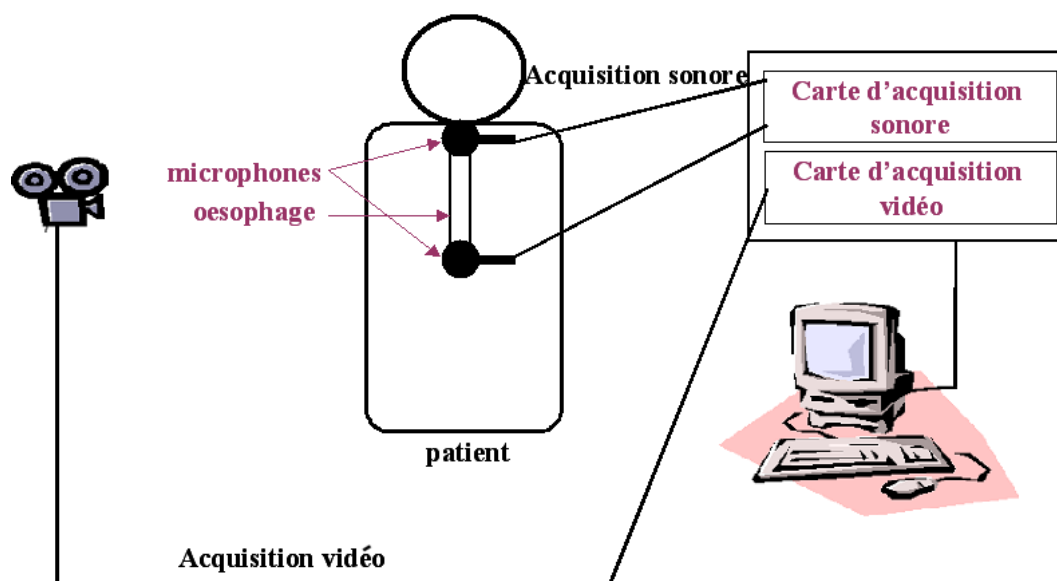


FIG. 1.1 – Système d'acquisition

sen, ainsi que l'évolution de la vitesse de déplacement du bolus lors de sa traversée. Le suivi d'objet est celui d'un objet déformable. Mais avant de le suivre, nous devons d'abord le détecter. Les algorithmes de suivi que nous proposons, reposent, pour le premier, sur une combinaison de suivi de région par un modèle de régions déformables et de contour par un modèle de contours actifs à minimisation locale présenté au chapitre 2 [DEL00], et pour le deuxième, sur un modèle de contours actifs non régularisés présenté au chapitre 3. La caractérisation du SIO grâce au suivi est présentée au chapitre 4 [DEL03c].

Les données radiologiques vont maintenant être décrites.

Principe physique

Les rayons X ont pour propriété de traverser la matière. Cependant, ils sont absorbés et diffusés par celle-ci. L'absorption dépend de l'épaisseur et de la densité du matériau traversé. Si la source émet un rayonnement d'énergie suffisante pour ne pas être complètement absorbé par l'objet, alors le rayonnement transmis est porteur d'une certaine information. Ce rayonnement, appelé rayonnement résiduel, est enregistré par un récepteur dont la réponse est fonction de la quantité de rayonnement reçue.

Plusieurs types de récepteurs sont présentés dans [GAN99]. L'écran radiosco-

pique, utilisé dans un premier temps, a été abandonné car seulement 5% des rayons X étaient transformés en lumière. Les images radioscopiques présentaient donc une très faible luminosité, et étaient peu exploitables.

L'amplificateur de luminance a alors été développé afin d'améliorer le rendement du processus. Grâce à un système d'accélération des électrons et de focalisation électronique, le signal reçu est beaucoup plus important. Un dispositif optique permet de le transformer en un signal vidéo qui peut être numérisé. Des séquences dynamiques peuvent ainsi être réalisées jusqu'à 25 images par seconde.

Caractéristiques des images radiologiques

Les images radiologiques sont caractérisées par un aspect granulaire, par de faibles contrastes entre les objets et le fond de l'image, ainsi que par l'absence de netteté du contour des objets.

Cela met en évidence les difficultés de la segmentation d'images radiologiques. Il peut être utile de procéder à des prétraitements tels que la modification de la dynamique et de la répartition des niveaux de gris sur toute la gamme disponible, afin d'augmenter les contrastes.

Présentation des séquences d'images étudiées

La durée des séquences est d'une quinzaine de secondes chacune, à la cadence de 25 images par seconde.

Les images se présentent sous forme rectangulaire de taille 1112x1024 pixels. Un cache se situe à la périphérie et entoure une fenêtre de visualisation dont la forme peut varier en fonction des séquences (forme circulaire ou rectangulaire). Une image extraite d'une séquence est présentée figure 1.2.

1.2.2 Analyse des données sonores

Les signaux sonores auxquels nous nous intéressons, sont des signaux stéréo de déglutition contenant les bruits cricoïdiens sur la première voie, et les bruits xiphoidiens sur la seconde. L'objectif de l'étude de ces signaux sonores est de déterminer si le comportement du SIO, peut être caractérisé par les bruits xiphoidiens. Un exemple de signal de déglutition est présenté figure 1.3.

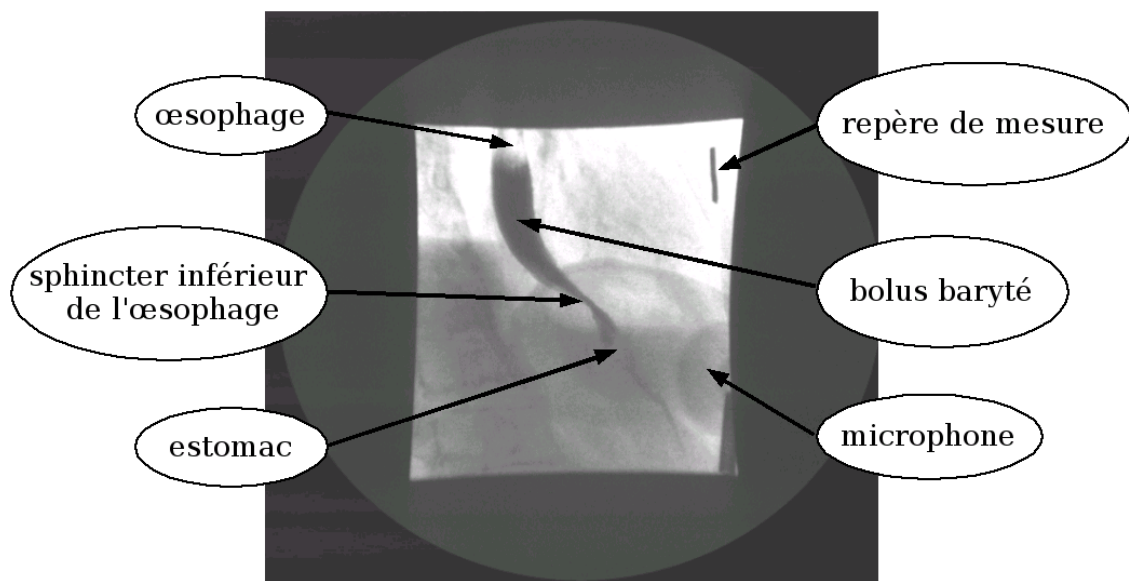


FIG. 1.2 – Image radiologique centrée sur la jonction gastro-œsophagienne extraite d'une séquence vidéo.

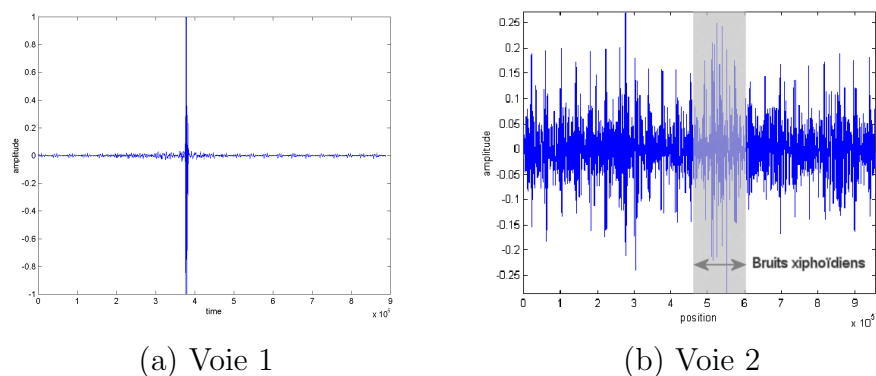


FIG. 1.3 – Signal de déglutition stéréo. La voie 1 (a) contient les bruits cricoïdiens, et la voie 2 (b) contient les bruits xiphoïdiens.

Le problème consiste donc d'une part à détecter les bruits xiphoïdiens au sein des signaux de déglutition, et d'autre part à caractériser les bruits xiphoïdiens en fonction des propriétés liées à la pathologie des patients.

Pour cela, différents ensembles de classes ont été considérés, dépendant de l'information médicale à prendre en considération. Ils sont présentés TAB. 1.1.

L'étude des bruits xiphoïdiens est un nouveau domaine de recherche, développé par Boiron et al [BOI97c,BOI97b,BOI97a,BOI98]. Les classes définies dans TAB. 1.1

Ensembles	Classes	Description	Nb Patients
ba	c1	avant opération	11
	c2	après opération	11
hiat1	c1	hernie hiatale	5
	c2	pas d'hernie hiatale	6
hiat2	c1	hernie hiatale avec une faible PSIO	5
	c2	pas d'hernie hiatale avec une faible PSIO	6
oo	c1	petite ouverture de l'œsophage	5
	c2	grande ouverture de l'œsophage	6
vh	c1	petite hauteur de valve	7
	c2	grande hauteur de valve	4
od	c1	petit diamètre œsophagien	5
	c2	grand diamètre œsophagien	6
ao	c1	court œsophage abdominal	4
	c2	long œsophage abdominal	7

TAB. 1.1 – Description des ensembles de classes liées aux caractéristiques de la pathologie des patients, et de leur effectif. (PSIO signifie Pression du SIO)

ont été établies à partir de l'analyse des données acoustico-radiologiques (ensemble xb) (Dr. M. Boiron, Dr. Z. Benchellal), des données chirurgicales (ensemble ba) (Dr. N. Hutten, Dr. Z. Benchellal), des explorations manométriques (ensembles hiat1, hiat2) (Dr. M. Boiron, Dr. E.H. Metman), ainsi que des images radiologiques (ensembles oo, vh, od, ao) (Dr B. Scotto, Dr. Z. Benchellal). Ainsi, l'intérêt de l'application d'un système d'identification sonore automatique est de déterminer si l'information contenue dans les données médicales peut être extraite à partir de l'analyse des données sonores.

Nous avons donc évalué, dans le chapitre 5, des systèmes d'identification de signaux sonores fréquemment utilisés notamment dans les domaines biologique, de reconnaissance de parole et de musique. Nous avons également développé un système d'identification de signaux sonores reposant sur l'analyse de la structure interne des signaux par les lois de Zipf [DEL03e, DEL03g, DEL03f, DEL03a]. Enfin, nous avons mis au point une méthode rapide de détection des bruits xiphoidiens au sein des signaux de déglutition, qui est présentée dans le chapitre 6.

1.3 Organisation du mémoire

La **première partie** de ce mémoire est consacrée à l'analyse de signaux vidéos. L'objectif est le suivi d'objets déformables dans des séquences d'images, et l'étude de l'évolution de leurs propriétés au cours du temps.

Ainsi, le **deuxième chapitre** propose un aperçu des méthodes de suivi et des modèles de mouvement les plus utilisés. La méthode de suivi que nous avons développée, combine un suivi de région par un modèle de régions déformables, et un suivi de contours par un modèle de contours actifs à minimisation locale. Cela rend possible le suivi d'un objet de forme et de texture quelconques. Nous décrirons ensuite son application au suivi d'un bolus baryté dans des séquences d'images radiologiques.

Le **troisième chapitre** est consacré à la présentation d'une seconde méthode de suivi développée pour apporter une solution aux problèmes non résolus par la méthode précédente : complexité calculatoire importante, réglage nécessaire de plusieurs paramètres interdisant une réelle automatisation de la méthode, non gestion du changement de topologie. La méthode de suivi développée, reposant sur la théorie des ensembles de niveaux, présente cependant une limite par rapport à la méthode précédente : elle ne peut pas s'appliquer à des objets de texture quelconque, mais à des objets de texture uniforme ce qui est le cas du bolus baryté dans les séquences d'images radiologiques, utilisées pour valider la méthode.

Dans le **quatrième chapitre**, un certain nombre de paramètres permettant la caractérisation d'objets dans des images sera présenté. L'objectif est de permettre la détermination de zones d'intérêt dans les images radiologiques centrées sur la jonction œso-gastrique et l'étude de l'évolution temporelle des caractéristiques d'un bolus baryté, permettant, de manière indirecte, l'étude des organes traversés par ce bolus, et en particulier le sphincter inférieur de l'œsophage. Les informations ainsi obtenues pourront alors être utilisées par les experts physiologues et médicaux afin d'aider à l'interprétation du reflux gastro-œsophagien.

La **deuxième partie** porte sur l'étude de signaux sonores. L'objectif consiste à identifier les signaux afin de permettre l'interprétation du contenu sonore.

Le **cinquième chapitre** présente donc le principe des systèmes dédiés à l'identification de signaux sonores. Trois méthodes d'extraction de l'information sonore fréquemment utilisées sont ensuite décrites : l'analyse de la couche bas-niveau des signaux sonores, l'analyse cepstrale et la modélisation des signaux par modèles

ARMA. Puis, deux méthodes que nous avons développées seront présentées, la première reposant sur une analyse par ondelettes multirésolution, et la seconde sur une adaptation aux signaux sonores de l'analyse de la structure de phénomènes par les lois de Zipf et Zipf Inverse. Nous verrons ensuite comment nous avons intégré ces primitives dans un système de reconnaissance sonore permettant la sélection des primitives les plus pertinentes par algorithme génétique, et la classification par un système multi-classifieurs. Enfin, l'application de cette méthode à des signaux sonores de déglutition a été réalisée, l'objectif étant de caractériser la pathologie de reflux gastro-œsophagien par l'identification des propriétés des signaux sonores.

Enfin, le **sixième chapitre** est consacré à la présentation d'une technique que nous avons développée afin de détecter des événements transitoires dans des signaux sonores bruités. Nous l'avons appliquée à la détection de bruits xiphoïdiens dans des signaux de déglutition bruités.

Première partie

Analyse de signaux vidéos

Chapitre 2

Suivi d'objets déformables

2.1 Introduction

Une des problématiques que doit résoudre l'analyse de séquences vidéos consiste à déterminer les objets en mouvement, puis à étudier leur évolution au cours du temps, dans les différentes images. La première étape consiste en une segmentation spatio-temporelle pour mettre en évidence l'image initiale, et la deuxième phase correspond au suivi des objets détectés dans la séquence.

Nous nous intéressons plus particulièrement dans ce chapitre à la phase de suivi. La problématique que nous nous proposons consiste à suivre un objet déformable de forme et de texture quelconques soumis à un mouvement quelconque pouvant être de forte amplitude, dans des séquences d'images en 256 niveaux de gris.

Un bref aperçu des algorithmes de suivi sera donc présenté, ainsi que les modèles de mouvement les plus utilisés. Nous présenterons ensuite l'algorithme de suivi que nous avons développé afin de permettre le suivi d'un objet déformable satisfaisant aux contraintes fixées. Enfin, l'application de la méthode au suivi d'un bolus baryté dans des séquences d'images radiologiques sera proposée.

2.2 Algorithmes de suivi

Le suivi d'objets dans une séquence d'images est un problème classique et largement étudié dans le domaine de l'analyse du mouvement, du fait de ses nombreuses applications parmi lesquelles on peut noter la surveillance vidéo [RIC97], la robo-

tique, la compression de vidéos, l'étude de la déformation d'objets, ... Ainsi, de nombreux algorithmes de suivi ont été proposés ces dernières années. Nous allons, dans les sections suivantes, présenter une classification possible de ces algorithmes, ainsi que leur structure générale.

2.2.1 Classification des algorithmes de suivi

Les algorithmes de suivi peuvent globalement être classés en deux catégories : les algorithmes orientés mouvement, et ceux orientés modèle.

Les **algorithmes orientés mouvement** s'appuient sur des méthodes de regroupement de régions homogènes au sens du mouvement. Ils ont l'avantage d'être rapides, cependant, les régions suivies n'ont en général pas de signification physique globale puisqu'elles ont été obtenues à partir d'un critère d'homogénéité et non d'une représentation d'objet. Le principe de ces algorithmes est de partir d'une segmentation spatio-temporelle initiale [FRA91, VM01]. Cette dernière peut être obtenue de différentes façons : à partir de pixels individuels (techniques Markoviennes fournissant un champ dense [GEL98]), de petits blocs de pixels (approche MPEG), d'une grille arbitraire [NIC92, NIC93], de l'image entière, ou encore d'une segmentation spatiale [BON98, BON94]. Les paramètres de mouvement sont ensuite estimés à partir de cette segmentation, puis des fusions ou des divisions de régions sont réalisées selon un critère d'uniformité du mouvement. Cette segmentation est ensuite suivie au cours de la séquence. Ce type d'algorithme est utilisé pour la compression [GG93, BON94], le codage [GG95], l'interpolation [BON98] et l'indexation [GEL98] de vidéos. [BAR96] y ont également recours pour réaliser la détection et le suivi de fouillis radar.

Les **algorithmes orientés modèle**, contrairement aux précédents, s'attachent à suivre des régions correspondant aux projections 2D des objets auxquels on s'intéresse, en les représentant soit par des modèles déformables 3D [LOW92], soit par des modèles déformables 2D [BAS94b, BAS94a, PAR99, NAS93, BIT98]. Les régions suivies sont donc plus significatives, mais la complexité calculatoire est plus importante du fait de la nécessité de déterminer d'éventuelles rotations, translations, homothéties ou autres déformations. Ces algorithmes permettent l'analyse de la déformation d'objets au cours de séquences vidéo : dans [NAS93], par exemple, est étudiée la déformation du ventricule gauche humain dans des séquences d'images échographiques.

Dans les deux cas, le suivi peut être réalisé par des méthodes basées soit sur les frontières, soit sur les régions. Il est à noter que les premières recherches dans le domaine s'appuyaient sur le suivi de primitives simples telles que des segments de contours ou des points de référence dans les objets [BRO86]. Cependant, ces techniques étaient peu fiables du fait de données trop éparées, et sensibles aux occlusions. C'est pourquoi les recherches se sont orientées vers des primitives plus complexes et plus globales telles que les contours et les régions.

Les **méthodes basées sur les frontières** reposent sur l'information apportée par le contour de l'objet à suivre. Leur avantage est d'être indépendantes du type de mouvement ainsi que de l'objet. De plus, elles sont efficaces pour extraire les contours de l'objet avec précision. Elles sont donc particulièrement bien adaptées au suivi d'objets déformables. Ce sont en général des modèles de contours déformables tels que les contours actifs (snakes) [KAS98,AMI90,WIL92,XU91], les contours actifs géodésiques [CAS93] ou encore des modèles statistiques [KER95,CHE00] qui sont utilisés. Ainsi, les méthodes proposées dans [BON94,BON98,BAR96] sont basées sur ces modèles. Les principaux inconvénients de ces modèles sont leur instabilité due à la prise en compte d'une information locale, ainsi que leur sensibilité à l'initialisation. En effet, le principe de ces algorithmes est de partir d'un contour initial et de converger de façon itérative vers le contour recherché, par minimisation d'une fonction d'énergie. Ils sont donc performants dans le cas de mouvements lents, mais mal adaptés aux déplacements importants.

Les **méthodes basées sur les régions** utilisent, quant à elles, l'information apportée par la région entière, comme par exemple des propriétés de texture ou de mouvement. Bien qu'ayant des difficultés à suivre avec précision les contours, ces méthodes sont plus robustes puisqu'elles utilisent une information plus globale et qu'elles sont moins dépendantes de l'initialisation. Elles sont ainsi capables de gérer des déplacements importants. De tels modèles ont été développés dans [PAR99,PAR00] en utilisant des modèles de régions déformables géodésiques, et dans [JB03] en utilisant des contours actifs basés région. De plus, en associant au suivi de région une segmentation de régions basée mouvement [MEY94], il est même possible de s'affranchir de toute initialisation puisque cette segmentation détecte les objets en mouvement dans l'image.

Ces deux types de méthodes possèdent donc des qualités complémentaires. En les combinant, certains algorithmes permettent un suivi optimisé [BAS94b,BAS94a,PAR99].

2.2.2 Structure générale des algorithmes de suivi

Le suivi des primitives présentées précédemment d'un instant t vers un instant $t + 1$, est généralement composé d'un minimum de trois étapes, après avoir calculé une segmentation spatiale ou spatio-temporelle initiale, selon le type d'algorithme de suivi considéré. Ces trois étapes sont les suivantes :

1. **Prédiction** du mouvement des primitives de l'instant t à l'instant $t + 1$
2. **Ajustement** de la segmentation prédite
3. **Estimation** du mouvement des régions suivies de t vers $t + 1$

Nous allons brièvement considérer chacune de ces trois étapes.

Phase de prédiction

Elle a pour objectif de relier la segmentation de l'image à l'instant t à celle de l'image à l'instant $t + 1$. Quatre modèles de prédiction peuvent être envisagés [GG93, GG95, BON98] :

- **Prédiction de mouvement nul** : on considère que la primitive ne se déplace pas entre t et $t + 1$. Sa position prédite est donc identique à celle qu'elle avait à l'instant précédent.
- **Prédiction à court-terme simple** : on prend uniquement en compte l'information du mouvement entre les deux images précédentes. Le terme *court-terme* vient du fait que dans ce cas, on n'utilise aucun modèle temporel d'évolution.
- **Prédiction à court-terme lissée** : on utilise alors un modèle d'évolution, tel que le filtre de Kalman. Le mouvement considéré est celui fourni par les équations de filtrage (ou lissage) de ce filtre.
- **Prédiction à long-terme** : dans ce cas, on prend en compte l'information du mouvement entre toutes les images précédentes. Les valeurs prédites sont alors issues de l'équation de prédiction du filtre de Kalman. Nous préciserons ultérieurement de manière plus détaillée, ces deux dernières méthodes.

Phase d'ajustement

La phase de prédiction réalisée, nous disposons d'une segmentation prédite de l'image à l'instant $t + 1$. Cependant, cette dernière n'est en général pas parfaite. Il est donc nécessaire d'ajuster les frontières des régions de cette segmentation sur les frontières réelles. C'est le rôle de la phase d'ajustement.

Selon le type de primitives suivies, deux types de méthodes d'ajustement peuvent être envisagés :

- **Ajustement basé contour** : ces méthodes utilisent des modèles de contours déformables afin d'ajuster les frontières des régions prédites, en se basant uniquement sur des informations spatiales, telles que le gradient spatial, dans l'image courante.
- **Ajustement basé région** : cette approche peut s'appuyer, par exemple, sur une modélisation markovienne des zones d'occlusion [GG93, GG95], ou encore sur des modèles de régions déformables [BAS94a, BAS94b]. Elle peut prendre en considération à la fois des informations spatiales dans l'image courante (gradient spatial, texture,...), et des informations de mouvement entre les images courante et précédente.

Phase d'estimation

Elle est réalisée après la phase d'ajustement et permet de déterminer quel a été le mouvement effectif des régions suivies entre les images aux instants t et $t + 1$. Il existe trois grands types de méthodes d'estimation du mouvement [NIC92] :

- **Les méthodes par mise en correspondance** : elles cherchent à déterminer des similarités entre des points, des contours ou des régions, présents dans deux images successives. Pour cela, une fonction de corrélation doit être maximisée. Ce sont en général des régions constituées de petits blocs carrés qui sont considérées. Seules les translations sont tolérées, car ce modèle suppose que tous les pixels appartenant à une même région subissent un déplacement identique. Les normes MPEG, par exemple, ont recours à ces méthodes.
- **Les méthodes par transformées** : elles utilisent les effets du mouvement

dans l'image sur les composantes ou les caractéristiques fréquentielles de transformées telles que la transformée de Fourier ou encore la décomposition en séries de Gabor. En effet, une translation spatiale, par exemple, modifie la phase de la transformée de Fourier de l'image.

- **Les méthodes différentielles** : elles reposent sur l'hypothèse d'invariance de la luminance des points au cours du temps, lors de leur déplacement. L'estimation est alors réalisée par la minimisation d'une fonction d'énergie prenant en compte soit le carré de la DFD mesurant la différence de luminance d'un point d'une image et son projeté dans l'image suivante (2.1), soit l'Equation de Contrainte du Mouvement Apparent ($ECMA$) obtenue par un développement de Taylor à l'ordre 1 de la DFD (2.2).

$$DFD^2(p, \vec{d}) = (I_{t+1}(p + \text{vecd}) - I_t(p))^2 \quad (2.1)$$

où p est un point de l'image I_t à l'instant t , et $p + \text{vecd}$ son projeté dans l'image I_{t+1} à l'instant $t + 1$.

$$ECMA \quad : \quad \frac{\partial I}{\partial x}(x, y, t).dx + \frac{\partial I}{\partial y}(x, y, t).dy + \frac{\partial I}{\partial t}(x, y, t).dt = 0 \quad (2.2)$$

où (x, y) sont les coordonnées d'un point de l'image I .

Conclusion

Notre objectif étant d'étudier le mouvement et surtout les déformations d'un objet de forme et de mouvement quelconques à partir d'une séquence vidéo, nous avons choisi une méthode orientée modèle que nous présenterons au chapitre 2.3. De plus, afin de rendre le suivi plus fiable, nous utiliserons une combinaison de suivi de région par un modèle de région déformable, et de suivi de contour par un modèle de contours actifs. Nous employerons également le filtre de Kalman comme modèle d'évolution temporelle afin de permettre une prédiction long-terme précise. Enfin, l'estimation du mouvement est réalisée par la minimisation de l'Erreur Quadratique Moyenne de Reconstruction de l'image qui sera décrite à la section 2.3.6, et basée sur le carré de la DFD .

Après avoir déterminé le type de méthode de suivi répondant le mieux au pro-

blème posé, il est possible de modéliser le mouvement des objets à suivre. C'est le rôle des modèles de mouvement. Nous allons présenter les différents modèles les plus couramment utilisés.

2.2.3 Modèles de mouvement

Cette partie est destinée à la présentation des modèles de mouvement les plus courants. Cependant, il est nécessaire avant de décrire ces derniers, d'introduire le concept de mouvement.

Concept de mouvement

Le mouvement d'un objet entre deux images consécutives I_t et I_{t+1} observées aux instants t et $t + 1$ correspond au déplacement des points (pixels) qui le constituent.

Soit p un point qui se déplace entre les images I_t et I_{t+1} de la position p_t à la position p_{t+1} . Il est possible d'exprimer son vecteur déplacement de deux manières différentes selon le sens de l'axe temporel considéré :

- Dans le sens croissant : $\vec{d}_t^+(p) = p_{t+1} - p_t$. Ce sens sera utilisé pour la prédiction du mouvement entre l'image I_t et l'image I_{t+1} .
- Dans le sens décroissant : $\vec{d}_{t+1}^-(p) = p_t - p_{t+1}$. Ce sens sera utilisé pour l'estimation du mouvement entre l'image I_t et l'image I_{t+1} .

En général, on considère les points dans l'image I_{t+1} et on cherche leur correspondant dans l'image I_t . La formule de reconstruction de l'image est la suivante :

$$\hat{I}_{t+1}(p) = I_t(p + \vec{d}_{t+1}^-(p))$$

où $\hat{}$ signifie "reconstruction".

Il est nécessaire de déterminer le vecteur déplacement pour chacun des points de l'objet. On utilise pour cela des modèles de mouvement.

Modèles paramétriques de mouvement 2D

Un modèle de mouvement permet de déterminer le vecteur déplacement de chaque point de la région considérée. Ce n'est pas au mouvement réel en trois di-

mensions de l'objet que nous nous intéressons, mais à sa projection dans un espace à deux dimensions : c'est ce qu'on appelle le *mouvement apparent* ou *mouvement 2D*. Les modèles de mouvement qui suivent ont donc pour objet la modélisation du mouvement apparent.

Pour un modèle de mouvement donné, le descripteur de mouvement Θ_R de la région R contient les paramètres du mouvement. Il permet d'obtenir le vecteur déplacement des points p , de coordonnées (x, y) , de la région R selon la formule générale :

$$\vec{d}_R \begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{M} \left(\Theta_R, \begin{pmatrix} x - x_G \\ y - y_G \end{pmatrix} \right)$$

où \mathbf{M} représente un modèle de mouvement, et (x_G, y_G) représente le couple des coordonnées du centre de gravité de la région R .

On distingue en particulier trois types de modèles : les modèles affines, quadratiques et homographiques.

Modèle affine complet

$$\vec{d} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} t_x \\ t_y \end{pmatrix} + \begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} x - x_G \\ y - y_G \end{pmatrix}$$

où le vecteur $\begin{pmatrix} t_x \\ t_y \end{pmatrix}$ représente les paramètres de translation donnés par le mouvement du centre de gravité de la région.

Ce modèle prend en compte les mouvements de translation, de rotation, de changement de facteur d'échelle, ainsi que les affinités et les cisaillements. C'est le modèle utilisé par [BAS94a, BAS94b, MEY94, PAR99].

Le descripteur de mouvement est le suivant :

$$\Theta = [t_x, t_y, a, b, c, d]$$

Des simplifications de ce modèle peuvent être obtenues :

Lorsque $a = d$ et $b = -c$, on obtient un modèle affine simplifié. Il permet de modéliser des mouvements de translation, de rotation et de grossissement. Son descripteur de mouvement est : $\Theta = [t_x, t_y, a, b]$.

Lorsque $a = b = c = d = 0$, on obtient un modèle constant, correspondant à des mouvements de translation uniforme. Son descripteur de mouvement est : $\Theta = [t_x, t_y]$.

Lorsque $t_x = t_y = a = b = c = d = 0$, on obtient un modèle nul correspondant à une région qui n'est pas en mouvement. Son descripteur de mouvement est : $\Theta = 0$.

Modèle quadratique

$$\begin{aligned} \vec{d} \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} t_x \\ t_y \end{pmatrix} + \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix} \begin{pmatrix} x - x_G \\ y - y_G \end{pmatrix} \\ &+ \begin{pmatrix} a_3 & a_4 \\ b_3 & b_4 \end{pmatrix} \begin{pmatrix} (x - x_G)^2 \\ (y - y_G)^2 \end{pmatrix} + \begin{pmatrix} a_5 \\ b_5 \end{pmatrix} (x - x_G)(y - y_G) \end{aligned}$$

Ce modèle complète le modèle affine en lui ajoutant des termes du second degré. Il permet de décrire les changements de forme d'objets non-rigides, cependant il est rarement utilisé car les termes quadratiques sont difficiles à estimer.

Le descripteur de mouvement est le suivant :

$$\Theta = [t_x, t_y, a_1, b_1, a_2, b_2, a_3, b_3, a_4, b_4, a_5, b_5]$$

Modèle homographique

$$\vec{d} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{N_{xx}(x-x_G) + N_{xy}(y-y_G) + N_x}{D_x(x-x_G) + D_y(y-y_G) + 1} \\ \frac{N_{yx}(x-x_G) + N_{yy}(y-y_G) + N_y}{D_x(x-x_G) + D_y(y-y_G) + 1} \end{pmatrix}$$

où N_{xx} , N_{xy} , N_{yx} , N_{yy} , N_x , N_y , D_x et D_y sont les paramètres du modèle.

Ce modèle a la capacité de décrire exactement tous les mouvements d'un objet 3D projeté sur un espace plan. Cependant, du fait de la difficulté de la mesure de ses paramètres, il est rarement employé.

Le descripteur de mouvement est le suivant :

$$\Theta = [N_{xx}, N_{xy}, N_{yx}, N_{yy}, N_x, N_y, D_x, D_y]$$

Le modèle de mouvement devra être choisi en fonction du type d'objet à suivre, et des contraintes de précision à satisfaire. Nous avons choisi d'utiliser le modèle de

mouvement affine à six paramètres car il permet de modéliser de façon satisfaisante la majorité des mouvements, tout en étant d'une relative simplicité.

Après cette courte présentation des méthodes que nous avons sélectionnées au cours de notre étude, nous allons montrer comment elles ont pu être mises en œuvre dans le cas du suivi d'un objet déformable dans des séquences d'images.

2.3 Suivi temporel d'un objet déformable

Cette section exposera la méthode que nous avons élaborée afin de suivre un objet déformable dans des séquences d'images.

2.3.1 Aperçu de la méthode de suivi

L'objet que nous nous attachons à suivre est un objet non-rigide dont nous devons étudier les déformations. C'est pourquoi nous nous sommes orienté vers un algorithme de suivi de type modèle. De plus, afin d'améliorer la robustesse du suivi, nous avons élaboré une méthode inspirée des travaux de [BAS94b] et de [BON94, BON98], combinant un suivi de région et un suivi de contour de manière à produire une plus grande précision de la région recherchée.

L'algorithme de suivi se compose de cinq modules principaux :

- **prédiction** des paramètres de mouvement Θ_t^+ de la région par filtrage de Kalman afin d'améliorer l'initialisation du modèle dans l'image à l'instant $t + 1$, connaissant sa position dans l'image à l'instant t .
- **optimisation** de la région prédite par un modèle de région déformable afin d'assurer une plus grande robustesse du suivi dans le cas de déplacements importants de l'objet et d'images bruitées.
- **ajustement** du contour de la région sur le contour réel par contours actifs afin d'améliorer la précision du contour de l'objet segmenté.
- **rééchantillonnage** des points de contour de la région afin de permettre une adaptation du contour de l'objet déformable.

- **estimation** des paramètres de mouvement Θ_{t+1}^- de la région afin de mesurer le mouvement de l'objet entre les images aux instants t et $t + 1$.

Le rôle de chaque étape du suivi est illustré par le schéma présenté figure 2.1. L'exemple adopté consiste en le suivi entre deux images successives d'un objet déformable ayant un mouvement de rotation.

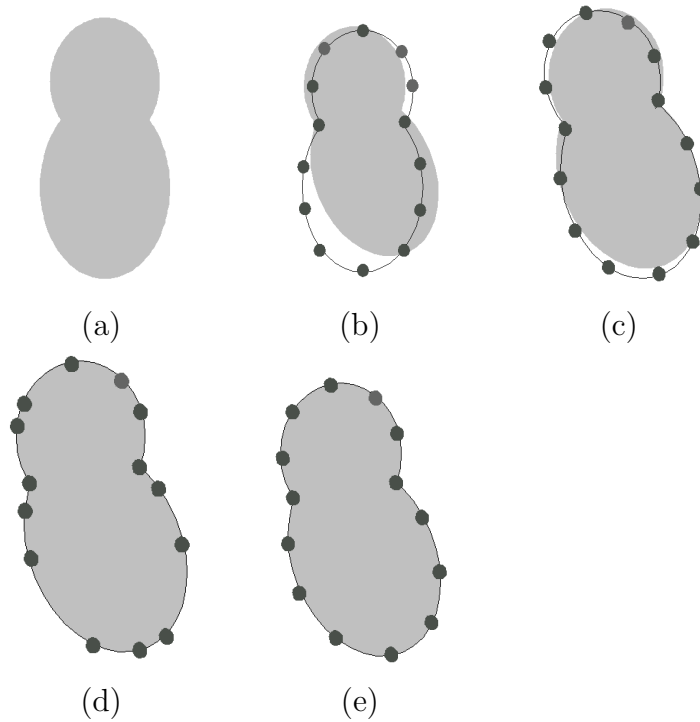


FIG. 2.1 – Illustration du rôle de chaque étape du suivi : segmentation de l'objet à l'instant t (a), prédiction de sa position à l'instant $t + 1$ par filtrage de Kalman (b), optimisation de la position par un modèle de région déformable (c), ajustement des contours par un modèle de contours actifs (d) et rééchantillonnage des points du contour (e).

2.3.2 Prédiction des paramètres de mouvement

La prédiction des paramètres de mouvement correspond à la première étape de l'algorithme de suivi. Son rôle est d'essayer de déterminer la position de la région à l'instant $t + 1$, connaissant sa position à l'instant t . Or, la phase d'initialisation est critique car il est souvent difficile voire impossible de corriger pendant la phase de segmentation l'erreur produite lors de l'initialisation si cette dernière est trop importante. C'est pourquoi un modèle d'évolution temporelle peut être envisagé, lorsque

le mouvement n'est pas uniforme, comme c'est le cas dans le cadre du problème que nous nous sommes fixé.

Le filtre de Kalman représente une possibilité intéressante puisqu'il offre l'avantage de tenir compte de toutes les mesures et de toutes les erreurs de mesure précédentes afin de fournir une bonne prédiction du mouvement. Nous allons rappeler son principe.

Le filtre de Kalman

Le filtre de Kalman, introduit par Kalman [KAL60], présenté et discuté dans [WEL97, SOR70] représente une solution récursive au problème du filtrage linéaire de données discrètes. Ses applications sont nombreuses, notamment dans le domaine de la navigation autonome ou assistée.

Son principe est d'estimer de façon optimale un vecteur d'état Ψ d'un système gouverné par une équation de la forme :

$$\Psi_{t+1} = A_t \Psi_t + \omega_t \quad (2.3)$$

à partir d'un vecteur de mesure bruitée s :

$$s_t = H_t \Psi_t + v_t \quad (2.4)$$

ω et v sont deux variables aléatoires représentant respectivement le bruit du système et le bruit des mesures. On fait l'hypothèse que ces variables sont indépendantes et suivent une loi normale de moyenne nulle :

$$P(\omega) \sim N(0, Q) \quad (2.5)$$

$$P(v) \sim N(0, R) \quad (2.6)$$

La matrice A_t représente la matrice de transition d'état de l'instant t à l'instant $t + 1$.

Enfin, la matrice H_t représente la matrice d'observation du vecteur de mesure s_t .

Equations de filtrage et de prédiction

Le problème se résume donc à déterminer le meilleur estimateur de Ψ_t et de

Ψ_{t+1} , connaissant le vecteur de mesure s_t et tous les précédents.

Soient :

- $\Psi_{t/t}$ l'estimateur a posteriori de Ψ_t
- $P_{t/t}$ la matrice de covariance de l'erreur d'estimation de Ψ_t .
- $\Psi_{t+1/t}$ l'estimateur a priori de Ψ_{t+1}
- $P_{t+1/t}$ la matrice de covariance de l'erreur d'estimation de Ψ_{t+1}
- K_t le gain de Kalman pondérant la différence entre la valeur mesurée et la valeur estimée
- Q la matrice de covariance du bruit du modèle
- R la matrice de covariance du bruit de mesure.

L'estimateur a posteriori et la matrice de covariance de l'erreur d'estimation de Ψ_t sont donnés par :

$$\Psi_{t/t} = E[\Psi_t/s_1 \dots s_t] \quad (2.7)$$

$$P_{t/t} = E[(\Psi_t - \Psi_{t/t})(\Psi_t - \Psi_{t/t})^T/s_1 \dots s_t] \quad (2.8)$$

où E représente l'espérance mathématique.

De même, l'estimateur a priori et la matrice de covariance de l'erreur d'estimation de Ψ_t et de Ψ_{t+1} sont donnés par :

$$\Psi_{t+1/t} = E[\Psi_{t+1}/s_1 \dots s_t] \quad (2.9)$$

$$P_{t+1/t} = E[(\Psi_{t+1} - \Psi_{t+1/t})(\Psi_{t+1} - \Psi_{t+1/t})^T/s_1 \dots s_t] \quad (2.10)$$

En dérivant les équations précédentes, on obtient les équations de filtrage et de prédiction.

Les équations de **filtrage** sont alors données par :

$$\begin{aligned} K_t &= P_{t/t-1}H^T(R + HP_{t/t-1}H^T)^{-1} \\ \Psi_{t/t} &= \Psi_{t/t-1} + K_t(s_t - H\Psi_{t/t-1}) \\ P_{t/t} &= (I - K_tH)P_{t/t-1} \end{aligned} \quad (2.11)$$

Les équations de **prédiction** sont :

$$\begin{aligned}\Psi_{t+1/t} &= A_t \Psi_{t/t} \\ P_{t+1/t} &= A_t P_{t/t} A_t^T + Q\end{aligned}\tag{2.12}$$

Application du filtre de Kalman à l'estimation des paramètres de mouvement

Ce sont les paramètres de mouvement qui vont être l'objet du filtrage de Kalman. Ainsi, le vecteur d'état du système sera composé du descripteur de mouvement Θ , ainsi que des dérivées temporelles du premier ordre des paramètres de mouvement.

En effet, pour la prédiction, nous considérons le mouvement comme non accéléré afin de simplifier la mesure et le calcul des paramètres. De cette façon, si δ_t représente un des paramètres de mouvement à l'instant t , $\dot{\delta}_t$ sa dérivée temporelle et $\ddot{\delta}_t$ sa dérivée seconde, nous disposons des relations suivantes :

$$\ddot{\delta}_t^{predict} = 0\tag{2.13}$$

$$\dot{\delta}_t^{predict} = \dot{\delta}_t\tag{2.14}$$

$$\delta_t^{predict} = \delta_t + T \cdot \dot{\delta}_t\tag{2.15}$$

où T représente la période d'acquisition des images.

Le vecteur d'état du système a donc pour expression :

$$\Psi_t = (t_{x_t}, t_{y_t}, a_t, b_t, c_t, d_t, \dot{t}_{x_t}, \dot{t}_{y_t}, \dot{a}_t, \dot{b}_t, \dot{c}_t, \dot{d}_t)^T\tag{2.16}$$

Nous considérons que seuls les paramètres de mouvement sont mesurables (les dérivées temporelles ne l'étant pas). Le vecteur d'observation a donc pour expression :

$$s_t = (t_{x_t}, t_{y_t}, a_t, b_t, c_t, d_t)^T\tag{2.17}$$

En considérant l'expression du vecteur d'état Ψ (2.16) et l'équation (2.15), il est alors possible de déterminer la matrice de transition d'état A utilisée pour la prédiction (2.12). Cette dernière est constante au cours du temps et a pour expres-

sion :

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & T & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & T & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & T & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & T & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & T & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & T \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (2.18)$$

Initialisation de la matrice d'observation H

Afin de permettre la sélection des seuls paramètres de mouvement à partir du vecteur d'état, leur dérivée temporelle étant considérée comme non observable, la matrice d'observation H est construite de manière à attribuer un poids de 1 aux paramètres de mouvement, et un poids de 0 à leur dérivée temporelle, les paramètres de mouvement étant déterminés à partir du vecteur d'état. De plus, la matrice d'observation H est également constante au cours du temps :

$$H = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad (2.19)$$

Initialisation des matrices de bruit Q et R

Les matrices Q et R représentent respectivement les matrices de covariance du bruit du système ω et des mesures v . De leur initialisation dépend le comportement du filtre. Il est souvent nécessaire de procéder par tâtonnement afin de déterminer des valeurs correctes.

Les valeurs que nous avons choisies sont présentées dans la partie expérimentale au paragraphe 2.4.2

Algorithme de prédiction par filtrage de Kalman

* INITIALISATION *

(1) On dispose de $\Psi_{t/t-1}$, de $P_{t/t-1}$ et de la mesure s_t

Tant que (non fin)

* PHASE DE FILTRAGE *

(2) Calcul du gain de Kalman :

$$K_t \leftarrow P_{t/t-1} H_t^T (R + H_t P_{t/t-1} H_t^T)^{-1}$$

(3) Calcul de l'estimateur a posteriori :

$$\Psi_{t/t} \leftarrow \Psi_{t/t-1} + K_t (s_t - H_t \Psi_{t/t-1})$$

(4) Calcul de sa matrice de covariance :

$$P_{t/t} \leftarrow (I - K_t H_t) P_{t/t-1}$$

* PHASE DE PREDICTION *

(5) Calcul de l'estimateur a priori :

$$\Psi_{t+1/t} \leftarrow A_t \Psi_{t/t}$$

(6) Calcul de sa matrice de covariance :

$$P_{t+1/t} \leftarrow A_t P_{t/t} A_t^T + Q$$

* PHASE DE MISE A JOUR *

(7) $\Psi_{t/t-1} \leftarrow \Psi_{t+1/t}$

(8) $P_{t/t-1} \leftarrow P_{t+1/t}$

finTq

La condition de sortie de la boucle est validée lorsque l'ensemble de la séquence d'images a été traité.

2.3.3 Optimisation de la région prédite par un modèle de région déformable

La phase de prédiction réalisée, nous disposons des paramètres de mouvement prédits. On peut alors obtenir un contour prédit de la région. Cependant, cette prédiction n'étant généralement pas parfaite, la région obtenue ne correspond pas toujours exactement à la région réelle. C'est pourquoi nous utilisons un modèle de région déformable inspiré de celui de [BAS94b], de manière à optimiser la position et la forme de la région. Afin d'assurer la compatibilité du modèle de région avec le modèle de contours actifs à minimisation locale qui sera présenté au paragraphe 2.3.4, nous utilisons ici un contour discret alors que l'article présente une

étude reposant sur un modèle de contour continu associé à une fonctionnelle d'énergie minimisée par une approche variationnelle.

Présentation du modèle de région déformable

Le modèle de région déformable que nous utilisons pour le suivi est défini par une courbe représentant le contour, et par une énergie.

L'optimisation de la région déformable est réalisée de façon itérative par minimisation de l'énergie mesurant la similitude des niveaux de gris entre les points de la nouvelle région et ceux de la région extraite de l'image précédente.

L'utilisation de cette information, fournie par l'ensemble de la région, améliore la fiabilité du suivi en cas de grands déplacements de l'objet, ou d'images bruitées.

La déformation de la région est entièrement contrainte par le modèle de mouvement utilisé (modèle affine dans notre cas). Ainsi, les paramètres de la région déformable qui sont optimisés, sont les paramètres de mouvement de la région.

Définition du contour de la région déformable

Le contour de la région est représenté par un ensemble ordonné de n points :

$$\omega = \{w_i = (x_i, y_i)/i \in [1, n]\} \quad (2.20)$$

Les points internes de la région sont alors obtenus par application d'un algorithme de remplissage après une interpolation linéaire du contour.

Définition de l'énergie de la région déformable

La définition de l'énergie de la région déformable repose sur la notion de *DFD* pour *Displaced Frame Difference*. Nous ferons l'hypothèse que l'intensité des points est invariante au cours du déplacement d'une image à l'autre. Cette énergie est définie par :

$$E = \frac{1}{S_{R_t}} \sum_{p \in R_t} DFD^2(p, \vec{d}_t^+(p)) \quad (2.21)$$

$$E = \frac{1}{S_{R_t}} \sum_{p \in R_t} [I_{t+1}(p + \vec{d}_t^+(p)) - I_t(p)]^2 \quad (2.22)$$

avec :

R_t la région extraite de l'image précédente à l'instant t ,
 S_{R_t} l'aire de la région R_t exprimée en nombre de pixels,
 p un point de l'image.

Cette énergie ne dépend que de $\vec{d}_t^+(p)$ et donc de Θ_t^+ .

Ainsi, minimiser l'énergie de la région déformable revient à déterminer le descripteur de mouvement $\hat{\Theta}_t^+$ pour lequel l'énergie est minimale :

$$\hat{\Theta}_t^+ = \arg \min_{\Theta_t^+} E$$

Nous avons utilisé deux méthodes de minimisation de l'énergie. La première, inspirée de celle de [NIC92, NIC93] s'appuie sur la méthode du gradient. Elle offre l'avantage d'être rapide mais elle est très sensible à l'initialisation et fournit très souvent un minimum local non optimum. C'est la raison pour laquelle nous nous sommes tourné vers la méthode du recuit simulé présentée dans [COR87] qui détermine plus souvent le minimum global. Son principal inconvénient est d'être très coûteuse en terme d'évaluations de fonction. Ainsi, suivant les besoins du suivi (rapidité ou précision), nous utiliserons l'une ou l'autre de ces méthodes. Ces dernières vont maintenant être décrites plus en détail.

Minimisation de l'énergie de la région par la méthode du gradient

La méthode du gradient est souvent utilisée lorsque l'on veut minimiser une fonction. Son principe est de s'approcher itérativement du minimum de la fonction en se déplaçant dans la direction du gradient de celle-ci, selon la formule :

$$\vec{d}^{i+1} = \vec{d}^i - (\lambda_i \epsilon \vec{G}^i)$$

avec

$$\vec{G}^i = \frac{1}{2} \begin{pmatrix} \frac{\partial DFD^2}{\partial x}(x, y, \vec{d}^i) \\ \frac{\partial DFD^2}{\partial y}(x, y, \vec{d}^i) \end{pmatrix}$$

où :

i représente la i^{ieme} itération, \vec{G}^i le gradient à l'itération i , λ_i la longueur de déplacement et ϵ la matrice de gain diagonale de taille 2x2.

La méthode du gradient peut être adaptée afin d'être appliquée au descripteur

de mouvement qui contient les variables de la fonctionnelle d'énergie à minimiser.

L'équation, dans le cas du modèle de mouvement affine, devient la suivante :

$$\Theta_R^{i+1} = \Theta_R^i - \lambda_i \epsilon \vec{G}^i$$

avec

$$\vec{G}^i = \frac{1}{S_R} \sum_{p \in R} \begin{pmatrix} \frac{\partial DFD^2}{\partial t_x}(p, \vec{d}^i) \\ \frac{\partial DFD^2}{\partial t_y}(p, \vec{d}^i) \\ \frac{\partial DFD^2}{\partial a}(p, \vec{d}^i) \\ \frac{\partial DFD^2}{\partial b}(p, \vec{d}^i) \\ \frac{\partial DFD^2}{\partial c}(p, \vec{d}^i) \\ \frac{\partial DFD^2}{\partial d}(p, \vec{d}^i) \end{pmatrix}$$

où :

Θ_R^i représente le descripteur de mouvement à l'itération i ,
 S_R l'aire de la region R .

Le gradient \vec{G}^i est alors appelé *terme correctif*.

A chaque itération, deux opérations doivent être réalisées :

1. Calculer le terme correctif \vec{G}^i
2. Déterminer le λ_i optimal pour lequel l'énergie est minimale dans la direction du gradient.

Ces deux étapes vont maintenant être présentées plus en détail.

Expression détaillée du gradient

Afin de simplifier l'expression du gradient, on utilise la relation suivante :

$$\frac{\partial DFD^2}{\partial k}(p, \vec{d}^i) = 2.DFD(p, \vec{d}^i). \frac{\partial DFD}{\partial k}(p, \vec{d}^i)$$

pour une variable k quelconque.

Or,

$$DFD(p, \vec{d}^i) = I_{t+1}(p + \vec{d}_t^+(p)) - I_t(p)$$

et

$$\vec{d}_t^+(p) = \begin{pmatrix} d_x \\ d_y \end{pmatrix} = \begin{pmatrix} t_x + a(x - x_G) + c(y - y_G) \\ t_y + b(x - x_G) + d(y - y_G) \end{pmatrix}$$

De plus, on a la relation suivante :

$$\frac{\partial I}{\partial k}(x, y) = \frac{\partial I}{\partial x}(x, y) \frac{\partial x}{\partial k}(k) + \frac{\partial I}{\partial y}(x, y) \frac{\partial y}{\partial k}(k)$$

d'où finalement :

$$\left\{ \begin{array}{l} \frac{\partial DFD}{\partial t_x}(p, \vec{d}^i) = \frac{\partial I_{t+1}}{\partial x}(p + \vec{d}^i) \\ \frac{\partial DFD}{\partial t_y}(p, \vec{d}^i) = \frac{\partial I_{t+1}}{\partial y}(p + \vec{d}^i) \\ \frac{\partial DFD}{\partial a}(p, \vec{d}^i) = (x - x_G) \frac{\partial I_{t+1}}{\partial x}(p + \vec{d}^i) \\ \frac{\partial DFD}{\partial b}(p, \vec{d}^i) = (x - x_G) \frac{\partial I_{t+1}}{\partial y}(p + \vec{d}^i) \\ \frac{\partial DFD}{\partial c}(p, \vec{d}^i) = (y - y_G) \frac{\partial I_{t+1}}{\partial x}(p + \vec{d}^i) \\ \frac{\partial DFD}{\partial d}(p, \vec{d}^i) = (y - y_G) \frac{\partial I_{t+1}}{\partial y}(p + \vec{d}^i) \end{array} \right.$$

avec :

$\frac{\partial I_{t+1}}{\partial x}$ la dérivée partielle de l'image I_{t+1} suivant x et $\frac{\partial I_{t+1}}{\partial y}$ la dérivée partielle de l'image I_{t+1} suivant y .

On obtient ainsi la relation itérative :

$$\Theta_R^{i+1} = \Theta_R^i - (\lambda_i \in \frac{2}{S_R} \sum_{p \in R} DFD(p, \vec{d}^i) \cdot \vec{C}^i)$$

avec :

$$\vec{C}^i = \begin{pmatrix} \frac{\partial I_{t+1}}{\partial x}(p + \vec{d}^i) \\ \frac{\partial I_{t+1}}{\partial y}(p + \vec{d}^i) \\ (x - x_G) \frac{\partial I_{t+1}}{\partial x}(p + \vec{d}^i) \\ (x - x_G) \frac{\partial I_{t+1}}{\partial y}(p + \vec{d}^i) \\ (y - y_G) \frac{\partial I_{t+1}}{\partial x}(p + \vec{d}^i) \\ (y - y_G) \frac{\partial I_{t+1}}{\partial y}(p + \vec{d}^i) \end{pmatrix}$$

Les paramètres de mouvement n'étant généralement pas du même ordre de grandeur, la matrice de gain ϵ doit donc être construite en fonction du problème. [NIC92]

propose dans le cas de paramètres de mouvement affine la matrice de gain suivante :

$$\epsilon = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0,01 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,01 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0,01 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0,01 \end{pmatrix}$$

Détermination du λ optimal

Une fois la direction du gradient calculée, il est nécessaire de déterminer la longueur de déplacement λ dans cette direction, afin de minimiser la fonction objectif.

L'algorithme du gradient proposé par [NIC92, NIC93] utilise une valeur initiale de λ et la divise par 2 à chaque itération, si le point obtenu ne minimise pas la fonction. En effet, au fur et à mesure des itérations, on se rapproche, en principe, du point solution, et donc le déplacement devrait être plus petit.

L'expérience a montré que sur nos séquences, le point initial est très souvent conservé comme solution, la diminution de λ étant trop brutale.

Nous avons donc utilisé une méthode de minimisation à une dimension, afin de déterminer le λ optimal. Plusieurs méthodes sont disponibles dans la littérature. Nous nous sommes orienté vers celle décrite dans [DAH74], qui s'appuie sur l'utilisation d'un polynôme d'interpolation quadratique.

Algorithme du gradient

(1) $i \leftarrow 0$

(2) *continue* $\leftarrow 1$

Tant que ($i < \text{nombre maximal d'itérations}$ et *continue* = 1)

(3) Calcul de \vec{G}^i

(4) Calcul de λ_i

(5) Calcul de Θ_{estim}^{i+1}

Si ($DFD(p, \vec{d}_{estim}^{i+1}) < DFD(p, \vec{d}^i)$) **alors**

(6) $\Theta^{i+1} \leftarrow \Theta_{estim}^{i+1}$

(7) $i \leftarrow i + 1$

Sinon

(8) *continue* $\leftarrow 0$

finSi
finTq

Minimisation de l'énergie de la région par la méthode du recuit simulé

La méthode du recuit simulé, présentée dans [COR87], est une technique de minimisation globale de fonctions à plusieurs variables. Son utilisation s'est, dans un premier temps, portée sur la résolution de problèmes combinatoires tels que le *voyageur de commerce*, ou encore la réalisation de circuits imprimés complexes, pour lesquels la fonction objectif est définie dans un domaine discret. Les utilisations de la méthode se sont ensuite étendues à l'optimisation de fonctions à variables continues.

L'avantage de la technique du recuit simulé est, qu'à la différence des méthodes traditionnelles, elle ne fournit pas le premier minimum local rencontré. Au contraire, par une recherche aléatoire itérative, avec un déplacement adaptatif le long de chaque direction des coordonnées de la fonction objectif, elle permet de trouver, sinon le minimum global, du moins un minimum local de meilleure qualité.

La méthode du recuit simulé doit son nom à une analogie avec la thermodynamique, et plus précisément avec le processus de refroidissement et de cristallisation de certains métaux. En effet, un métal change d'état lorsque l'on diminue son énergie en le refroidissant. Si l'on diminue sa température lentement et étape par étape, le métal passe naturellement à un état cristallin stable d'énergie minimale. Si, au contraire, le refroidissement est brutal, le métal n'atteint pas cet état, mais un état métastable polycristallin ou amorphe d'énergie plus élevée et donc non minimale.

L'analogie avec les méthodes d'optimisation se fait de la façon suivante : une méthode de recherche itérative acceptant uniquement les points pour lesquels la fonction objectif est la plus faible (la température du système est diminuée rapidement), va rapidement atteindre un état métastable, en un minimum local. A l'opposé, les méthodes de recuit simulé acceptent sous certaines conditions, de considérer des points non minimaux afin de s'approcher du minimum global.

L'algorithme se déroule de la façon suivante : la température est initialement élevée. Une séquence de points est alors générée jusqu'à ce que s'établisse un équilibre, c'est-à-dire, jusqu'à ce que la valeur moyenne de la fonction objectif des nouveaux points générés atteigne une valeur stable au fur et à mesure des itérations. Après cet équilibre "thermal", la température est réduite, et de nouvelles séquences de points sont générées jusqu'à ce qu'un nouvel équilibre soit obtenu, et ainsi de

suite. Le processus est arrêté lorsque la température est suffisamment basse pour qu'aucune amélioration ne puisse plus être envisageable.

2.3.4 Ajustement du contour de la région par contours actifs

L'étape précédente de l'algorithme de suivi nous a permis d'optimiser la région prédite par le filtre de Kalman. Cependant les modèles de régions déformables utilisés pour le suivi de régions, bien qu'assurant une fiabilité importante du suivi, ont des difficultés à déterminer avec précision les contours de la région. Aussi, il est intéressant d'associer au modèle de région, un modèle de contour déformable qui permettra d'ajuster le contour de la région aux contours réels de l'objet dans l'image traitée I_{t+1} . Le modèle de contour que nous avons choisi est le modèle des *contours actifs* ou *snakes*.

Présentation des contours actifs

Le modèle des contours actifs, ou snakes, présenté initialement par [KAS98] et appliqué au suivi temporel par [BAS94a, BON94, BON98] est défini à l'aide d'une courbe et d'une fonctionnelle d'énergie. La courbe est déformée de manière itérative afin de minimiser la fonctionnelle d'énergie, de manière à s'ajuster sur les contours de l'image.

Considérons une courbe paramétrée v pouvant être définie par (2.23). Nous confondrons, afin de simplifier son écriture, la courbe et sa représentation paramétrique.

$$\begin{aligned} v : [a, b] \subset \mathbb{R} &\rightarrow \mathbb{R}^2 \\ s &\mapsto v(s) = \begin{bmatrix} x(s) \\ y(s) \end{bmatrix} \end{aligned} \quad (2.23)$$

où s désigne l'abscisse curviligne le long de la courbe, a et b les extrémités de la courbe ($a < b$), et $v(s)$ le point courant de coordonnées $(x(s), y(s))$.

L'approche classique des snakes [KAS98] associe la courbe V à la réalisation du minimum de l'énergie définie par l'équation (2.24).

$$E(v) = \alpha \int_a^b |v'(s)|^2 ds + \beta \int_a^b |v''(s)|^2 ds - \lambda \int_a^b |\nabla I(v(s))| ds \quad (2.24)$$

où α , β et λ sont trois constantes positives et I est une image qui contient un objet à détecter.

Les deux premiers termes de (2.24) contrôlent la régularité du contour, tandis que le troisième l'attire sur les bords de l'objet.

Kass et al [KAS98] ont proposé une méthode de minimisation basée sur les techniques du calcul variationnel. Amini et al [AMI90] ont mis en évidence son caractère instable, et ont élaboré un algorithme fondé sur la programmation dynamique. Bien que plus efficace, cette méthode se révèle être très lente.

Nous avons utilisé le modèle proposé par [WIL92] appelé *greedy algorithm* qui a l'avantage d'être stable et plus rapide, ce qui en fait l'algorithme le plus souvent employé [PAP95].

Ce modèle consiste en la représentation d'un contour par un ensemble ordonné discret de n points :

$$\omega = \{w_i = (x_i, y_i) / i \in [1, n]\} \quad (2.25)$$

Ces points vont s'approcher du contour réel de l'objet de façon itérative par minimisation locale d'une fonction d'énergie.

Pour chaque point ω'_i dans un voisinage réduit (3x3, 5x5 ou 7x7) de ω_i , le terme d'énergie suivant est calculé :

$$\tilde{E}(\omega'_i) = E_{int}(\omega'_i) + E_{ext}(\omega'_i) + E_{ctxt}(\omega'_i)$$

où E_{int} est le terme d'énergie interne au snake, lui imposant une certaine régularité, E_{ext} est le terme d'énergie lié à l'image globale et E_{ctxt} est l'énergie de contexte liée aux propriétés du contour.

Chaque point ω_i est alors déplacé vers le point de son voisinage correspondant à la valeur minimale de \tilde{E} .

Les énergies choisies

L'efficacité du modèle dépend des termes d'énergie considérés, qui doivent être adaptés aux images traitées, et au type d'objet à segmenter.

Nous avons choisi d'utiliser cinq termes d'énergie.

L'énergie interne sera composée de deux termes : l'énergie de continuité et

l'énergie de courbure, selon la formule :

$$E_{int} = aE_{cont} + bE_{courb}$$

L'énergie externe, quant à elle, est composée de deux termes : l'énergie de gradient et l'énergie d'intensité selon la formule :

$$E_{ext} = cE_{grad} + dE_{ndg}$$

Enfin, l'énergie de contexte correspond par exemple à une énergie ballon :

$$E_{ctxt} = eE_{ballon}$$

Les coefficients de pondération a , b , c , d et e permettent de régler l'importance relative des différents termes d'énergie.

Energie de continuité

L'énergie de continuité E_{cont} a pour rôle d'encourager l'équidistance entre les points du contour. Elle a pour expression :

$$E_{cont}(\omega'_i) = abs(\bar{d}^2 - |\omega'_i - \omega_{i-1}|^2) \quad (2.26)$$

$|\omega'_i - \omega_{i-1}|$ représente la distance entre le point ω'_i correspondant au voisin en cours d'exploration du point ω_i , et le point ω_{i-1} qui est le prédécesseur de ω_i dans la courbe.

\bar{d} représente la distance moyenne entre les points du contour.

Energie de courbure

L'énergie de courbure E_{courb} a pour rôle d'éviter la formation d'angles aigus formés par les points consécutifs du contour. Elle correspond à l'approximation discrète du carré de la dérivée seconde de la fonction associée à la courbe au point ω_i , et a pour expression :

$$E_{courb}(\omega'_i) = |\omega_{i-1} - 2\omega'_i + \omega_{i+1}|^2 \quad (2.27)$$

Energie de gradient

L'énergie de gradient a pour rôle d'attirer le contour vers les frontières des objets de l'image, c'est-à-dire vers les zones de fort gradient spatial.

Son expression est la suivante :

$$E_{grad}(\omega'_i) = |\nabla I(\omega'_i)| \quad (2.28)$$

où $|\nabla I(\omega'_i)|$ correspond à la norme du gradient spatial de l'image I au point ω'_i .

Energie d'intensité

L'énergie d'intensité a pour rôle d'attirer le contour vers les zones de forte intensité de niveau de gris (ou faible suivant le signe du coefficient de pondération).

Son expression est la suivante :

$$E_{ndg}(\omega'_i) = I(\omega'_i) \quad (2.29)$$

où $I(\omega'_i)$ correspond au niveau de gris de l'image I au point ω'_i .

Energie ballon

L'énergie ballon, introduite dans [COH91], encourage le contour à s'étendre ou à se contracter suivant le signe du coefficient de pondération. Cette énergie est utile car, en l'absence de contrainte extérieure, le contour soumis aux seules forces de continuité et de courbure, a tendance à se contracter. L'utilisation de l'énergie ballon permet l'initialisation du contour à l'intérieur de la région d'intérêt, car le contour va alors s'étendre jusqu'aux frontières de l'objet.

L'énergie ballon pousse en général les points du contour à se déplacer dans la direction de la normale extérieure à la courbe en ces points. Cela ne convient pas à notre cas puisqu'il est nécessaire, notamment lors de la segmentation initiale, de permettre aux points de glisser le long des frontières de l'objet (le bolus), et donc de se déplacer dans une direction différente de celle de la normale. C'est pourquoi nous avons dû introduire une nouvelle expression de l'énergie basée sur la distance des points au centre de gravité.

L'énergie ballon a alors pour expression :

$$E_{ballon}(\omega'_i) = |\omega'_i - g|^2 \quad (2.30)$$

où g représente le centre de gravité du contour.

Normalisation des énergies

Il est nécessaire de normaliser les énergies précédentes afin qu'elles aient des valeurs du même ordre de grandeur. Leurs valeurs seront donc réajustées afin d'appartenir aux intervalles $[0, 1]$ ou $[-1, 0]$.

Il convient de distinguer deux manières de normaliser les énergies.

La première méthode consiste à normaliser la valeur de l'énergie au point ω'_i par la valeur maximale dans le voisinage. C'est le cas pour les énergies de continuité E_{cont} et de courbure E_{courb} .

Cependant, lorsque dans un voisinage la différence entre les valeurs d'énergie prises en deux points distincts, bien que significative à nos yeux, est très faible en comparaison de la valeur même de l'une ou de l'autre, il est judicieux de réaliser un rapport de différence. C'est le principe de la deuxième méthode qui sera utilisée pour les énergies ballon E_{ballon} , de gradient E_{grad} et d'intensité E_{ndg} .

L'expression des trois énergies précédentes normalisées est donc de la forme :

$$E'(\omega'_i) = \frac{E - \text{valmin}(\omega'_i)}{\text{valmax} - \text{valmin}} \quad (2.31)$$

Les valeurs valmin et valmax correspondent aux valeurs respectivement minimale et maximale de l'énergie considérée dans le voisinage du point ω_i .

Algorithme des contours actifs à minimisation locale

Nous donnons ici les grandes lignes de l'algorithme des contours actifs.

- (1) Initialisation des points du contour
 - (2) Initialisation des coefficients de pondération des énergies
- Tant que** ($nbPointsDeplaces > critereMinimum$)

Pour ($i \leftarrow 1$ à n)

(3) $E_{min} \leftarrow \infty$

Pour ($j \leftarrow 1$ à m)

* m représente la taille du voisinage*

(4) Calcul de E_j

Si ($E_j < E_{min}$) **alors**

(5) $E_{min} \leftarrow E_j$

```

        (6)  $j_{min} \leftarrow j$ 
    finSi
  finPour
  Si ( $j_{min}$  n'est pas le point courant) alors
    (7) Déplacer le point  $\omega_i$  vers  $j_{min}$ 
    (8)  $nbPointsDeplaces \leftarrow nbPointsDeplaces + 1$ 
  finSi
finPour
finTq

```

2.3.5 Rééchantillonnage des points du contour

Objectifs du rééchantillonnage

Le modèle de régions déformables défini dans la partie 2.3.3 ainsi que le modèle de contours actifs défini dans la partie 2.3.4 s'appuient sur la représentation du contour de la région par un ensemble ordonné de n points ω . Or, cet ensemble de points est le même pour les deux modèles. Il est important que ces points soient répartis de façon homogène autour de la région. Cependant, après la phase d'ajustement du contour par l'algorithme des contours actifs, les points, soumis à différentes forces, ne conservent pas nécessairement leur équidistance. De plus, l'objet suivi est un objet déformable dont les dimensions varient au cours du temps. Le nombre de points du contour ω doit donc pouvoir varier.

C'est pourquoi la phase d'ajustement du contour par contours actifs est suivie d'une phase de rééchantillonnage des points du contour ω .

Principe du rééchantillonnage

Le rééchantillonnage du contour est réalisé de la façon suivante : une interpolation linéaire entre les points de ω permet de déterminer l'ensemble des pixels de l'image appartenant au contour de la région. Ces derniers forment une boucle fermée de largeur 1 pixel. Celle-ci est alors parcourue dans le sens horaire, à partir du point le plus haut dans l'image. Les points retenus, qui vont composer le nouveau contour ω' , sont séparés d'un certain nombre de pixels (4 par exemple), correspondant au pas d'échantillonnage. Ce dernier doit être suffisamment faible pour assurer une bonne approximation du contour réel, tout en étant assez élevé afin d'éviter une perturba-

tion du comportement de l'algorithme des contours actifs pouvant apparaître lorsque le nombre de points du contour présent dans le voisinage considéré est trop élevé. Le sens de parcours de la boucle donne l'ordre des points de ω' .

2.3.6 Estimation des paramètres de mouvement

A ce stade de l'algorithme, la région suivie est supposée parfaitement segmentée dans la nouvelle image I_{t+1} . Il reste alors à mesurer les paramètres de mouvement de la région de I_t à I_{t+1} afin d'alimenter le filtre de Kalman, pour que ce dernier puisse fournir une nouvelle prédiction. C'est l'estimation du mouvement de la région qui est l'objet de cette partie.

Principe de l'estimation

Afin d'estimer les paramètres de mouvement de la région suivie, nous utilisons un algorithme dont le principe est de minimiser l'erreur quadratique moyenne de reconstruction EQM de la région.

Cette erreur repose sur la notion de DFD présentée dans la section 2.3.3. Cependant, la DFD aura ici un rôle inverse par rapport au sens de l'axe temporel considéré.

En effet, dans la section 2.3.3, nous connaissions la position de la région R à l'instant t (R_t), et nous cherchions sa position à l'instant $t + 1$ (R_{t+1}). Nous considérons donc l'axe temporel dans le sens croissant : $t \rightarrow t + 1$.

La situation est différente à présent : le suivi nous a permis d'obtenir la position de la région à $t + 1$ (R_{t+1}) et nous avons pour objectif de déterminer le mouvement qu'elle a eu depuis sa position à l'instant t (R_t). Nous considérons donc l'axe temporel dans le sens décroissant : $t + 1 \rightarrow t$. L'expression de la DFD est alors la suivante :

$$DFD(p, \vec{d}_{t+1}^-(p)) = I_t(p + \vec{d}_{t+1}^-(p)) - I_{t+1}(p) \quad (2.32)$$

L'erreur quadratique moyenne de reconstruction a donc pour expression :

$$EQM_{R_{t+1}} = \frac{1}{S_{R_{t+1}}} \sum_{p \in R_{t+1}} DFD^2(p, \vec{d}_{t+1}^-(p)) \quad (2.33)$$

$$EQM_{R_{t+1}} = \frac{1}{S_{R_{t+1}}} \sum_{p \in R_{t+1}} [I_t(p + \vec{d}_{t+1}^-(p)) - I_{t+1}(p)]^2 \quad (2.34)$$

où $S_{R_{t+1}}$ représente l'aire de la région R_{t+1} .

Cette erreur ne dépend que de $\vec{d}_{t+1}^-(p)$ et donc de $\Theta_{R,t+1}^-$.

Ainsi, minimiser l'erreur quadratique moyenne de reconstruction de la région revient à déterminer le descripteur de mouvement $\hat{\Theta}_{R,t+1}^-$ pour lequel l'erreur est minimale :

$$\hat{\Theta}_{R,t+1}^- = \arg \min_{\Theta_{R,t+1}^-} EQM_{R,t+1}$$

La minimisation de l' EQM est réalisée de la même façon que dans la partie 2.3.3. En effet, il est possible d'utiliser soit la méthode du gradient, soit la méthode du recuit simulé, selon les besoins du suivi.

Dans le cas de la méthode du gradient, l'expression de la DFD ayant été modifiée par rapport à la partie 2.3.3, la relation itérative utilisée est la suivante :

$$\Theta_R^{i+1} = \Theta_R^i - \lambda_i \quad \epsilon \quad \frac{2}{S_R} \sum_{p \in R} DFD(p, \vec{d}^i) \cdot \vec{C}^i$$

avec :

$$\vec{C}^i = \begin{pmatrix} \frac{\partial I_t}{\partial x}(p + \vec{d}^i) \\ \frac{\partial I_t}{\partial y}(p + \vec{d}^i) \\ (x - x_G) \frac{\partial I_t}{\partial x}(p + \vec{d}^i) \\ (x - x_G) \frac{\partial I_t}{\partial y}(p + \vec{d}^i) \\ (y - y_G) \frac{\partial I_t}{\partial x}(p + \vec{d}^i) \\ (y - y_G) \frac{\partial I_t}{\partial y}(p + \vec{d}^i) \end{pmatrix}$$

Relation entre Θ_t^+ et Θ_{t+1}^-

Nous avons alors déterminé la mesure du mouvement Θ_{t+1}^- correspondant au sens décroissant de l'axe temporel. Or, le filtre de Kalman nécessite une mesure de

mouvement dans le sens croissant Θ_t^+ pour réaliser sa prédiction. Il faut donc définir une fonction F telle que : $\Theta_t^+ = F(\Theta_{t+1}^-)$.

Dans le cadre du modèle affine, le vecteur déplacement des points d'une région correspondant au descripteur de mouvement Θ_{t+1}^- est donné par :

$$\vec{d}_{t+1}^- \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} t_x \\ t_y \end{pmatrix} + \begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} x - x_G \\ y - y_G \end{pmatrix}$$

ce qui se traduit sous forme matricielle par :

$$\vec{d}_{t+1}^-(p) = T + M(p_{t+1} - g_{t+1})$$

Or $\vec{d}_{t+1}^-(p) = p_t - p_{t+1}$ et $\vec{d}_t^+(p) = p_{t+1} - p_t$ d'où : $\vec{d}_t^+(p) = -\vec{d}_{t+1}^-(p)$.

On aboutit ainsi à l'équation :

$$\vec{d}_t^+(p) = -T - M(\vec{d}_t^+(p) + p_t - g_{t+1})$$

Après calcul, on obtient la relation suivante :

$$\begin{aligned} \vec{d}_t^+(p) &= -(M + I)^{-1}T + [(M + I)^{-1} - I](p_t - g_{t+1}) \\ &= T' + M'(p_t - pg_{t+1}) \end{aligned}$$

Nous obtenons finalement l'expression des composantes de Θ_t^+ en fonction de celles de Θ_{t+1}^- :

$$\begin{aligned} T' &= -(M + I)^{-1}T \\ \begin{pmatrix} t'_x \\ t'_y \end{pmatrix} &= - \begin{pmatrix} a + 1 & c \\ b & d + 1 \end{pmatrix}^{-1} \begin{pmatrix} t_x \\ t_y \end{pmatrix} \end{aligned}$$

et

$$M' = (M + I)^{-1} - I$$

$$\begin{pmatrix} a' & c' \\ b' & d' \end{pmatrix} = \begin{pmatrix} a + 1 & c \\ b & d + 1 \end{pmatrix}^{-1} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

2.4 Résultats

Nous allons présenter dans cette partie l'application du suivi au problème médical décrit dans la section 1.2. L'objectif est de suivre un bolus baryté dans des séquences d'images radiologiques. Nous décrirons tout d'abord la méthode que nous avons développée afin d'initialiser de manière automatique l'algorithme de suivi. Les valeurs utilisées pour les paramètres seront ensuite indiquées. Enfin, nous présenterons les images illustrant la position du contour du bolus à différents instants, pour quelques séquences représentatives des séquences étudiées.

2.4.1 Initialisation de l'algorithme de suivi

Les séquences radiologiques dont nous disposons ont été présentées section 1.2. Elles durent une dizaine de secondes chacune, et sont cadencées à 25 images par seconde. Elles présentent le déplacement d'un bolus baryté dans l'œsophage et le SIO.

Les images se présentent sous forme rectangulaire de taille 1112x1024 pixels. Un cache se situe à la périphérie et entoure une fenêtre de visualisation au centre de l'image.

Les premières images des séquences sont très sombres, car elles correspondent à la phase de mise sous tension de l'appareil à rayons X. Les images suivantes sont de plus en plus claires. Au bout de quelques secondes, l'intensité lumineuse se stabilise. Apparaît alors le bolus en haut de l'image. Lorsque finalement celui-ci disparaît de l'image, l'acquisition se termine, ce qui se caractérise par un assombrissement progressif des images.

Le but de cette première étape du traitement est de déterminer la première image de la séquence où apparaît le bolus et de localiser ce dernier dans l'image.

Notre algorithme de détection repose sur deux hypothèses :

- le bolus apparaît par le haut de la fenêtre de visualisation, sur le bord du cache
- l'apparition du bolus entraîne un assombrissement local (une diminution de la valeur des niveaux de gris) de la zone de l'image concernée.

Afin de réduire le temps de calcul, nous utilisons la première hypothèse qui nous permet de réduire l'espace de recherche à une zone réduite de l'image appelée superbloc située au niveau de la partie supérieure de la fenêtre de visualisation. Son aire est environ égale au quart de celle de la fenêtre. Du fait de la variabilité de la position de la fenêtre de visualisation, d'une séquence à l'autre (figure 2.2), la position du superbloc doit être adaptative.

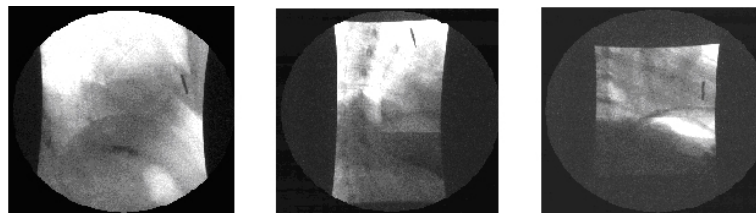


FIG. 2.2 – Différentes configurations de la fenêtre de visualisation

Pour déterminer cette zone (figure 2.3), nous avons besoin de connaissances sur les caractéristiques globales des conditions d'acquisition. Elles sont plus accessibles dans une zone temporelle pour laquelle le contraste des images est stable (au milieu de la séquence). L'image correspondante est binarisée en utilisant comme valeur seuil la valeur moyenne des niveaux de gris des pixels d'un bloc de petite taille (10*10 pixels) situé au centre de la moitié inférieure de l'image (zone sombre située au niveau de l'estomac). Une ouverture morphologique sur l'image binaire en utilisant un élément structurant carré de taille 6x6, par exemple, est ensuite opérée afin de diminuer le bruit.

La partie supérieure du superbloc est obtenue en considérant la position de la première zone blanche rencontrée sur plusieurs colonnes verticales situées au centre de l'image, en partant du haut. La hauteur du superbloc est fixée expérimentalement à 10 % de la hauteur totale de l'image. Sa largeur, c'est-à-dire le côté horizontal du superbloc, est délimitée à chaque extrémité du côté inférieur par le premier pixel noir rencontré (figure 2.3).

Le superbloc est ensuite divisé en blocs adjacents de même hauteur, et de largeur 10 pixels (environ la moitié de la largeur du bolus baryté) afin d'utiliser la deuxième hypothèse pour la détection.

Notre méthode repose sur l'étude des différences du niveau de gris moyen du

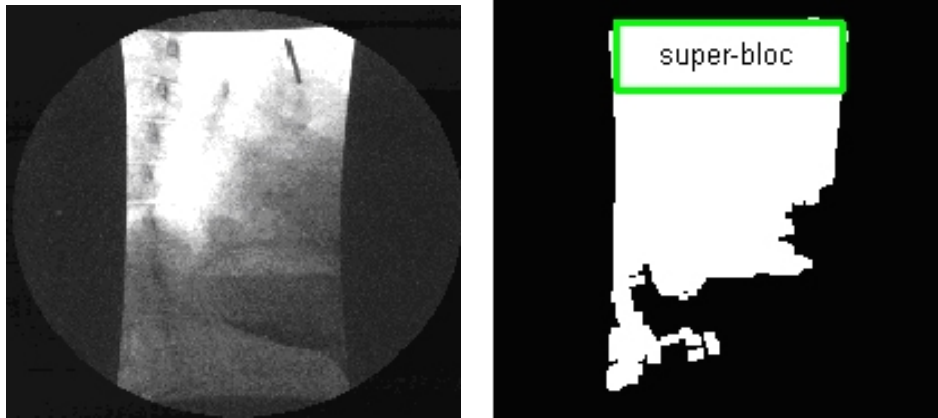
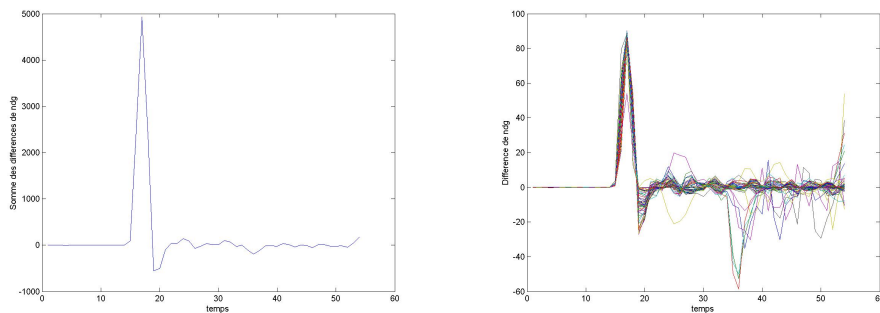


FIG. 2.3 – Construction du superbloc

superbloc entre deux images successives, ainsi que de chacune de ses zones. L'étude au niveau du superbloc permet de détecter le moment de mise en marche de la caméra, correspondant à une phase d'éclaircissement global de l'image, alors que l'étude au niveau des zones du superbloc permet de détecter l'arrivée du bolus baryté dans l'image correspondant à une phase d'assombrissement local.

L'évolution des différences du niveau de gris moyen du superbloc ainsi que les évolutions de chacune de ses zones sont présentées figure 2.4.



(a) Différences de niveau de gris du superbloc

(b) Différences de niveau de gris de chaque zone

FIG. 2.4 – Etude des différences des niveaux de gris des zones du superbloc

Le maximum des différences des niveaux de gris du superbloc correspond à l'éclaircissement général (passage du noir au blanc), l'image correspondante étant la première à partir de laquelle la recherche du bolus débute.

La zone dont la différence des niveaux de gris est négative et donc la plus faible

(passage du blanc au noir) est alors identifiée. Elle correspond à l'endroit dans l'image où le bolus apparaît. Le résultat est présenté figure 2.5.

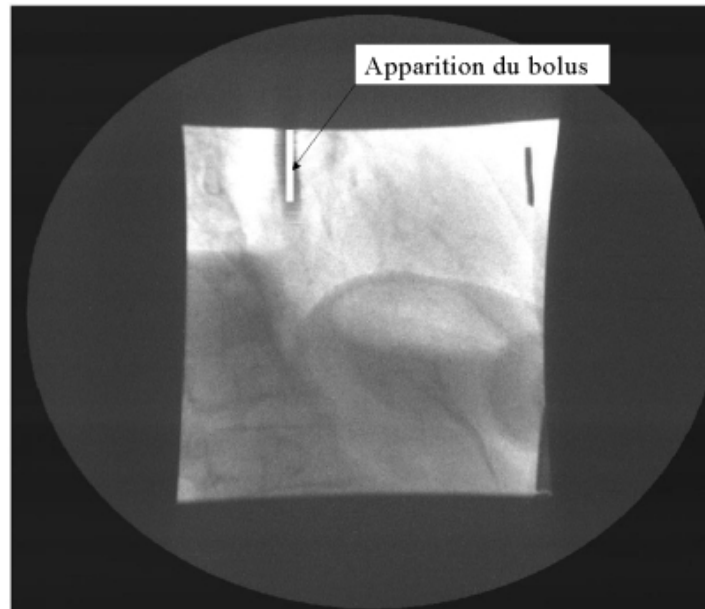


FIG. 2.5 – Détection du bolus

Le contour de cette zone permet l'initialisation d'un algorithme de contours actifs présenté au paragraphe 2.3.4 permettant de déterminer avec précision le contour du bolus. La méthode de suivi peut alors être appliquée afin de suivre cet objet dans la séquence d'images.

2.4.2 Paramétrage de l'algorithme de suivi

L'initialisation des paramètres, que ce soit pour le filtre de Kalman ou pour les contours actifs, est une phase très importante. En effet, de la valeur des paramètres dépend le comportement de l'algorithme. En l'absence d'une méthode générale permettant de déterminer les combinaisons de valeurs optimales, il est souvent nécessaire de procéder de manière empirique.

Initialisation des paramètres de Kalman

Les valeurs des matrices de covariance des bruits du système Q et des mesures R utilisées sont les suivantes :

$$Q = (q_{ij}), \quad q_{ij} = 10^{-4} \quad \forall i \text{ et } j = 1 \text{ à } 12$$

et $R = 10^{-6} * Id_{6 \times 6}$ où $Id_{6 \times 6}$ est la matrice identité de taille 6×6 .

Initialisation des paramètres des contours actifs

Les valeurs des coefficients de pondération des différentes énergies sont les mêmes pour la phase de segmentation initiale et pour les phases d'ajustement, à l'exception du coefficient d'énergie ballon. Ces valeurs sont présentées ci-dessous :

$a = 1$ (énergie de continuité)

$b = 1$ (énergie de courbure)

$c = 1, 2$ (énergie de gradient)

$d = 1, 2$ (énergie d'intensité)

$e = 1, 9$ (énergie ballon) pour la segmentation initiale

et $e = 1, 1$ pour les segmentations d'ajustement.

La minimisation locale de l'énergie se fait sur un voisinage 7×7 .

La méthode de minimisation utilisée pour l'optimisation de la région et l'estimation des paramètres de mouvement est celle du recuit simulé.

Le pas d'échantillonnage utilisé pour la phase de rééchantillonnage du contour est de 4 pixels.

2.4.3 Résultats du suivi

La figure 2.6 présente des images extraites du suivi du bolus sur une séquence radiologique représentative. La figure 2.7 correspond au suivi du bolus dans le cas d'une séquence particulière, caractérisée par un faible contraste.

Le contour vert correspond au contour obtenu après optimisation de la région par le modèle de région déformable, et le contour rouge, au contour ajusté rééchantillonné.

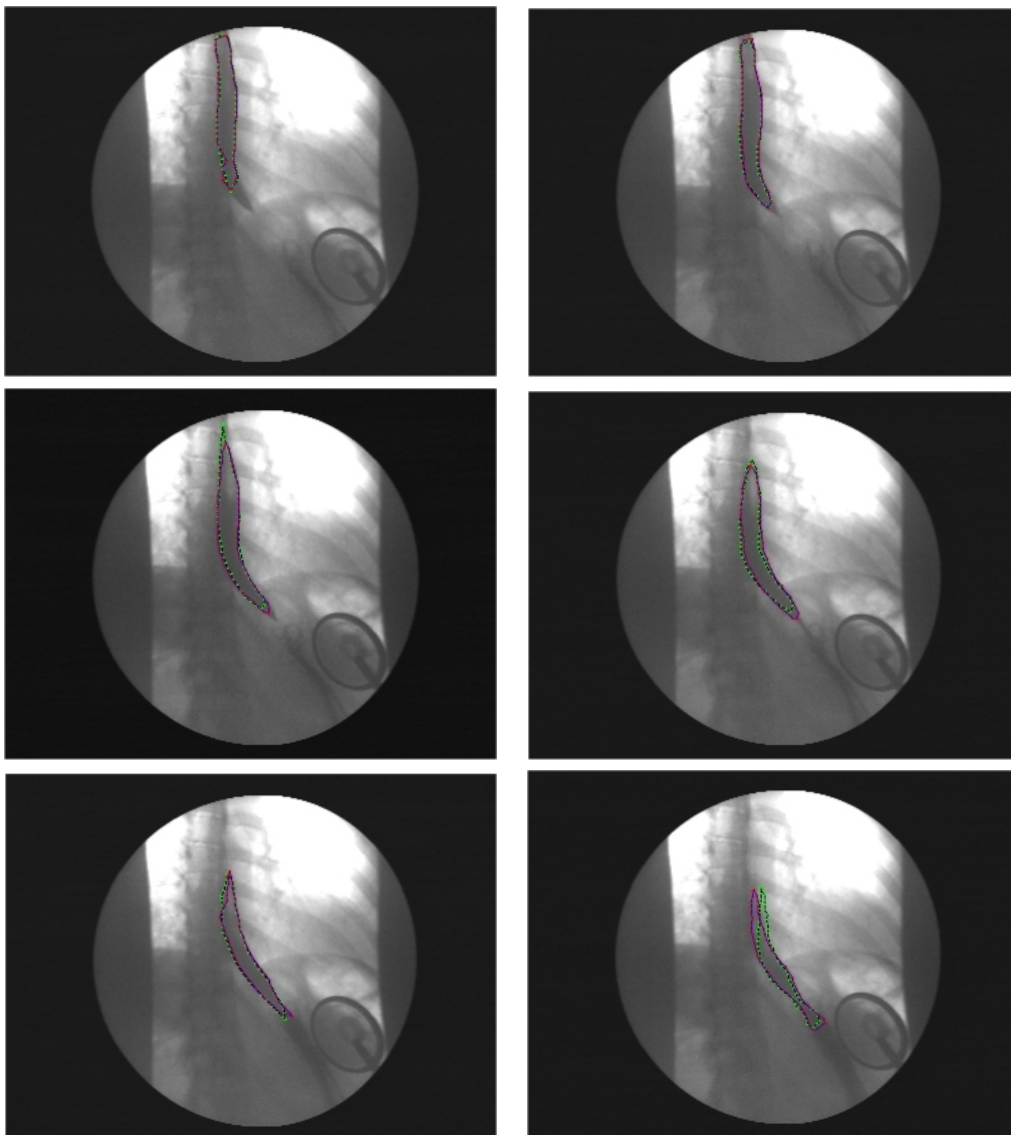


FIG. 2.6 – Images extraites du suivi du bolus pour la première séquence

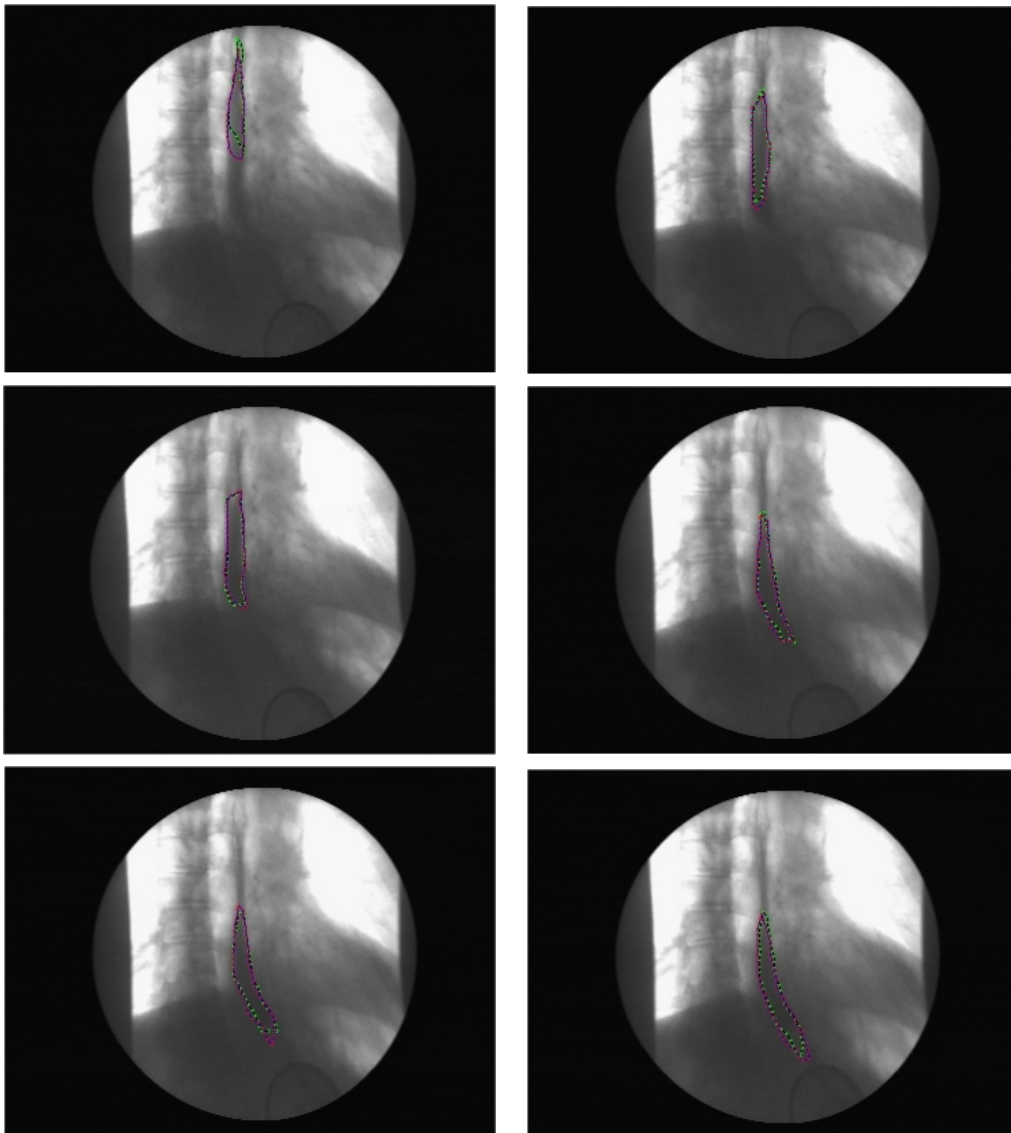


FIG. 2.7 – Images extraites du suivi du bolus pour la deuxième séquence

2.5 Conclusion

La méthode de suivi d'objets déformables exposée dans ce chapitre nous a permis de suivre un bolus baryté dans des séquences d'images radiologiques, lors de son transit à travers l'œsophage et le SIO.

Les résultats obtenus montrent que la méthode est efficace, le bolus étant bien suivi au cours des séquences d'images, malgré le faible contraste et le bruit de ces images. De plus, du fait de sa structure, la méthode pourrait être généralisée au suivi d'objets quelconques, aucune hypothèse relative à la forme ni à la texture de la région n'étant réalisée, et le modèle de mouvement utilisé permettant la modélisation de la plupart des mouvements.

Cependant, la méthode présente trois limitations. En effet, le nombre de paramètres à régler est relativement important, ce qui empêche une réelle automatisation du suivi. De plus, la méthode ne permet pas la gestion du changement de topologie, pouvant se produire lorsque le bolus se sépare en deux parties au niveau du SIO, ou lorsque plusieurs parties du bolus viennent à fusionner. Enfin, le coût calculatoire de l'algorithme est élevé du fait des nombreux modules et de l'utilisation de la méthode du recuit simulé comme méthode d'optimisation.

C'est pourquoi nous avons développé une méthode plus rapide ne présentant pas ces défauts, et reposant sur les ensembles de niveaux. Elle fait l'objet du chapitre suivant.

Chapitre 3

Suivi d'objets et ensembles de niveaux

3.1 Introduction

Ce chapitre est consacré à la description d'une méthode de suivi d'objets déformables que nous avons développée afin de s'affranchir des limites caractérisant la méthode présentée au chapitre précédent.

Notre objectif est donc d'élaborer une méthode de suivi d'objets déformables permettant un suivi rapide d'objets pouvant être soumis à un changement de topologie en réduisant autant que possible le paramétrage de l'algorithme afin de permettre une automatisation du suivi.

Pour cela, nous faisons l'hypothèse que les objets à suivre sont homogènes, c'est-à-dire que leur texture est relativement uniforme. Nous perdons de ce fait la généralité caractérisant la méthode de suivi présentée au chapitre précédent, la contrepartie étant la possibilité de mettre en œuvre des techniques permettant d'atteindre l'objectif fixé.

Ainsi, la méthode proposée repose sur un modèle de contours actifs dont l'évolution est basée sur une technique rapide issue de la théorie des ensembles de niveaux.

Nous allons présenter dans les sections suivantes un rapide historique des améliorations successives du modèle original de contours actifs destiné à la segmentation d'images, et aboutissant aux contours actifs reposant sur les ensembles de niveaux. Une présentation de cette technique sera ensuite réalisée. La méthode de

suivi développée sera finalement décrite, ainsi que son application au suivi d'un bolus baryté dans des séquences d'images radiologiques.

3.2 Les ensembles de niveaux

Le modèle des contours actifs a été l'objet de nombreuses études durant ces dernières années. En effet, les perspectives offertes par cette technique de segmentation sont très intéressantes puisqu'elle permet d'intégrer une modélisation à la segmentation, ce qui faisait défaut aux méthodes utilisées jusqu'alors dans le domaine de la segmentation d'images, ces méthodes ne prenant pas en compte la forme globale des contours, mais procédant par étude locale. Le modèle des contours actifs souffre néanmoins d'un certain nombre de limitations. Ainsi, la difficulté liée au choix des paramètres, l'influence importante de l'initialisation sur la convergence ainsi que les problèmes liés à la minimisation de la fonctionnelle d'énergie ont été discutés dans [BER91].

De nombreux travaux ont donc eu pour objectif l'amélioration du modèle. Une force de ballon [COH91], ainsi qu'une force appelée *gradient vector flow* [XU91] ont été introduites afin d'améliorer la segmentation, la première permettant au contour de se dilater, alors qu'il tend naturellement à se contracter, et la deuxième permettant au contour d'atteindre des zones concaves malgré l'influence des forces internes de régularité. Dans [WIL92], un algorithme rapide à minimisation locale, utilisé dans notre méthode (section 2.3.4), a été développé afin de réduire les problèmes liés à la minimisation de la fonctionnelle d'énergie.

Malgré ces améliorations du modèle des contours actifs, des problèmes importants demeurent. En effet, la contrainte de régularité interdit les changements de topologie. Il n'est donc pas possible, sans modification importante de l'algorithme, de détecter et de suivre plusieurs objets. De plus, afin de permettre une segmentation acceptable, le modèle doit être initialisé à proximité du contour de l'objet. Enfin, la paramétrisation du modèle, dépendant à la fois de l'objet à segmenter, et du type d'image, reste délicate à réaliser.

Une approche originale, proposée par [CAS93] et [MAL95] permet de s'affranchir de ces limites. Elle repose sur la théorie de l'évolution des courbes et la méthode des ensembles de niveaux. Les contours sont déplacés en utilisant seulement des mesures géométriques. Cette caractéristique est un avantage qui leur permet de s'affranchir

d'un paramétrage souvent difficile. De plus, l'évolution des courbes est couplée avec les données des images afin d'en extraire les contours.

3.2.1 Présentation générale

La méthode des ensembles de niveaux (ou level set), dont une présentation est fournie dans [OSH01a, OSH01b], a été développée à l'origine par Osher et Sethian dans [OSH88] pour résoudre les problèmes de propagation de fronts. Elle a été employée dans de nombreux types de problèmes mettant en jeu des contours en mouvement comme l'étude de la dynamique des fluides incompressibles multiphase [SUS94, ZHA96], des fluides compressibles [KOR02], l'étude de la structure cristalline des métaux [ELL03] et le déplacement de bulles [ZHA98c, SUS03]. Cette méthode numérique trouve également des applications dans des domaines tels que le traitement d'images, permettant la réduction de bruit et l'amélioration de la qualité des images [MAL96], ainsi que la segmentation de structures complexes ce qui est particulièrement recherché en médecine [SUR01]. Ainsi, l'étude des mouvements cardiaques est réalisée dans [CHA03], [HAS03] s'intéresse à la segmentation cérébro-vasculaire, [SCH00] applique sa méthode à la segmentation microscopique biomédicale et, plus généralement, [XU00] montre comment la méthode des ensembles de niveaux permet la segmentation de nombreuses structures organiques.

L'intérêt important porté à cette méthode a conduit à de nombreuses améliorations du modèle initial. Ainsi, une adaptation au suivi d'objets dans des séquences d'images est réalisée dans [MAN02], une approche de type région a été intégrée de manière à augmenter la robustesse de la segmentation dans [JB03] et dans [PAR99]. Enfin, des améliorations concernant la vitesse d'exécution ont été apportées dans [MAL98b, PEN99, NIL03].

L'idée générale de cette méthode est de décrire une surface à n dimensions par le niveau zéro d'une fonction à $n + 1$ dimensions. Ainsi, dans le cas du traitement d'images bidimensionnelles, le contour de l'objet étudié sera décrit par le niveau zéro d'une surface tridimensionnelle. L'évolution de ce contour est liée à l'évolution de la surface de dimension supérieure. Cette évolution correspond à une propagation de fronts. Un front est une courbe qui se déplace et qui réalise l'interface entre deux régions.

L'évolution est généralement réalisée en déplaçant chaque point du contour avec une vitesse F en direction du vecteur normal au contour \vec{N} , soit vers l'intérieur de

la forme, soit vers l'extérieur.

On peut par exemple repérer la séparation entre la partie solide (partie foncée) et la partie liquide (partie claire) d'un bloc de glace en train de fondre, et déterminer l'évolution de cette interface (figure 3.1).

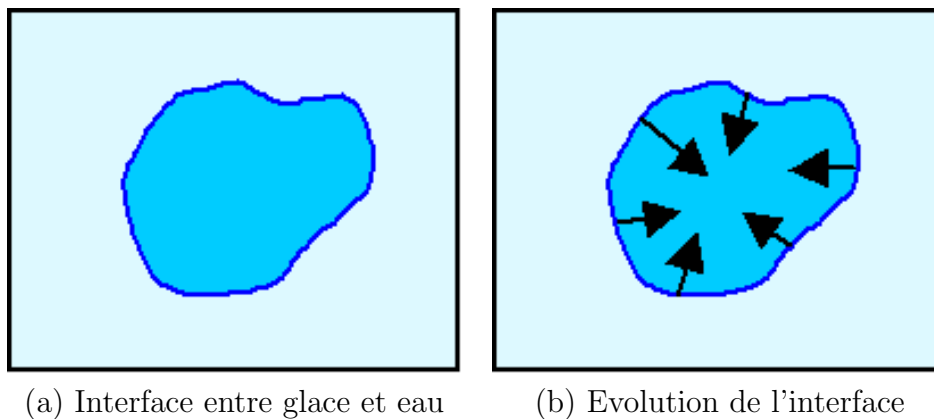


FIG. 3.1 – Evolution de l'interface entre de la glace et de l'eau.

La plupart des méthodes qui traitent ce problème discrétisent l'interface en plusieurs points particuliers et font déplacer ces points. L'interface peut alors être reconstituée en reliant les points par des interpolations linéaires ou non linéaires (β -spline).

Ce type de résolution est efficace pour des cas simples, mais il devient beaucoup moins performant en cas de changement de topologie, lorsque la région à l'intérieur du front se sépare en deux sous-régions ou que plusieurs régions fusionnent.

La méthode des ensembles de niveaux résout ce problème d'une tout autre façon. Au lieu de suivre l'interface en elle-même, elle la considère comme une courbe tracée sur une surface. Cette surface a la propriété suivante : à chaque instant, son niveau 0, pouvant être considéré comme son intersection avec un plan d'équation $z = 0$ dans le cas d'un espace à 3 dimensions (x, y, z) , correspond à la position du front que l'on suit. Cette surface est appelée fonction d'ensemble de niveaux.

Par exemple, pour représenter un cercle avec la méthode des ensembles de niveaux, on définit une surface dont le niveau 0 correspond à ce cercle. Ainsi, la figure 3.2 représente un contour circulaire et une représentation en 3 dimensions de la fonction d'ensembles de niveaux associée.

L'évolution du contour est alors réalisée de manière indirecte par l'évolution de la surface, de manière à ce que le niveau 0 de cette surface coïncide avec le front

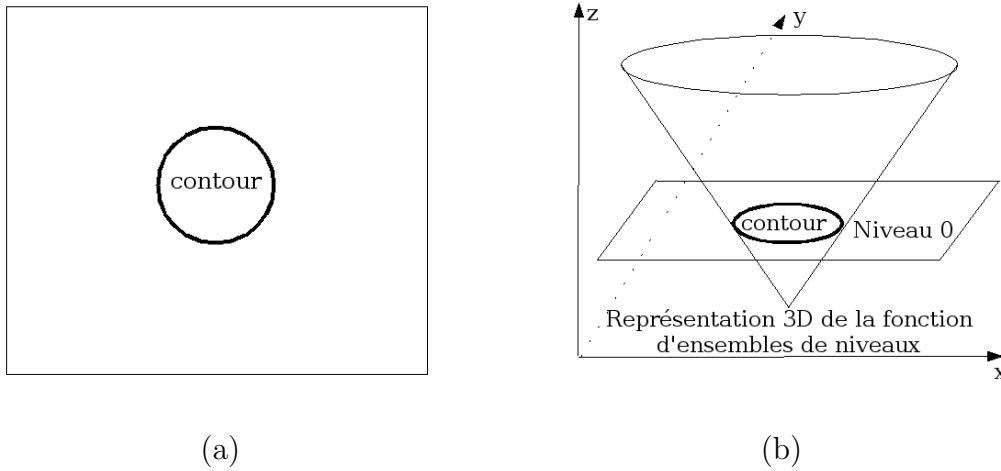


FIG. 3.2 – Un contour (a) et la représentation en 3 dimensions de la fonction d'ensembles de niveaux associée (b).

suivi.

Une des propriétés intéressantes de cette méthode est de prendre naturellement en compte les changements de topologie.

3.2.2 Formulation mathématique

Evolution de la surface

Considérons une courbe paramétrée v définie par la relation (2.23), que nous confondrons avec sa représentation paramétrique.

La courbe v représente le contour initial associé à l'objet à suivre. On définit alors $v(s, t)$ comme une famille de courbes générées par le mouvement de v en direction de son vecteur normal interne \vec{N} en chaque point du contour, au cours du temps t .

En définissant la valeur de la projection de la vitesse $F(v(s))$ sur la normale en chaque point $v(s)$ de la courbe, il est alors possible d'exprimer le déplacement au cours du temps de ce contour par l'équation (3.1).

$$\frac{\partial v}{\partial t}(s, t) = F(v(s))\vec{N} \quad (3.1)$$

Afin de mettre en œuvre la méthode des ensembles de niveaux, la surface peut être définie à partir de la fonction u (3.2), qui est la fonction d'ensemble de niveaux

correspondant à une représentation implicite des contours.

$$\begin{aligned} u : \mathbb{R}^2 \times [O, T] &\rightarrow \mathbb{R} \\ p, t &\mapsto u(p, t) \end{aligned} \quad (3.2)$$

où T est le temps que met la courbe à atteindre le contour de l'objet à segmenter, et p un point de coordonnées (x, y) .

La propriété de la surface associée à la fonction u est que initialement sa courbe de niveau 0 coïncide avec le contour initial (3.3).

$$u(v(s, 0), 0) = 0, \quad \forall s \in [a, b] \quad (3.3)$$

Cette propriété est conservée lors de l'évolution. Ainsi, à chaque instant t , la relation (3.4) est vérifiée.

$$u(v(s, t), t) = 0, \quad \forall s \in [a, b], \quad \forall t \in [0, T] \quad (3.4)$$

La fonction u est usuellement choisie comme la carte des distances signées au contour $v(s, 0)$, c'est-à-dire que pour tout point p et à tout instant t , $u(p, t)$ représente la distance algébrique du point p au contour $\Gamma(s, t)$. Les valeurs associées aux points situés à l'intérieur du contour sont choisies négatives par convention. Elles sont positives pour les points situés à l'extérieur du contour.

En dérivant partiellement l'expression (3.4) par rapport à t , en utilisant le fait que le vecteur unitaire normal à la courbe peut être exprimé par $\vec{N} = -\frac{\nabla u}{|\nabla u|}$, et en considérant que la fonction u a des valeurs négatives pour les points dont la projection sur le plan contenant v est à l'intérieur du contour et positives à l'extérieur, on démontre alors que l'évolution de $v(s, t)$ peut être représentée par la courbe de niveau 0 de la surface associée à la fonction u , évoluant suivant l'équation (3.5).

$$\frac{\partial u}{\partial t}(p, t) = F(p)|\nabla u| \quad (3.5)$$

Discrétisation

L'équation d'évolution de la fonction d'ensembles de niveaux u nécessite le calcul de ses dérivées premières et parfois secondes, selon l'expression de F . Or, dans

un espace discrétisé, l'évaluation de ces valeurs doit être correctement réalisée afin d'éviter la génération d'erreurs d'approximation qui entraîneraient un comportement incorrect du processus d'évolution de u . Ainsi, un schéma numérique, basé sur les équations d'Hamilton-Jacobi, a été développé par [OSH88] afin de permettre la résolution numérique de l'équation d'évolution.

L'équation (3.6) correspond à la forme discrète de l'équation d'évolution (3.5) en utilisant cette méthode numérique. Elle permet de calculer la position de la surface à l'instant $t + 1$ à partir de la position de la surface à l'instant t .

$$u_{x,y}^{t+1} = u_{x,y}^t + \Delta t (\max(F_{x,y}, 0) \nabla^+ + \min(F_{x,y}, 0) \nabla^-) \quad (3.6)$$

où $u_{x,y}^{t+1}$ et $u_{x,y}^t$ représentent la valeur de u au point de coordonnées (x, y) respectivement aux instants $t + 1$ et t . $F_{x,y}$ est la valeur de la fonction vitesse au point de coordonnées (x, y) à l'instant t .

Les gradients discrets ∇^+ et ∇^- s'expriment de la manière suivante :

$$\nabla^+ = [\max(D_{x,y}^{-x}u, 0)^2 + \min(D_{x,y}^{+x}u, 0)^2 + \max(D_{x,y}^{-y}u, 0)^2 + \min(D_{x,y}^{+y}u, 0)^2]^{1/2} \quad (3.7)$$

$$\nabla^- = [\max(D_{x,y}^{+x}u, 0)^2 + \min(D_{x,y}^{-x}u, 0)^2 + \max(D_{x,y}^{+y}u, 0)^2 + \min(D_{x,y}^{-y}u, 0)^2]^{1/2} \quad (3.8)$$

Ces gradients font intervenir les dérivées partielles du premier ordre de u , dont les expressions sont :

$$D_{x,y}^{+x}u = \frac{u_{x+1,y} - u_{x,y}}{\Delta x} \quad (3.9)$$

$$D_{x,y}^{-x}u = \frac{u_{x,y} - u_{x-1,y}}{\Delta x} \quad (3.10)$$

$$D_{x,y}^{+y}u = \frac{u_{x,y+1} - u_{x,y}}{\Delta y} \quad (3.11)$$

$$D_{x,y}^{-y}u = \frac{u_{x,y} - u_{x,y-1}}{\Delta y} \quad (3.12)$$

De plus, il est parfois nécessaire d'utiliser les approximations centrées des dérivées partielles du premier ordre ($D_{x,y}^{0x}u$ et $D_{x,y}^{0y}u$), ainsi que du second ordre de u ($D_{x,y}^{0xx}u$, $D_{x,y}^{0yy}u$ et $D_{x,y}^{0xy}u$) :

$$D_{x,y}^{0x}u = \frac{u_{x+1,y} - u_{x-1,y}}{2\Delta x} \quad (3.13)$$

$$D_{x,y}^{0y}u = \frac{u_{x,y+1} - u_{x,y-1}}{2\Delta y} \quad (3.14)$$

$$D_{x,y}^{0xx}u = \frac{u_{x+1,y} - 2u_{x,y} + u_{x-1,y}}{2\Delta x} \quad (3.15)$$

$$D_{x,y}^{0yy}u = \frac{u_{x,y+1} - 2u_{x,y} + u_{x,y-1}}{2\Delta y} \quad (3.16)$$

$$D_{x,y}^{0xy}u = \frac{u_{x+1,y+1} - u_{x+1,y-1} - u_{x-1,y+1} + u_{x-1,y-1}}{2\Delta x\Delta y} \quad (3.17)$$

Initialisation de la fonction d'ensembles de niveaux

La fonction d'ensembles de niveaux u est généralement choisie comme une carte de distance, correspondant à la distance algébrique entre le point considéré et le point de contour le plus proche. Les points intérieurs se verront assigner une distance négative et les points extérieurs une distance positive.

Pour calculer cette fonction, plusieurs algorithmes ont été développés, donnant lieu à des vitesses d'exécution et des exactitudes des résultats très différentes. Deux de ces méthodes vont être présentées dans les sections suivantes.

Algorithme avec calcul des distances euclidiennes pour tous les points

Cet algorithme consiste à déterminer dans un premier temps les points de l'image appartenant au contour, puis à calculer en tout point la distance au point de contour le plus proche.

Cette méthode a l'avantage d'être très simple et précise. Cependant, elle est caractérisée par un temps de calcul très important. Elle pourra donc être utilisée lorsque une bonne précision sera nécessaire et que la contrainte de temps sera faible.

Algorithme de transformation de distance

L'opération *transformation de distance* (ou *distance transform*) [ROS95] permet de déterminer la distance à un objet en chaque point de l'image, et repose sur l'hypothèse qu'il est possible de déterminer la valeur de distance en un point à

partir de la valeur de distance de ses voisins. Cette méthode présente donc l'intérêt d'être très rapide, puisqu'elle nécessite uniquement deux balayages de l'image. Elle peut être vue comme un opérateur qui, appliqué sur une image binaire, donne une image dans laquelle chaque point prend une valeur proche de la distance au point du contour le plus proche.

Un exemple est présenté figure 3.3. Dans un souci de simplification, seul l'intérieur du contour a été considéré. La première image représente l'image initiale, la partie blanche étant la partie interne dans laquelle les distances vont être calculées. La deuxième image correspond aux distances obtenues après application de l'opérateur de *transformation de distance*. Les points internes ont été remplacés par une valeur de gris proportionnelle à leur éloignement au contour.

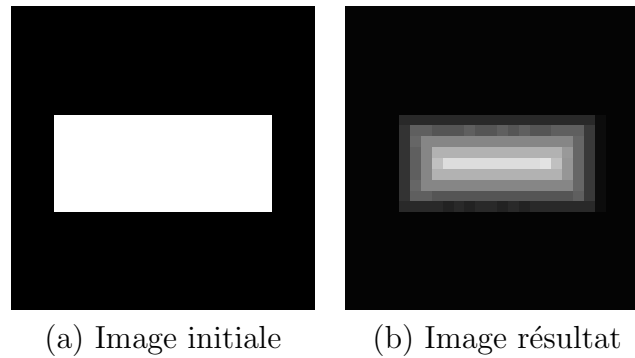


FIG. 3.3 – Illustration de l'algorithme de *transformation de distance*.

L'opération de *transformation de distance* comporte plusieurs variantes dans lesquelles on utilise différentes métriques pour estimer les distances. Une métrique couramment utilisée est la métrique *chamfer* [ROS95]. La distance entre deux points q et p est alors exprimée par la relation (3.18).

$$d_{d_1, d_2}(p, q) = d_1(|q_x - p_x| - |q_y - p_y|) + d_2|q_y - p_y| \quad (3.18)$$

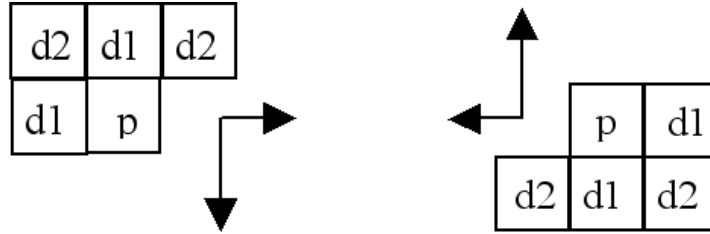
où d_1 et d_2 sont deux constantes.

Cet algorithme reposant sur l'hypothèse que la distance d'un point aux contours peut être déduite à partir de la valeur de ses voisins, deux étapes sont définies : initialisation de la valeur de distance associée aux points de l'image, puis calcul de la distance des points aux contours lors de deux balayages successifs de l'image.

La première étape de cette méthode consiste donc en l'initialisation de la valeur associée aux points de la surface correspondant aux contours, à la valeur 0. Tous les

autres points seront affectés d'une valeur infinie.

Deux masques, présentés figure 3.4, sont ensuite définis.



(a) Masque 1 : balayage avant (b) Masque 2 : balayage arrière

FIG. 3.4 – Masques utilisés par l'algorithme de *transformation de distance*.

L'utilisation de ces masques permet de déterminer la valeur de distance d au contour du point $p_{x,y}$ de coordonnées (x, y) :

La distance calculée par rapport au masque 1 est la suivante :

$$d = \min(p_{x,y} + 0, p_{x-1,y-1} + d_2, p_{x,y-1} + d_1, p_{x+1,y-1} + d_2, p_{x-1,y} + d_1) \quad (3.19)$$

Dans le cas du masque 2, la distance s'exprime par :

$$d = \min(p_{x,y} + 0, p_{x-1,y+1} + d_2, p_{x,y+1} + d_1, p_{x+1,y+1} + d_2, p_{x+1,y} + d_1) \quad (3.20)$$

Le principe de l'algorithme est alors de parcourir dans un premier temps l'image de gauche à droite et de haut en bas à partir du point supérieur gauche et en utilisant le masque 1 pour la marche en avant. Le parcours inverse est ensuite réalisé à partir du point inférieur droit en utilisant alors le masque 2 pour la marche en arrière.

Les valeurs des paramètres d_1 et d_2 déterminent la métrique qui sera utilisée. Si $d_1 = 1$ et $d_2 = 1,351$, la distance correspondante est proche de la distance euclidienne.

3.2.3 Les techniques d'évolution de front

L'implémentation de la méthode des ensembles de niveaux permet le calcul de la fonction d'ensembles de niveaux u au cours des itérations successives jusqu'à convergence. La méthode la plus simple consiste à calculer u en chaque point de son ensemble de définition, et à chaque itération. L'inconvénient de ce procédé est le

temps de calcul très élevé. C'est pourquoi des méthodes rapides de propagation ont été développées. Trois d'entre elles, correspondant aux méthodes les plus utilisées, vont être présentées dans les sections suivantes. Il s'agit de l'algorithme de la *bande étroite*, de la méthode à *progression rapide* et de l'algorithme *Hermès*.

Algorithme de la *bande étroite*

Le principe de l'algorithme de la *bande étroite* (*narrow band*) [ADA95, MAL95] est de ne calculer la fonction d'ensembles de niveaux u que pour les points situés dans une zone limitée entourant le niveau 0 de u , une bande étroite, et non pour la totalité des points de l'ensemble de définition de u . Cette méthode est illustrée figure 3.5. L'avantage de cet algorithme est qu'il permet de diminuer de manière importante le nombre de calculs réalisés à chaque itération pour faire avancer le front. Ainsi, le surcroît de calculs induits par le traitement du problème dans une dimension supérieure peut être compensé au moins partiellement.

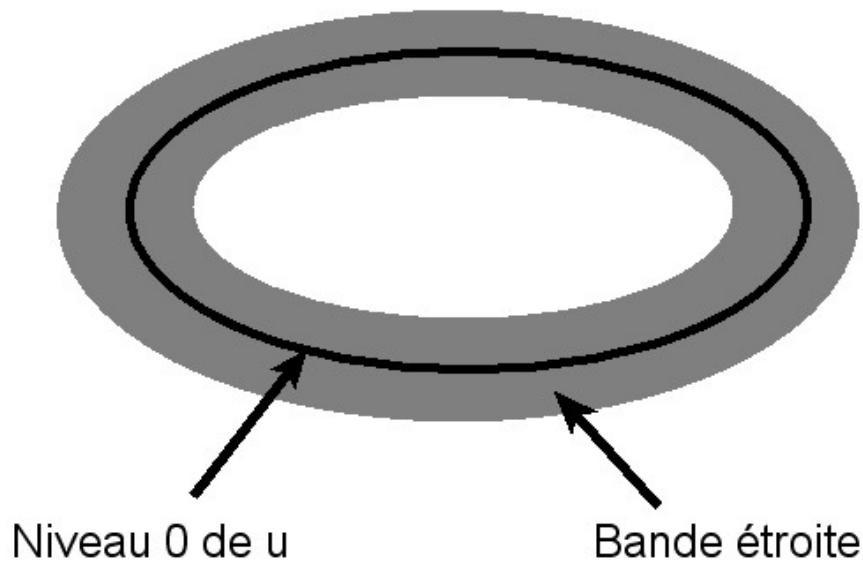


FIG. 3.5 – Principe de l'algorithme de la *bande étroite*.

Lorsque le front se propage, il est indispensable de mettre à jour la limite de la bande étroite en fonction de la position de l'interface. Cela est réalisé en considérant la valeur de la fonction d'ensembles de niveaux pour les points localisés au voisinage de l'interface, puisque cette fonction correspond à la carte des distances par rapport au niveau 0. Typiquement, la largeur de la bande est de 12 à 24 pixels.

Algorithme à *progression rapide*

L'algorithme à *progression rapide* (*fast marching*) [SET96,SET98], qui sera présenté de manière plus détaillée à la section 3.3.1 aborde le problème d'une manière différente. Plutôt que de chercher à suivre directement l'interface, une approche stationnaire est utilisée.

Le principe est de transformer l'équation d'évolution des ensembles de niveaux (3.5) en une équation (3.31) mettant en jeu le temps de passage du front $T(p)$ au point p , correspondant à l'instant de passage par zéro de la fonction u associée aux contours.

Le front va se propager depuis un ensemble de points constituant un germe, à l'ensemble de l'image. Le temps de passage du front est calculé pour chaque point, à partir de la valeur associée à ses voisins.

Dans l'exemple proposé figure 3.6, le front se propage vers l'extérieur.

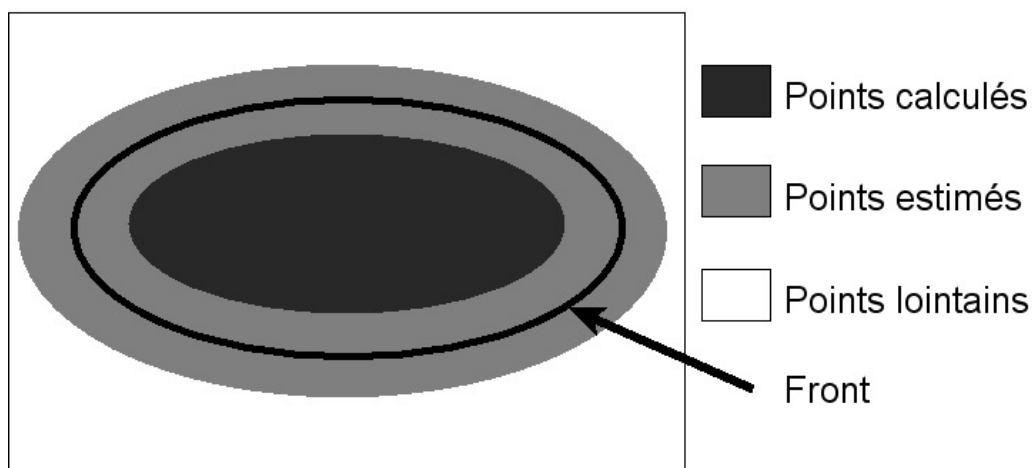


FIG. 3.6 – Illustration de la méthode d'évolution à *progression rapide*.

Bien que cette méthode soit très rapide, elle ne peut être utilisée que pour un nombre limité d'applications. La contrainte vient du fait que la fonction de vitesse doit être toujours positive ou toujours négative.

Algorithme Hermès

Cet algorithme proposé dans [PAR98,PAR00] est inspiré par l'algorithme HPF (Highest Confidence First) et combine les atouts des algorithmes à progression rapide

et à bande étroite en utilisant les idées de propagation sélective (bande étroite) et de fenêtre relativement petite (progression rapide).

Le temps d'exécution de l'algorithme Hermès est inférieur à celui de l'algorithme à bande étroite, cependant il reste supérieur à l'algorithme à progression rapide, mais ne possède pas la contrainte d'une fonction vitesse à signe constant.

3.2.4 Détection de contours

L'idée principale de la segmentation d'objets dans des images par la méthode des ensembles de niveaux est de coupler la vitesse de déformation avec les données de l'image considérée pour que l'évolution de la courbe s'arrête sur le bord des objets. Les recherches dans le domaine des contours actifs géométriques portent donc sur la définition de fonctions de vitesse.

Il existe plusieurs types de fonctions de vitesse F à valeur dans \mathbb{R} , qui vont définir le type du modèle de contours actifs dont l'équation générale d'évolution est (3.21).

$$\frac{\partial u}{\partial t}(p, t) = F(I, l, L, U, \nabla u) |\nabla u| \quad (3.21)$$

$$u(p, t = 0) = u_0 \quad (3.22)$$

où I est l'image initiale, u est la représentation implicite des contours à l'instant t , l et L sont des caractéristiques respectivement locales et globales de u . La fonction de vitesse F dépend donc des caractéristiques géométriques de u et de l'image I .

Les caractéristiques locales ou globales l et L permettent d'imposer des contraintes sur la forme des objets à segmenter. Elles se calculent à partir de la fonction u ou de ses dérivées d'ordres supérieurs.

Une des premières fonctions vitesse utilisées (3.23) ne dépend que de la courbure de u . Elle a pour effet de lisser la courbe et de la faire rétrécir jusqu'à devenir un point. L'utilisation de la courbure donne le même effet que la force interne élastique dans le modèle des contours actifs déformables paramétriques (snakes).

$$F(p) = \alpha \kappa(p) \quad (3.23)$$

où α est une constante positive et $\kappa(p)$ est la courbure de la courbe au point p .

Lorsque la vitesse est de norme constante (3.24), l'effet est semblable à celui de la force de pression (force ballon) des contours actifs paramétriques. Une valeur positive entraîne un rétrécissement de la courbe alors qu'une valeur négative entraîne une dilatation de cette dernière.

$$F = V_0 \quad (3.24)$$

où V_0 est une constante.

La première fonction vitesse conçue pour la détection de contours (3.25) a été proposée dans [CAS93].

$$F = g(\kappa + V_0) \quad (3.25)$$

où g est une fonction positive strictement décroissante utilisée pour détecter les gradients d'une image (3.26) et V_0 est une constante.

$$g = \frac{1}{1 + |\nabla \hat{I}|^q} \quad (3.26)$$

où \hat{I} est une version floue de l'image à segmenter, obtenue par application d'un filtre gaussien, par exemple.

Dans ce cas, la vitesse du front est couplée à l'image à segmenter par le terme "stoppeur" multiplicatif g . Cette méthode est adaptée aux images qui présentent un bon contraste. Cependant, lorsque le bord de l'objet est diffus, le contour risque de "fuir". Le terme multiplicatif g n'est pas assez proche de 0 pour arrêter complètement la progression de la courbe. Tout au plus, il la ralentit. Une fois les bords de l'objet dépassés, le contour ne sera plus retenu et continuera à évoluer.

Pour contrer ce problème, Caselles et al [CAS93] et Maladi et al [MAL95] ont développé une fonction vitesse inspirée du problème de minimisation d'énergie de la méthode des snakes : les contours actifs géodésiques. Ces derniers permettent d'apporter une réponse aux limitations que les contours déformables paramétriques comme les "snakes" génèrent :

- Il est nécessaire de déterminer empiriquement les valeurs adaptées des coefficients de pondération des différentes énergies,
- Les changements de topologie ne sont pas pris en compte,
- La minimisation de la fonctionnelle d'énergie peut mener à des instabilités

numériques.

Les contours actifs géodésiques reposent sur la théorie de l'évolution des courbes et la méthode des ensembles de niveaux. Les contours sont déplacés en utilisant seulement des mesures géométriques. Cette caractéristique est un avantage qui leur permet de s'affranchir d'un paramétrage souvent difficile. Comme dans la méthode des snakes, l'évolution des courbes est liée aux données d'une image pour en extraire les contours.

Caselles et al [CAS93, CAS97] ont montré comment, à partir de l'équation générale de la méthode des "snakes", il est possible d'aboutir à une équation d'évolution pouvant être implémentée en utilisant les ensembles de niveaux, et donnant un résultat équivalent. Ce résultat pour le cas d'images à 2 dimensions a été étendu dans [AUB98] pour les problèmes à 3 dimensions. La nouvelle équation obtenue a l'avantage de supprimer les limites de la méthode des snakes. De plus, une méthode d'implémentation rapide a été proposée récemment dans [GOL01].

Ainsi, la minimisation de la fonctionnelle d'énergie associée au modèle initial des contours actifs (2.24) est équivalente à l'équation d'évolution de la courbe décrite dans (3.27) [CAS93].

$$\frac{\partial v}{\partial t} = g(V_0 + \kappa)\vec{N} - (\nabla g \cdot \vec{N})\vec{N} \quad (3.27)$$

où κ représente la courbure et \vec{N} est le vecteur normal intérieur. La position de la courbe correspondant aux contours de l'objet est donnée par la solution stationnaire de cette équation.

On peut alors en déduire que la surface associée aux ensembles de niveaux, qui permet de suivre la courbe définie par cette précédente équation, se déforme selon l'équation (3.28).

$$\frac{\partial u}{\partial t} = g(V_0 + \kappa)|\nabla u| - (\nabla g \cdot \nabla u) \quad (3.28)$$

Le terme V_0 représente la force ballon de la méthode des snakes. Il assure au contour une certaine vitesse de propagation. Le choix de ce paramètre est délicat, et des méthodes d'estimation ont été proposées dans [CAS93].

Le terme κ représente la force élastique dans la méthode des snakes. Il a pour fonction de lisser la courbe pour éviter que des angles aigus ne se forment.

Le terme g a pour fonction de stopper la progression de la courbe quand elle passe sur le bord d'un objet.

Le terme $\nabla g \cdot \nabla u$ a pour but d'empêcher le front de dépasser le bord de l'objet. En effet, la fonction g ne peut pas être rigoureusement égale à 0 sur le bord d'un objet. La courbe ne s'arrête donc pas complètement. Le produit oblige le contour à rester dans la "vallée" formée autour du contour.

La courbure κ en un point de la surface u peut être exprimée par l'équation (3.29).

$$\kappa = \operatorname{div} \left(\frac{\nabla u}{|\nabla u|} \right) \quad (3.29)$$

Enfin, lorsque la fonction vitesse F dépend uniquement de l'image I et ne change pas de signe au cours de l'évolution, on aboutit à un autre modèle de contours actifs présentant un grand intérêt dans différentes applications en traitement d'images [SCH00]. Ce modèle correspond à la version non régularisée du modèle général (3.21), et sera présenté plus en détail au paragraphe suivant.

3.3 Suivi par contours actifs non régularisés

L'algorithme de suivi présenté à la section 2.3 a l'avantage d'être robuste au suivi d'objets complexes dans des images bruitées et lors de déplacements importants de l'objet à suivre. Cependant, il présente l'inconvénient d'être lent, du fait des nombreuses étapes d'optimisation qui le composent. De plus, il nécessite le réglage de nombreux paramètres et ne permet pas la gestion du changement de topologie.

Ainsi, lorsque l'objet à suivre est homogène, c'est-à-dire que sa texture est relativement uniforme, et que la contrainte de temps doit être prise en compte, il peut être intéressant d'utiliser les propriétés des modèles de contours actifs non régularisés. Ils ont l'avantage d'être rapides à calculer puisqu'ils ne sont soumis à aucune contrainte de forme de l'objet, et que leur évolution peut être réalisée par l'implémentation d'une technique de propagation de front rapide.

La méthode de suivi reposant sur les contours actifs non régularisés que nous avons développée va donc être présentée dans les paragraphes suivants.

3.3.1 Les contours actifs non régularisés

Les contours actifs non régularisés sont une famille de contours actifs dont l'évolution repose sur la méthode des ensembles de niveaux. L'équation d'évolution de la fonction d'ensembles de niveaux u est donnée par (3.30).

$$\frac{\partial u}{\partial t}(p, t) = F(p, I)|\nabla u| \quad (3.30)$$

Les contours actifs non régularisés sont donc caractérisés par l'hypothèse que la fonction vitesse F ne dépend que des informations contenues dans l'image I et ne change pas de signe au cours de l'évolution.

Il est alors possible de convertir l'équation des ensembles de niveaux en une équation stationnaire, permettant l'utilisation de la méthode de propagation de front rapide, dont le principe général a été présenté à la section 3.2.3.

Pour cela, on définit T qui est une fonction qui associe à tout point p l'instant de passage par zéro de la fonction $t \mapsto u(p, t)$ associée aux contours. $T(p) = \inf\{t/u(p, t) = 0\}$ est donc le temps qu'a mis le front à atteindre le point p , dans le cas où F garde un signe constant.

On peut montrer [SET98] que le temps de passage T du front en chaque point p est gouverné par l'équation (3.31).

$$\begin{aligned} |\nabla T(p)| \cdot F(p) &= 1 \\ T(\Gamma_0) &= 0 \end{aligned} \quad (3.31)$$

où Γ_0 représente le contour initial.

Cette équation et ses variantes sont d'un grand intérêt dans différents problèmes de traitement et d'analyse d'images. Par exemple, elles sont à la base de la formulation continue de certaines transformations morphologiques telles que la ligne de partage des eaux, du calcul des distances pondérées, de certaines méthodes de squelettisation, et des méthodes d'extraction et de localisation d'objets.

L'équation (3.31) est adaptée aux valeurs discrètes en considérant le schéma numérique présenté à la section 3.2.2 :

$$\begin{aligned} F_{x,y}^{-2} &= \max(D_{x,y}^{-x}T, 0)^2 + \min(D_{x,y}^{+x}T, 0)^2 \\ &\quad + \max(D_{x,y}^{-y}T, 0)^2 + \min(D_{x,y}^{+y}T, 0)^2 \end{aligned} \quad (3.32)$$

Le principe de l'algorithme de propagation de front rapide est donc de calculer le temps de passage du front en chaque point, correspondant à la solution maximale de l'équation (3.32).

La fonction vitesse F est choisie de manière à permettre la segmentation d'objets. Une possibilité fréquemment retenue est d'utiliser l'information de gradient dans l'image. Ainsi, plus le gradient est important et plus la vitesse doit être faible, ce qui peut être formalisé par la relation (3.33).

$$F = \frac{1}{1 + |\nabla I|} \quad (3.33)$$

où ∇I est le gradient de l'image obtenue par application d'un filtre de Sobel [COC95].

L'algorithme le plus rapide permettant le calcul du temps de passage du front en chaque point de l'image à partir de l'équation (3.31) a été proposé dans [SET98]. Cet algorithme est basé sur le fait que cette équation impose que l'information soit propagée uniquement dans une direction (des petites vers les grandes valeurs de T) et que la solution puisse être obtenue par une méthode de propagation vers l'intérieur ou l'extérieur à partir de la plus petite valeur de T . Une implémentation rapide peut être obtenue en créant une mince bande de construction en avant du front. L'algorithme pousse le front en ne considérant que les pixels à l'intérieur de la bande.

Les points de l'image sont répartis en trois classes :

- points calculés : les points de l'image qui ont déjà été rejoints par le front et donc dont le temps de passage est connu
- points estimés : les points qui peuvent être atteints par le front dans l'itération courante
- points lointains : les points qui n'ont pas encore été considérés et dont le temps de passage n'est pas connu.

L'algorithme d'évolution est composé d'une étape d'initialisation suivie d'une étape de propagation.

1. Phase d'initialisation :

- (a) Les points p calculés sont initialisés à un temps de passage égal à zéro ($T(p) = 0$).
- (b) Les points p estimés sont initialisés à un temps de passage qui est inversement proportionnel à leur vitesse de propagation ($T(p) = 1/F(p)$).

- (c) Les points p lointains sont affectés d'un temps de passage égal à l'infini ($T(p) = \infty$).

2. Phase de propagation :

- (a) Sélection parmi les points estimés, du point associé à la plus petite valeur de T , noté p_{min} .
- (b) Ce point p_{min} passe de l'ensemble des points estimés à l'ensemble des points calculés.
- (c) Pour chaque point p voisin de p_{min} (4-connexe en 2D) :
- Si p est un point lointain, on l'ajoute à l'ensemble des points estimés et on calcule T en résolvant l'équation (3.31)
 - Si p est un point estimé, on recalcule T en ce point et on le remet à jour
 - Si p est un point déjà calculé, on ne fait rien.

L'algorithme s'arrête lorsque tous les points sont calculés.

3.3.2 Méthode de suivi

Le suivi par contours actifs non régularisés que nous avons développé permet de suivre un ou plusieurs objets uniformes (dont la texture est relativement homogène) dans une séquence d'images. Afin d'illustrer les différentes phases de la méthode, l'exemple du suivi d'un bolus baryté (petite quantité d'eau barytée opaque aux rayons X) dans des séquences d'images radiologiques centrées sur la jonction gastro-œsophagienne présenté à la section 1.2 sera utilisé.

La méthode de suivi est composée de deux étapes : une phase d'initialisation et une phase de propagation du front et de segmentation des objets à détecter. L'objet étant segmenté à l'instant t , l'initialisation à l'instant $t + 1$ pourra alors être réalisée. La procédure est donc appliquée à chaque image de la séquence, jusqu'à ce que l'ensemble de la séquence soit traité.

La phase d'initialisation consiste en l'initialisation du statut des points de l'image (calculés, estimés ou lointains).

Le front pourra ensuite se propager dans l'ensemble de l'image et les temps de passage T seront calculés en chaque point. Il sera alors possible de déterminer les points localisés à l'intérieur de l'objet à segmenter.

Cette segmentation réalisée, l'initialisation dans l'image suivante pourra être effectuée.

Ces différentes étapes vont être présentées de manière détaillée dans les sections suivantes.

Initialisation

La phase d'initialisation consiste en l'attribution d'un statut à chaque point de l'image. Afin de permettre la segmentation de l'objet, il est nécessaire d'initialiser un ensemble de points calculés et estimés à l'intérieur de celui-ci. Le front se propagera alors de ces points jusqu'à l'ensemble des points de l'image.

Ainsi, après avoir déterminé un ensemble de points situés à l'intérieur de l'objet, cet ensemble est divisé en deux sous-ensembles : l'ensemble des points calculés formant une composante connexe correspondant à un germe à partir duquel se propagera le front, et l'ensemble des points estimés, situés à la périphérie. En choisissant les points estimés de manière homogène autour de la composante connexe formée par les points calculés, on permet au front de se propager dans toutes les directions à partir du germe.

Deux phases d'initialisation sont à distinguer : l'initialisation sur la première image de la séquence considérée, et l'initialisation sur l'image à l'instant $t + 1$, connaissant la position de l'objet dans l'image à l'instant t .

L'initialisation sur la première image de la séquence peut être réalisée en déterminant un ensemble de points situés à l'intérieur de l'objet par l'utilisation d'une méthode de détection automatique, comme celle décrite à la section 2.4.1. Le procédé d'initialisation correspondant est illustré figure 3.7.

Au cours du suivi, la segmentation de l'objet à l'instant t permet l'initialisation des points de l'image à l'instant $t + 1$. Le germe constitué des points calculés devant nécessairement être situé à l'intérieur de l'objet, nous avons procédé à une squelettisation de l'objet à l'instant t . En effet, le squelette d'une forme, qui sera présenté à la section 4.2.2, correspond à un sous-ensemble de largeur 1 pixel de l'objet, situé en son "centre". Les points du squelette seront donc affectés du statut calculé, et les points situés à sa périphérie seront affectés du statut estimé, tous les autres points étant des points lointains. Cela est illustré figure 3.8.

Comme il sera mentionné à la section 4.2.2, le processus de squelettisation est

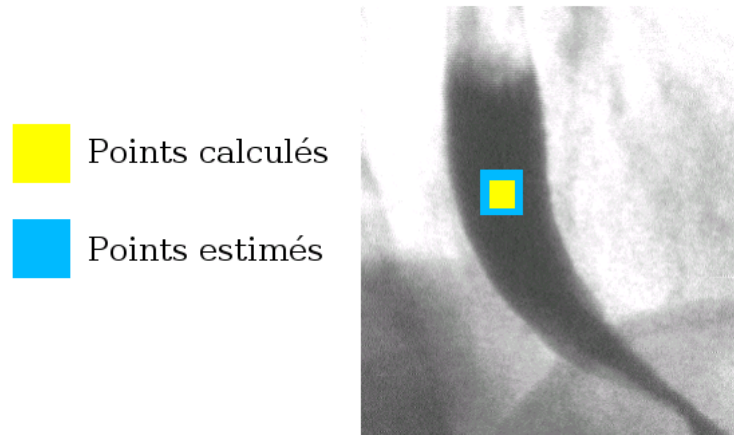


FIG. 3.7 – Exemple d’initialisation des points calculés et estimés sur la première image de la séquence à considérer.

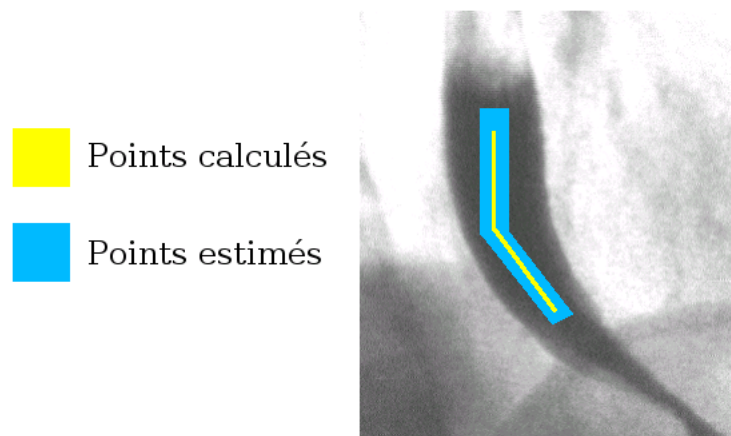


FIG. 3.8 – Exemple d’initialisation des points calculés et estimés sur une image de la séquence à l’instant $t + 1$, à partir de la segmentation et de la squelettisation de l’objet dans l’image à l’instant t .

sensible aux irrégularités du contour. En effet, des branches parasites du squelette, appelées *barbules*, peuvent apparaître. Pour réduire ce phénomène, on peut envisager une régularisation du contour avant transformation, telle qu’une fermeture morphologique (dilatation morphologique suivie d’une érosion morphologique) [COC95]. La figure figure 3.9 illustre ces différentes étapes.

Au cours de la séquence d’images, la topologie de l’objet peut être modifiée : l’objet suivi peut être amené à se séparer en plusieurs composantes, ou plusieurs objets peuvent être amenés à fusionner.

Afin de gérer la première situation, la distance de chaque point du squelette au

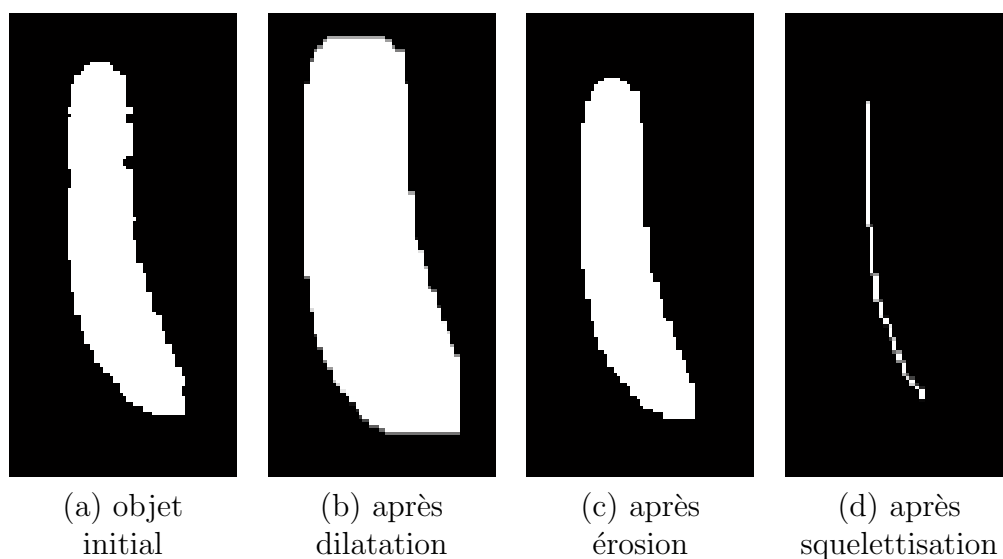


FIG. 3.9 – Exemple de squelettisation d'un objet, précédée d'une phase de lissage des contours par fermeture morphologique (l'élément structurant utilisé est un carré de dimension 3x3).

point de contour le plus proche est calculée à chaque instant t . Cela permet d'évaluer la proximité du contour aux points du squelette. Ainsi, lorsque cette distance est faible pour certains points du squelette, celui-ci est découpé au niveau des points correspondants.

La fusion des objets, quant à elle, est réalisée lorsque leur intersection est non-nulle. Le squelette global est alors recalculé.

Propagation et segmentation

Au cours du processus de propagation du front, le temps de passage T de ce dernier est calculé en chaque point de l'image selon l'équation (3.32) en utilisant l'algorithme à progression rapide présenté section 3.2.3.

Or, la vitesse F étant liée au gradient de l'image selon la relation (3.33), le front est freiné au niveau des zones de fort gradient dans l'image, ce qui est le cas des contours de l'objet. Puisque le front est initialisé à l'intérieur de l'objet, le franchissement de ses contours correspondra donc à une augmentation importante du temps de passage du front.

Afin de visualiser graphiquement la propagation du front, les valeurs du temps de passage peuvent être normalisées sur l'intervalle $[0; 255]$. Ainsi, plus le temps

de passage du front en un point sera élevé, et plus le niveau de gris de ce point dans l'image sera clair. La figure figure 3.10 présente l'image initiale et l'image des temps de passage du front obtenue après normalisation. La zone de propagation a été réduite afin de réduire les temps de calcul.

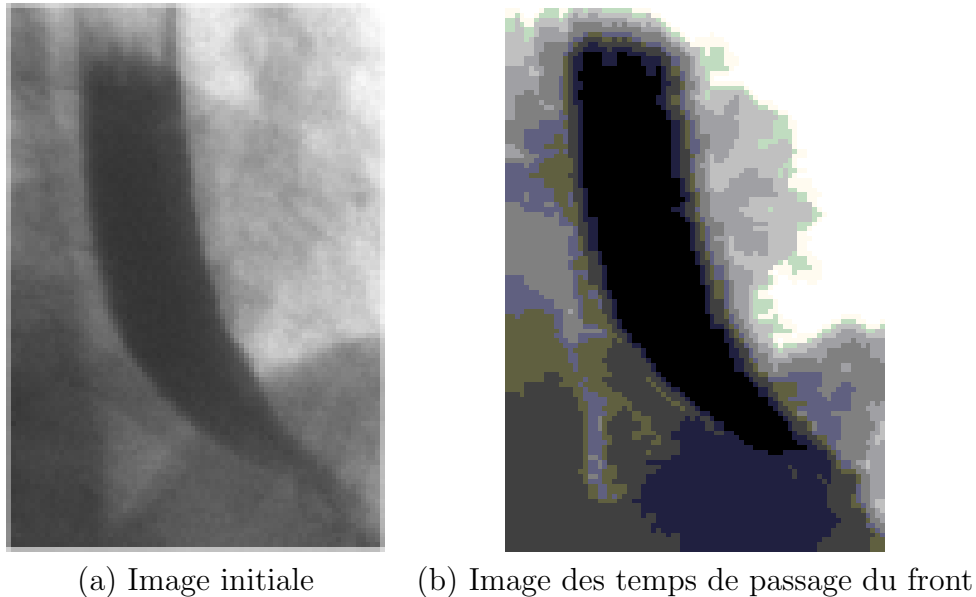


FIG. 3.10 – Exemple de représentation du temps de passage du front (b) à partir de l'image initiale (a).

Comme cela peut être observé dans la figure figure 3.10, les points situés à l'intérieur de l'objet sont caractérisés par une faible valeur du temps de passage du front. En effet, les points calculés ayant été utilisés pour l'initialisation sont associés à une valeur nulle. Le front ne rencontrant pas d'obstacle dans un premier temps, il va se propager rapidement aux points situés à l'intérieur de l'objet. Puis, freiné par les frontières de l'objet, il va être retardé, pour finalement franchir les contours. Cela se traduit au niveau de l'histogramme associé aux temps de passage du front par un nombre important de points associés à une valeur faible (points intérieurs à l'objet), puis un nombre réduit de points associés à une valeur légèrement plus importante (points de la partie interne aux contours), et enfin une augmentation importante du nombre de points associés à une valeur plus élevée du temps de passage (points extérieurs à l'objet). Ce phénomène peut être observé sur l'histogramme présenté figure 3.11. Les points de l'objet peuvent donc être déterminés en étudiant l'histogramme, et plus précisément, en détectant la première évolution croissante. Cela indique en effet que le front a alors franchi les frontières de l'objet.

Le point de rupture de l'histogramme correspondant à la première évolution crois-

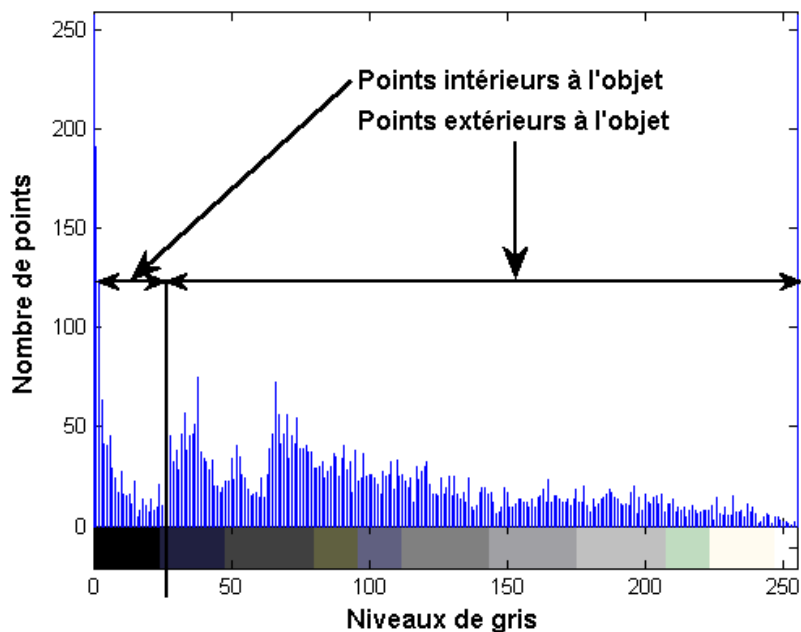


FIG. 3.11 – Histogramme des temps de passage du front normalisés dans l'intervalle $[0;255]$, correspondant à l'image présentée figure 3.10.

sante est déterminé par l'étude du gradient de l'histogramme, après avoir procédé à un lissage par filtre moyenneur afin de réduire le bruit. Les points associés aux valeurs inférieures à celle du point de rupture sont alors extraits, et correspondent aux points de l'objet. Cette segmentation de l'objet, ainsi que son squelette permettant l'initialisation dans l'image suivante sont présentés figure 3.12.

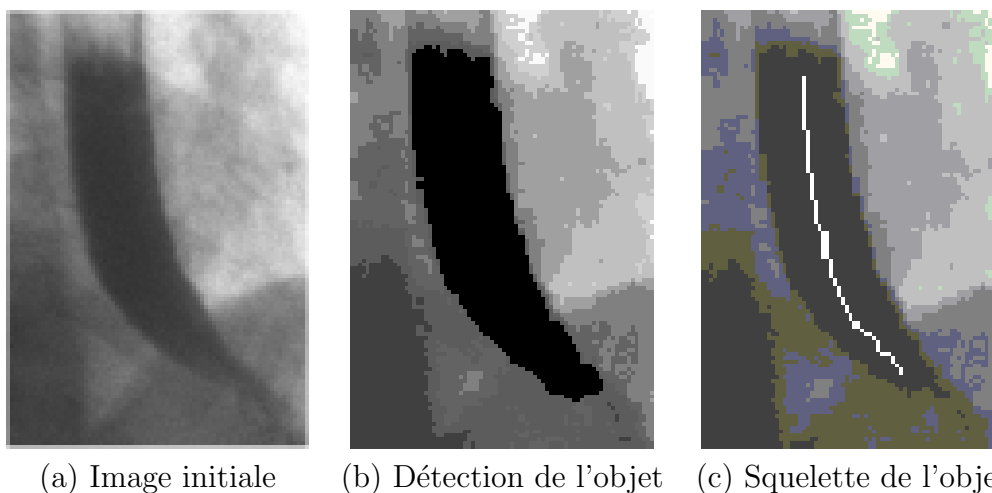


FIG. 3.12 – Segmentation de l'objet et squelettisation permettant l'initialisation dans l'image suivante.

Grâce à la segmentation de l'objet dans l'image à l'instant t , un certain nombre de paramètres caractéristiques, tels que ceux qui seront présentés au chapitre 4, vont pouvoir être calculés. De plus l'information de position de l'objet à l'instant t pourra être utilisée afin de permettre l'initialisation dans l'image à l'instant $t + 1$.

Les résultats du suivi par cette méthode basée sur les contours actifs non régularisés, sont présentés dans la section suivante.

3.4 Résultats

La figure 3.13 présente le suivi du bolus baryté dans une séquence d'images radiologiques. La zone sombre correspond à la segmentation du bolus.

La figure 3.14 est présentée afin d'illustrer le phénomène de changement de topologie de l'objet suivi, géré de manière automatique par la méthode de suivi.

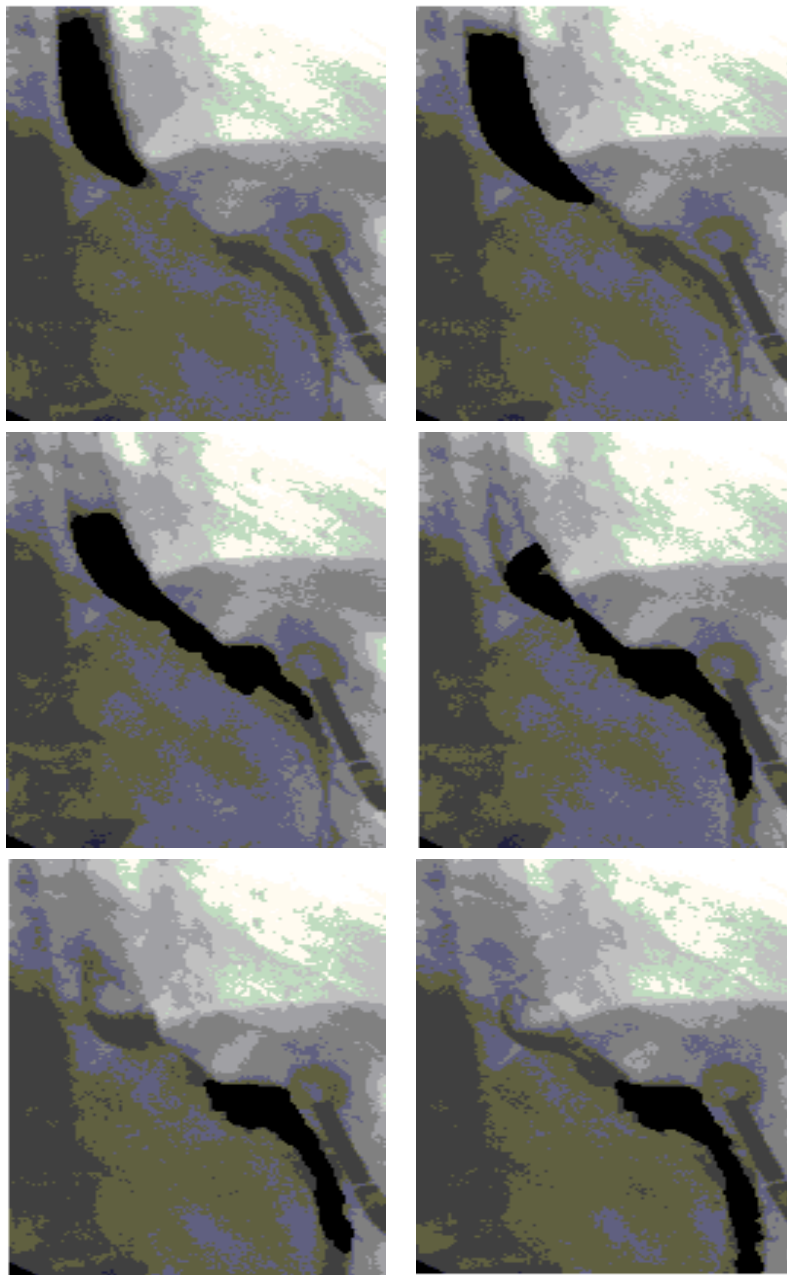


FIG. 3.13 – Images extraites du suivi du bolus pour la première séquence

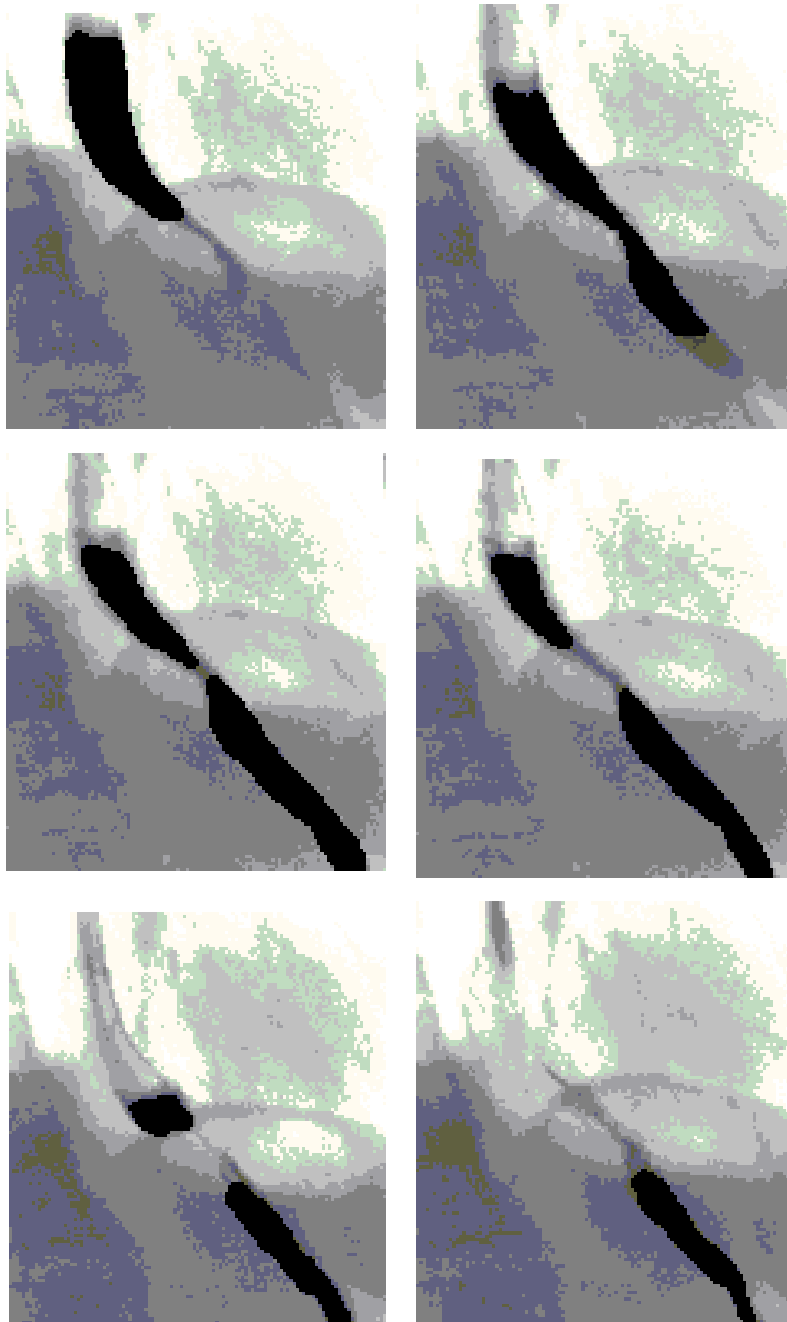


FIG. 3.14 – Images extraites du suivi du bolus mettant en évidence la gestion du changement de topologie de l'objet suivi.

3.5 Conclusion

Les résultats ont montré que la méthode de suivi par contours actifs non régularisés permet une bonne segmentation du bolus baryté lors de son passage à travers l'œsophage et le SIO. De plus, la gestion du changement de topologie peut permettre une étude plus précise des propriétés du SIO responsable de ce phénomène. Enfin, l'absence de réglage et l'association de la méthode de suivi à la méthode de détection de la première image de la séquence dans laquelle apparaît le bolus, présentée à la section 2.4.1 permet l'automatisation du suivi du bolus baryté.

Grâce au suivi du bolus, les caractéristiques de certains organes tels que le SIO vont pouvoir être déterminés. C'est l'objet du chapitre suivant.

Chapitre 4

Caractérisation des objets suivis

4.1 Introduction

Le suivi d'objets dans des séquences d'images n'est pas un objectif en soi, son intérêt étant de permettre l'étude de l'évolution des propriétés d'un objet au cours du temps. Les applications sont nombreuses notamment en physique par exemple pour l'étude du comportement de matériaux subissant certaines contraintes, et en médecine pour l'étude de l'évolution d'organes tels que le coeur.

Ainsi, les méthodes de suivi que nous avons développées et présentées dans les sections précédentes ont été appliquées au suivi d'un bolus baryté dans des images radiologiques lors de son passage à travers l'œsophage et le sphincter inférieur de l'œsophage (SIO). Le but de ce suivi est de permettre, de façon indirecte, la mesure de paramètres caractérisant les différents organes pouvant intervenir dans la pathologie du reflux gastro-œsophagien. Notre objectif est donc de localiser des zones intéressantes au sein des images, telles que l'œsophage dans sa globalité, le coude de l'œsophage et le SIO. De plus, grâce à la localisation de ces régions, le comportement du bolus peut être analysé afin de déterminer, de manière indirecte, le comportement de l'organe situé dans la zone correspondante. Pour cela, un ensemble de paramètres a été défini et sera présenté dans les sections suivantes.

4.2 Extraction de primitives

Briand dans [BRI91] distingue deux types de paramètres, qu'il applique à la caractérisation de défauts dans des radiographies industrielles. Tout d'abord, des paramètres de taille vont apporter une information sur les dimensions de la région. Ensuite, des paramètres de forme vont être définis afin de décrire la forme. Ces paramètres sont extraits soit directement à partir de la région segmentée, soit à partir du squelette de la région.

La méthode de segmentation ainsi que l'extraction des paramètres vont maintenant être exposées.

4.2.1 Extraction de paramètres de taille et d'orientation

L'*aire* de la région est obtenue en comptant le nombre de pixels qui la composent.

Le *périmètre*, quant à lui, est obtenu en comptant le nombre de pixels formant le contour de la région. Ce dernier est obtenu par soustraction entre la région et érodé par un élément structurant correspondant à un disque de rayon 1 pixel.

L'*orientation* de la région est obtenue par la mesure des angles formés par l'axe vertical et les axes d'inertie de la région.

Ces axes d'inertie correspondent aux axes de l'ellipse d'inertie de la région. Ils sont obtenus par le calcul des vecteurs propres associés aux valeurs propres de la matrice de variance-covariance A , exprimée ci-dessous :

$$A = \begin{pmatrix} Var_x & Cov_{xy} \\ Cov_{xy} & Var_y \end{pmatrix} \quad (4.1)$$

4.2.2 Extraction de paramètres issus de la squelettisation de la région

Principe de la squelettisation

La squelettisation est une approche très intéressante pour étudier un objet [ATT95, FV96]. En effet, elle apporte des informations sur sa forme globale, son contour, sa taille et son orientation.

Une présentation du squelette ainsi qu'un état de l'art sur les méthodes de squelettisation sont présentés dans [MAL98a]. Il existe plusieurs définitions du squelette. Il peut être introduit par le concept de *feu de prairie* : si l'on considère qu'un feu se propage à l'intérieur de l'objet à vitesse constante, à partir de chaque point du contour, le squelette peut être défini comme le lieu où ces fronts s'interceptent (les points d'extinction). Il peut également être défini comme l'ensemble des centres des boules maximales contenues dans l'objet.

Ainsi, le squelette est un sous-ensemble très fin de l'objet (de largeur 1 pixel), situé en son "centre". Il aura l'apparence d'un ensemble de lignes. Cette structure apporte l'information sur la forme de l'objet, comme nous le verrons dans les sections suivantes.

Le processus de squelettisation ayant pour but d'apporter une bonne description de l'objet, il doit satisfaire aux conditions suivantes :

- être homotopique : il doit préserver le nombre de composantes connexes de l'objet,
- assurer une bonne localisation du squelette.

Les méthodes de squelettisation peuvent être classées en quatre catégories [MAL98a] :

- algorithmes d'amincissement morphologique,
- algorithmes calculant le squelette d'un objet à partir de sa carte de distance,
- algorithmes basés sur la simulation de la propagation du feu de prairie,
- algorithmes calculant le squelette à partir du diagramme de Voronoï de l'objet.

L'algorithme que nous avons utilisé est un algorithme d'amincissement, présenté dans [COC95].

Algorithme de squelettisation

Soit X l'objet considéré et S un élément structurant formé de deux parties disjointes S^0 et S^1 . L'image est supposée binaire et les deux niveaux sont notés 0 et 1.

On note S_x^0 (respectivement S_x^1) le translaté de S^0 (respectivement S^1) par x .

On note $X \otimes S$ l'ensemble des pixels x pour lesquels S translaté par x vérifie :

- $\forall x_1 \in S_x^1, x_1 \in X$,

$$- \forall x_0 \in S_x^0, x_0 \notin X.$$

L'amincissement est la transformation, notée \circ , qui consiste à supprimer d'un ensemble X l'ensemble $X \otimes S$:

$$X \circ S = X \setminus X \otimes S$$

où \setminus représente la différence ensembliste.

L'élément structurant L défini ci-dessous, conduit à un amincissement homotopique. Il pourra donc être utilisé pour une squelettisation. Les pixels appartenant à S^0 sont notés 0, et ceux appartenant à S^1 sont notés 1. L'état des pixels notés * n'est pas pris en compte.

$$L = \begin{array}{ccc} 1 & 1 & 1 \\ * & 1 & * \\ 0 & 0 & 0 \end{array}$$

L'amincissement est itéré avec l'élément structurant L et les sept autres configurations obtenues par rotation. Après un certain nombre d'itérations, plus aucune modification n'est observée. L'objet obtenu est un squelette de X .

Le processus de squelettisation est sensible aux irrégularités du contour. En effet, des branches parasites du squelette, appelées *barbules*, peuvent apparaître. Pour remédier à ce problème, on peut envisager une régularisation du contour avant transformation (par une fermeture morphologique par exemple), et/ou un amincissement du squelette à partir de l'élément structurant E :

$$E = \begin{array}{ccc} * & * & * \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{array}$$

Après avoir déterminé et ébarbulé le squelette de la région, il est utile d'ordonner les points qui le composent. Cet ordre est obtenu en considérant le squelette comme une chaîne liant tous les points qui le composent, le premier de la liste étant le point dont la coordonnée verticale est la plus petite (le plus haut dans l'image).

Un certain nombre de paramètres peut alors être calculé.

Longueur de la région

La longueur de la région est obtenue en ajoutant au nombre de points constituant le squelette, la distance entre les points extrêmes du squelette et les points du contour les plus proches. Les deux points extrêmes correspondent aux points du squelette le plus haut dans l'image, et le plus bas.

La distance du squelette au contour

La distance du squelette au contour est obtenue en associant à chaque point du squelette, sa distance au point du contour le plus proche.

Cela permet d'étudier l'évolution du diamètre de la région le long du squelette. La fonction représentant cette évolution est tracée dans un repère dont les abscisses indiquent l'indice des points ordonnés du squelette, et les ordonnées, le diamètre de la région en ces points.

Cette fonction est extrêmement utile car elle permet de déterminer le diamètre sur l'ensemble de la région segmentée.

Evolution de la coordonnée horizontale des points du squelette

Il est également intéressant d'étudier l'évolution de la coordonnée horizontale des points le long du squelette. La courbe correspondante indique la relation entre l'indice des points ordonnés du squelette, et la coordonnée horizontale de ces points.

Evolution de la coordonnée verticale des points du squelette

De même, l'évolution de la coordonnée verticale des points le long du squelette est une information intéressante. La courbe correspondante traduit la relation entre l'indice des points ordonnés du squelette, et la coordonnée verticale de ces points.

4.2.3 Détermination de la trace des objets

Les paramètres présentés précédemment peuvent être calculés directement à partir de l'objet segmenté dans chaque image de la séquence, ou sur une région globale, que nous appelons trace de l'objet dans la séquence d'images. Cette trace permet

de déterminer le lieu de déplacement global de l'objet au cours du temps. Elle est obtenue en utilisant l'information de position de l'objet à chaque instant et en superposant cette information pour l'ensemble des images. Cela est réalisé par application de l'opérateur booléen *ou inclusif* sur l'ensemble des images de la séquence segmentées (les pixels appartenant à l'objet ont pour valeur 1, et ceux appartenant au fond de l'image ont pour valeur 0).

4.2.4 Extraction de primitives de vitesse

Une information de vitesse peut également être extraite. En effet, connaissant la position de points caractéristiques de l'objet, il est possible de calculer leur vitesse de déplacement en prenant en considération la fréquence d'acquisition des images.

4.3 Application à la caractérisation du SIO

Les paramètres présentés dans les sections précédentes ont été définis de manière à permettre la caractérisation des différents organes pouvant intervenir dans la pathologie du reflux gastro-œsophagien, grâce au suivi d'un bolus baryté dans des séquences d'images radiologiques (section 1.2).

En effet, le suivi d'objets déformables permet la segmentation du bolus lors de son passage dans l'œsophage et dans le SIO, au cours de la séquence. Les paramètres définis dans la section précédente donnent donc la possibilité de quantifier et de caractériser ces régions. Notre objectif est de localiser des zones intéressantes au sein des images, telles que l'œsophage dans sa globalité, le coude de l'œsophage et le SIO. De plus, grâce à la localisation de ces régions, le comportement du bolus peut être analysé afin de déterminer, de manière indirecte, le comportement de l'organe situé dans la zone correspondante.

La localisation de ces zones d'intérêt, ainsi que la caractérisation du SIO vont être présentées dans les sections suivantes.

4.3.1 Détection de la position du coude de l'œsophage

La trace du bolus est utilisée afin d'obtenir la segmentation de l'ensemble *œsophage + SIO*, grâce à l'utilisation de la segmentation du bolus baryté dans chaque

image de la séquence.

Lorsque le patient subit une exploration radiologique, il est orienté légèrement de biais par rapport à la source de rayons X. Le bolus baryté qui est conduit par l'œsophage, présente donc un mouvement vertical vers le bas dans un premier temps. Ensuite, peu avant son entrée dans le SIO, le bolus change de direction, et se dirige vers le bas à droite. Cette modification de la direction de déplacement du bolus entraîne une variation des courbes représentant l'évolution des coordonnées horizontales et verticales des points ordonnés du squelette de l'ensemble *œsophage + SIO*. Il est ainsi possible de détecter le "coude" de l'œsophage situé avant la zone du SIO.

Afin d'illustrer cela, les squelettes des ensembles *œsophage + SIO* obtenus à partir des résultats du suivi du bolus baryté réalisé sur les deux séquences présentées au chapitre 2 ont été déterminés.

La figure 4.1 présente la segmentation de ces ensembles. (a) et (b) correspondent respectivement aux séquences présentées figures 2.6 et 2.7.

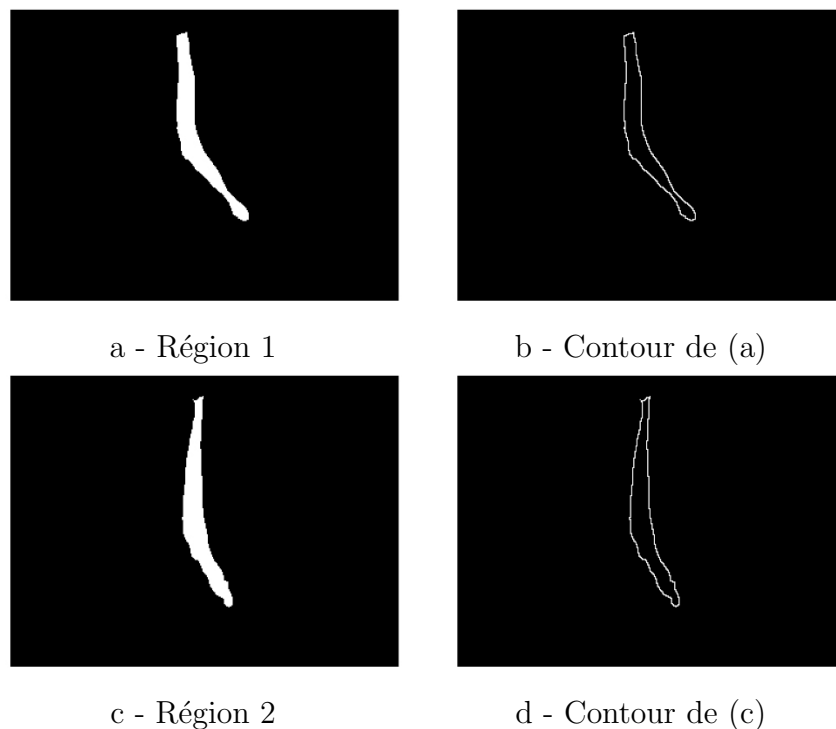


FIG. 4.1 – Segmentation de l'œsophage et du SIO

Les squelettes de ces deux régions sont présentés figure 4.2.

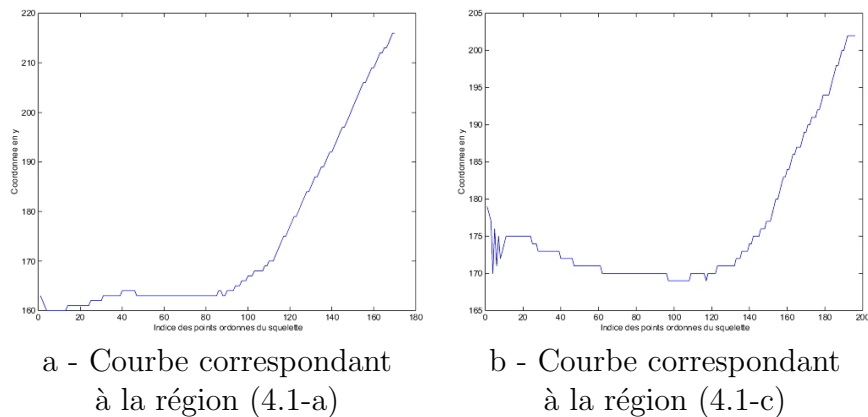
Les courbes représentant l'évolution de la coordonnée horizontale le long des



a - Squelette de la région (4.1-a) b - Squelette de la région (4.1-c)

FIG. 4.2 – Squelettisation des régions

squelettes sont présentées figure 4.3.



a - Courbe correspondant
à la région (4.1-a)

b - Courbe correspondant
à la région (4.1-c)

FIG. 4.3 – Evolution de la coordonnée horizontale des points le long des squelettes

A partir de ces courbes, il est donc possible de déterminer le coude de l'œsophage, correspondant à l'augmentation brutale de la pente.

4.3.2 Détection et caractérisation du SIO

Le SIO correspond à une zone de resserrement par rapport à l'œsophage. Ainsi, la courbe représentant l'évolution du diamètre de la région segmentée (œsophage et SIO) le long de son squelette comporte une diminution du diamètre de la région au niveau du SIO. En détectant cette variation de la courbe, il est possible de déterminer la position du SIO dans l'image. Cela est illustré figure 4.4. De plus, le diamètre de la région au niveau du SIO est une information importante, correspondant au diamètre maximal du SIO.

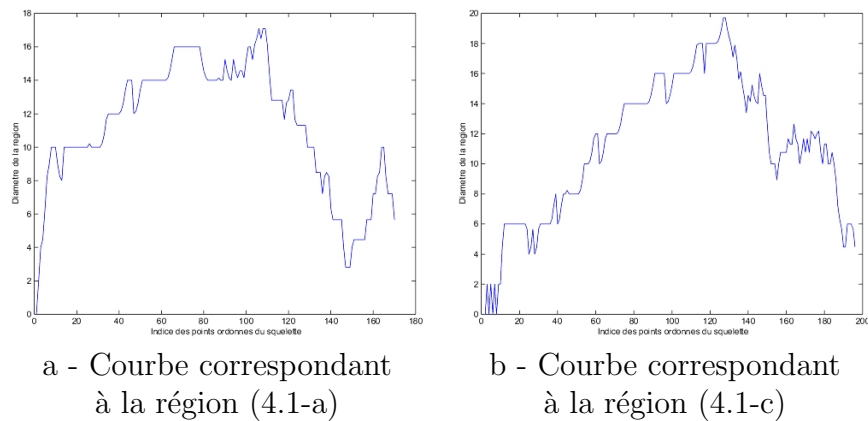


FIG. 4.4 – Evolution du diamètre de la région le long des squelettes

Une autre méthode permettant la détection de la zone correspondant au SIO consiste à étudier l'évolution de la vitesse de déplacement du bolus, celui-ci étant freiné et parfois arrêté à l'entrée du SIO, avant de subir une accélération importante au moment de son ouverture.

Ayant déterminé la zone correspondant au SIO dans l'image, il est alors possible de calculer les paramètres de taille et d'orientation du bolus, présentés précédemment, lors de son passage dans cette zone, afin de caractériser cet organe de manière indirecte.

4.4 Conclusion

Nous avons présenté dans cette partie un certain nombre de paramètres permettant de détecter de manière automatique des zones d'intérêt dans les images telles que les zones correspondant au coude de l'œsophage et au SIO. Il est alors possible de d'étudier ces organes de manière indirecte en déterminant les caractéristiques du bolus lors de son passage dans ces zones, ce qui permet en particulier, grâce à l'information de vitesse, d'apporter une information relative à la motricité du SIO.

Ces différents paramètres pourront ensuite être calculés de manière automatique en situation réelle de manière à assister l'expert lors de son diagnostic, en lui donnant accès à des informations qu'il n'aurait pu obtenir de manière aisée en analysant manuellement les séquences.

Toutes les mesures de distances que nous avons réalisées sont exprimées en

nombre de pixels. En connaissant le pas de numérisation, il est possible de les convertir en unités métriques, plus significatives. Une manière de déterminer ce pas consisterait en l'utilisation de l'information apportée par le microphone placé au niveau du SIO et présent sur toutes les images. En effet, ce microphone est rond, et son diamètre réel parfaitement connu. Il apparaît dans les images sous forme d'ellipse du fait de son inclinaison par rapport au plan de la caméra. En mesurant le plus grand axe (en nombre de pixels), et en réalisant le rapport avec le diamètre réel, nous obtenons le pas de numérisation.

Deuxième partie

Analyse de signaux sonores

Chapitre 5

Identification de signaux sonores

5.1 Introduction

L'identification du contenu de signaux sonores est un large domaine de recherche du fait des nombreuses difficultés à surmonter, et des nombreuses applications possibles. Une présentation des systèmes assurant cette fonction peut être trouvée dans [GER00a, GER00b, FOO99]. L'analyse de la parole est certainement un des domaines les plus prolifiques [AO97]. L'objectif est dans ce cas de reconnaître les mots prononcés [CHA99, THO02], et/ou d'identifier le locuteur [CAM97, CHA97, PAT97]. Un autre domaine important est l'indexation de documents sonores [PFE96, ZHA98b, FOO97] dont le but est de déterminer le contenu des signaux sonores permettant la recherche automatique d'information dans des bases de données importantes. Une approche intéressante est la discrimination musique/parole [JAR02, SCH97] ou encore la discrimination musique/parole/sons environnementaux/silence [LU01, ZHA98a]. Une autre application consiste en l'analyse de scènes audio [LIU97] où les auteurs distinguent quatre types de scènes : publicités, commentaires de matchs de basketball et de football, journaux d'actualité et présentations météorologiques. Dans [PEL02], des scènes d'extérieur (rues, parcs, ...) et d'intérieur (restaurants, maison, ...) sont discriminées. L'analyse de scènes musicales [GER98] a pour but l'identification du style musical [TZA01, TZA03] et du type d'instruments [DIX00, HEI02]. De plus, une application intéressante consiste en la transcription musicale automatique [KLA98, MAR96]. L'identification de signaux sonores médicaux est également un problème intéressant dans un contexte médical [DEL02b, DEL02a, DEL03d, DEL03h] pouvant apporter une aide au diagnostic et à la prise de décision.

Enfin, l'analyse du contenu de signaux sonores est de plus en plus utilisée pour l'indexation de documents vidéo [ROA01, PAT96, FIS95].

Les ordinateurs sont donc d'une grande aide pour les humains, leur permettant l'analyse de signaux de plus en plus nombreux dans différents domaines. Cependant, si la tâche paraît naturelle et d'une facilité apparente pour les humains, elle nécessite l'analyse de la méthodologie implicite afin de développer des systèmes d'identification automatiques. La classification de signaux sonores (ASC pour Audio Signal Classification) [GER00a, GER00b] est un domaine particulier de tels systèmes, destinée à l'analyse de signaux sonores.

Généralement, le principe d'un système ASC est d'extraire des primitives à partir de signaux sonores, et ensuite d'identifier la classe à laquelle ils ont la plus grande probabilité d'appartenir.

Le choix des primitives est très important parce que ce sont elles qui contiendront l'information utilisée pour la classification.

Les primitives sont généralement divisées en deux catégories [GER00a] : les primitives de type physique et les primitives de type perceptuel. Les primitives de type physique sont basées sur les propriétés statistiques et mathématiques des signaux, alors que les primitives de type perceptuel sont liées à la manière dont les humains perçoivent les sons.

Les primitives de type physique peuvent comporter notamment l'énergie qui est une mesure de l'importance de l'amplitude d'un signal, le *Zero-Crossing-Rate* qui est une mesure du nombre de passages par 0 du signal par unité de temps, ainsi que des primitives spectrales qui indiquent le contenu fréquentiel des signaux, telles que la fréquence fondamentale, grâce à la transformée de Fourier, ou aux transformées en ondelettes.

On peut souvent lier les primitives de type perceptuel aux primitives de type physique. Le *pitch* par exemple est fortement lié à la fréquence fondamentale. L'évolution temporelle du *pitch* est représentée par le *prosody*. Le timbre est une autre primitive importante qui est définie par la qualité du son permettant par exemple la distinction entre les instruments et la voix ayant le même *pitch*. Enfin, le rythme caractérise les événements perceptibles individuellement dans le son, qui apparaissent de manière répétitive et prévisible.

De plus, il est possible de décomposer la structure d'un signal sonore en trois couches [LIU97] : le niveau bas contenant les caractéristiques acoustiques, le niveau

intermédiaire contenant la signature auditive associée aux différents sons des objets, et le niveau haut contenant les modèles sémantiques du signal des différentes classes scéniques.

La couche des caractéristiques acoustiques correspond à des particularités générales telles que l'énergie du signal, la période du "pitch" ou encore la largeur de bande du signal sonore.

La couche des signatures acoustiques associées aux objets a pour rôle de déterminer l'objet qui produit un son particulier. En effet, les sons produits par des objets différents ont des signatures différentes. Par exemple, chaque instrument de musique a sa propre "réponse impulsive". Ainsi, en stockant ces signatures dans une base de données et en les comparant avec un signal à classer, il est possible de déterminer la classe à laquelle appartient ce signal.

La couche de niveau haut repose sur l'utilisation de règles sémantiques connues concernant la structure des signaux sonores dans différents types de scènes. Par exemple, un reportage d'information ne contient que de la parole, alors que dans une annonce commerciale, la parole est mélangée à une musique de fond. Les commentaires d'un match de tennis, quant à eux seront composés notamment de parole, de bruits de foule et de chocs de balle. A partir de l'extraction des différents composants d'un signal sonore, il est alors possible de déterminer la nature de la scène.

En général, lorsque les caractéristiques acoustiques associées à chaque signal ne sont pas suffisantes pour assurer une bonne reconnaissance, les primitives considérées pourront alors être associées à des primitives de niveau supérieur.

Une fois les primitives extraites et structurées pour former un unique vecteur entité, une méthode de sélection de primitives peut être utilisée afin de réduire la dimension de ce vecteur. Cela permet d'améliorer l'efficacité de la phase suivante de classification dont le rôle est de déterminer la classe à laquelle le vecteur de primitives a la plus grande probabilité d'appartenir.

De nombreux algorithmes de classification peuvent être appliqués [JAI88], en fonction du type de données à classer et des connaissances relatives à ces données. Deux grandes familles de méthodes de classification peuvent être distinguées : les méthodes supervisées lorsque les classes sont connues a priori (réseaux de neurones, k -means, chaînes de Markov cachées, ...), et les méthodes non supervisées lorsque les classes sont découvertes au fur et à mesure de l'évolution de l'algorithme (classification ascendante hiérarchique, cartes neuronales auto-organisées, ...).

Les signaux analysés étant généralement de nature complexe, une des difficultés rencontrées pour le problème d'identification est l'extraction de l'information pertinente, sous forme de primitives, à partir des signaux. Nous rappellerons donc dans les parties suivantes différentes analyses souvent utilisées dans les systèmes ASC. Nous avons choisi, en raison du manque d'information sur les caractéristiques des sons que nous voulons analyser selon différentes classes, de recourir à des primitives multiples permettant de mettre en valeur des informations de nature différente. Elles concernent l'étude de la couche de niveau bas des signaux, une analyse *Cepstrale* du signal et une étude des signaux par modélisation *ARMA*. Enfin deux ensembles de primitives que nous avons développés seront décrits. Le premier est lié à une analyse en ondelettes des signaux, et le second à une analyse suivant les lois de *Zipf* et *Zipf Inverse*.

La méthode de classification utilisée sera ensuite décrite. Elle est composée de cinq méthodes de classification différentes, ainsi qu'une méthode de fusion de classifieurs permettant de prendre en considération les résultats de chacun des algorithmes de classification.

Enfin, l'application de ces méthodes à des signaux médicaux sera présentée dans une dernière partie. Ces signaux correspondent aux bruits xiphoidiens produits lors du passage d'un bolus alimentaire à travers l'œsophage et le sphincter inférieur de l'œsophage.

5.2 Etude de la structure bas-niveau des signaux

Les primitives bas-niveau sont souvent utilisées pour la segmentation de signaux sonores [LIU97]. Afin d'étudier l'évolution des propriétés des signaux au cours du temps, un découpage des signaux en fenêtres chevauchantes de longueur N est réalisé.

Afin de mesurer les variations temporelles de l'amplitude des signaux, le calcul de la distribution de l'énergie des signaux est réalisé. L'énergie de la fenêtre n est donnée par la formule (5.1).

$$E(n) = \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} s_n^2(i)} \quad (5.1)$$

où $s_n(i)$ est l'amplitude de l'échantillon i de la fenêtre n de taille N .

A partir de ces valeurs, le *ESTD* (Energy Standard Deviation) est utilisé, et correspond à l'écart type de l'énergie du signal, normalisé par l'énergie maximum sur l'ensemble du signal, ainsi que le *EDR* (Energy Dynamic Range) défini par la formule (5.2).

$$EDR = (\max(E) - \min(E)) / \max(E) \quad (5.2)$$

où $\max(E)$ et $\min(E)$ sont respectivement les valeurs maximale et minimale de l'énergie sur l'ensemble des fenêtres définies dans le signal.

La valeur moyenne *ZCRM* et l'écart-type *ZCRSD* du *ZCR* (Zero Crossing Rate) sont également calculés. Le *ZCR* représente le nombre de fois où l'enveloppe du signal sonore coupe l'axe des abscisses par unité de temps. Le *NSR* (Non Silence Ratio) quant à lui correspond au rapport du nombre de fenêtres considérées comme non silencieuses, sur le nombre total de fenêtres du signal. Les fenêtres sont silencieuses lorsque l'énergie et le *ZCR* sont tous deux inférieurs à un seuil.

Une autre caractéristique intéressante est le pitch qui est la période fondamentale du signal sonore. De nombreux algorithmes de calcul du pitch ont été développés. Celui que nous avons retenu, en raison de sa facilité d'utilisation et de sa fiabilité, repose sur l'*AMDF* (Average Magnitude Difference Function), [LIU97] défini pour chaque fenêtre n par la formule (5.3).

$$AMDF_n(l) = \frac{\sum_{i=0}^{N-l-1} |s_n(i+l) - s_n(i)|}{N-l} \quad \forall l \in [0, N] \quad (5.3)$$

Le principe de cet algorithme repose sur le fait que si le signal est périodique de période T , alors $s_n(i) = s_n(i+T)$ pour tout i . Ainsi, le pitch correspond à la distance entre l'origine et le premier minimum local de l'*AMDF*. Les primitives extraites à partir du pitch sont le *PSTD* (Pitch Standard Deviation) qui est l'écart-type du pitch sur l'ensemble du signal, le *SPR* (Smooth Pitch Ratio), ainsi que le *NPR* (None Pitch Ratio). Le *SPR* est défini comme le rapport entre le nombre de fenêtres ayant le même pitch que leur fenêtre précédente (la différence de valeur étant en deça d'un seuil) et le nombre total de fenêtres du signal. Cela permet d'évaluer les segments de pitch constants. Le *NPR* est défini comme le rapport entre le nombre de fenêtres sans pitch (le signal étant alors non périodique) et l'ensemble des fenêtres.

Plusieurs primitives relatives aux caractéristiques fréquentielles du signal sont

également calculées. Elles font appel à la transformée de Fourier à fenêtres glissantes [BRI88]. Il s'agit de la valeur moyenne FCM et de l'écart-type $FCSD$ de la fréquence du centroïde FC (5.4) de la fenêtre considérée.

$$FC(n) = \frac{\int_0^\infty \omega |T^{fen} s(\omega, n)(\omega)|^2 d\omega}{\int_0^\infty |T^{fen} s(\omega, n)(\omega)|^2 d\omega} \quad (5.4)$$

où $T^{fen} s(\omega, n)$ est la transformée à fenêtre glissante du signal s , calculée sur la fenêtre d'indice n .

A ces primitives fréquentielles, s'ajoutent la valeur moyenne BWM et l'écart-type $BWSD$ de la bande de fréquence BW (5.5).

$$BW^2(n) = \frac{\int_0^\infty (\omega - FC(n))^2 |T^{fen} s(\omega, n)(\omega)|^2 d\omega}{\int_0^\infty |T^{fen} s(\omega, n)(\omega)|^2 d\omega} \quad (5.5)$$

La distribution de l'énergie des différentes bandes de fréquences variant en fonction des signaux, les rapports d'énergie $SBERi$ (subband energy ratios) des sous-bandes d'indice i ont également été calculés. Ils correspondent au signal obtenu par application d'un filtre passe-bande, normalisé par l'énergie du signal d'origine. La définition de ces sous-bandes dépend des signaux et de l'information recherchée.

Pour chaque $SBER$, la valeur moyenne $SBERM$ et l'écart-type $SBERSD$ sont ensuite calculés.

Les primitives présentées précédemment ont pour objectif la description bas-niveau des signaux sonores. La section suivante est consacrée à la présentation de primitives issues d'une analyse cepstrale du signal, et liées de manière plus importante à la perception humaine des sons, raison pour laquelle cette méthode est couramment utilisée notamment en reconnaissance de la parole.

5.3 Etude des signaux par analyse cepstrale

L'analyse cepstrale est très fréquemment utilisée pour l'analyse et la reconnaissance de signaux sonores [LI01], et plus particulièrement de signaux de parole [MAL03, MOK96, THO02]. Dans [PAT99], les auteurs y ont également recours pour la modélisation de l'activité électrique musculaire.

L'intérêt de l'analyse cepstrale réside dans le fait qu'elle permet de déterminer la

fréquence fondamentale du signal, et d'isoler le signal propre, du bruit environnant.

Son principe repose sur l'hypothèse que le signal x mesurable correspond à la convolution du signal de la source u par une réponse impulsionnelle b correspondant au bruit :

$$x = u \otimes b \quad (5.6)$$

où \otimes est l'opérateur de convolution.

Le cepstre permet de séparer ces deux sources par déconvolution. La première étape consiste en la transformation de la convolution en produit, par application de la transformée en Z (T^Z) :

$$T^Z(x) = T^Z(u).T^Z(b) \quad (5.7)$$

Ce produit est ensuite transformé en somme par application du logarithme népérien. Seul le module de la transformée en Z est considéré, le rôle de l'information de phase étant négligé :

$$\text{Ln}(|T^Z(x)|) = \text{Ln}(|T^Z(u)|) + \text{Ln}(|T^Z(b)|) \quad (5.8)$$

Enfin, le cepstre est obtenu par application de la transformée en Z inverse. Puisque dans la pratique, la transformée en Z est remplacée par la transformée de Fourier rapide, le cepstre d'un signal x ($T^{\text{cepstre}}(x)$) peut donc être exprimé par la relation (5.9).

$$T^{\text{cepstre}}(x) = T^{\text{fft}^{-1}}(\text{Ln}(|T^{\text{fft}}(x)|)) \quad (5.9)$$

L'espace de représentation du cepstre est homogène au temps et il est possible, par un filtrage temporel, de séparer le signal de la source et la réponse impulsionnelle correspondant au bruit. Dans le cas de signaux de parole, cette réponse impulsionnelle est associée au conduit vocal. Les premiers coefficients cepstraux contiennent alors l'information relative au conduit. Cette contribution devient négligeable à partir d'un échantillon P . Les pics périodiques visibles au-delà de P , reflètent les impulsions de la source.

Ainsi, le cepstre fournit une méthode pour réduire la quantité d'information puisqu'en conservant les coefficients en deçà de P , il est possible de lisser la représentation spectrale du signal afin de ne conserver que l'enveloppe spectrale caractéristique du signal u . De plus, le cepstre permet de déterminer la fréquence fondamentale de

la source u en détectant les pics périodiques au-delà de P . A titre d'exemple, les ordres de grandeur des fréquences fondamentales pour la parole sont de 120 Hz pour les hommes, 250 Hz pour les femmes et 450 Hz pour les enfants.

Récemment, une extension de l'analyse cepstrale a été réalisée. Il s'agit de l'analyse *Mel-cepstrale* utilisée pour la modélisation de la parole [FOO97] et de la musique [LOG00]. La différence réside dans le fait que dans le cas de l'analyse *Mel-cepstrale*, l'application de la transformée inverse est précédée d'une phase de lissage du spectre et de mise en valeur des fréquences pertinentes par un recalage des fréquences dans une échelle logarithmique (échelle MEL), plus fidèle à la perception humaine.

Les deux ensembles de primitives présentés dans cette section et dans la section précédente sont utilisés afin de décrire certaines caractéristiques temporelles ou fréquentielles des signaux sonores. Dans la section suivante, une approche différente est envisagée, et consiste à modéliser les signaux par des modèles ARMA. Les signaux sonores seront alors représentés par les coefficients du modèle associé.

5.4 Etude des signaux sonores par modélisation ARMA

L'approche ARMA (Auto-Regressive, Moving Average) [BLA98] a pour but la modélisation de signaux stationnaires. Son champ d'application est très vaste : on peut citer par exemple l'estimation et la détection de signaux [EZZ91], la modélisation de la parole [VER99, YAS99], ainsi que la modélisation de phénomènes biologiques [PAT99].

Un processus ARMA correspond à la mise en série d'une structure AR et d'une structure MA, permettant la modélisation d'un signal x à partir du signal observable w . Ces structures vont être définies dans les sections suivantes.

5.4.1 Processus AR

Soit l'équation récurrente suivante :

$$\sum_{i=0}^{n_1} a_i x(t-i) = w(t) \quad (5.10)$$

où w est un signal centré stationnaire de bruit blanc, $\{a_i\}$ une suite de coefficients réels et n_1 l'ordre du modèle.

Le signal x peut être vu comme la sortie du filtre de fonction de transfert T définie par (5.11).

$$T(z) = \frac{1}{1 + \sum_{i=1}^{n_1} a_i z^{-i}} \quad (5.11)$$

On peut montrer que l'équation (5.10) admet une unique solution x s'exprimant causalement en fonction de w si et seulement si les pôles de T sont de module inférieur à 1. Ainsi, dans ce cas, la valeur $x(t)$ peut être déterminée par la connaissance des valeurs du signal d'entrée w à l'instant t et aux instants passés :

$$x(t) = \sum_{i=0}^{\infty} h_i w(t-i) \quad (5.12)$$

où la suite $\{h_i\}$ est la suite des coefficients du développement en série de Fourier de T .

5.4.2 Processus MA

Soit le signal x défini par :

$$x(t) = \sum_{i=0}^{n_2} b_i w(t-i) \quad (5.13)$$

où n_2 est l'ordre du modèle, w est un signal centré, de bruit blanc, et stationnaire et $\{b_i\}$ une suite de coefficients réels.

Le signal ainsi construit correspond à la moyenne des $n_2 + 1$ dernières valeurs du signal d'entrée w , pondérées par la suite $\{b_i\}$. Schématiquement, cela revient à déplacer cette suite de pondération le long du signal w , d'où le nom de *Moving Average* (moyenne mobile) donné à ce type de modèle.

De plus, x peut être vu comme la sortie d'un filtre linéaire dont la réponse impulsionnelle est la suite finie b_m , et dont la fonction de transfert est définie par (5.14).

$$H(z) = \sum_{i=0}^{n_2} b_i z^{-i} \quad (5.14)$$

5.4.3 Processus ARMA

Comme indiqué précédemment, les processus ARMA sont obtenus en mettant en série une structure AR et une structure MA.

Soit l'équation récurrente suivante :

$$\sum_{i=0}^{n_1} a_i x(t-i) = \sum_{i=0}^{n_2} b_i w(t-i) \quad (5.15)$$

où w est un signal centré stationnaire de bruit blanc, et $\{a_i\}$ et $\{b_i\}$ deux suites de coefficients réels.

On montre que cette équation admet une unique solution x s'exprimant causalement en fonction de w si et seulement si les pôles de la fonction S définie par (5.16) sont de module inférieur à 1.

$$S(z) = \frac{\sum_{i=0}^{n_2} b_i z^{-i}}{1 + \sum_{j=1}^{n_1} a_j z^{-j}} \quad (5.16)$$

Un processus ARMA est par définition la solution stationnaire d'une équation récurrente de la forme (5.15).

Les primitives retenues à partir de la modélisation ARMA dans le but de représenter les signaux sonores sont donc les coefficients du modèle d'ordre n_2 pour la structure MA et n_1 pour la structure AR, ce qui correspond à $(n_1 + n_2 + 2)$ coefficients.

Les modèles ARMA, comme indiqué précédemment, permettent la modélisation de signaux stationnaires. Or, cela est rarement le cas des signaux naturels. Afin de résoudre cette difficulté, l'hypothèse de stationnarité est réalisée, lorsque cela est possible, sur des intervalles temporels de faible durée des signaux sonores. La méthode de caractérisation de signaux sonores présentée dans la section suivante est au contraire, particulièrement bien adaptée à l'étude de signaux au comportement non stationnaire présentant de fortes discontinuités. Il s'agit de l'analyse en ondelettes.

5.5 Etude des signaux par analyse en ondelettes

Le champ d'application des ondelettes est vaste : en mathématique (analyse, probabilités, fractales, ...), en physique (mécanique quantique, turbulence, ...), et dans le domaine du traitement du signal (compression, astronomie, études sismiques, ...) [TRU98] où cette transformée est très appréciée car elle permet d'étudier les fréquences d'une fonction tout en conservant la localisation des événements.

5.5.1 Définition des ondelettes

Les ondelettes forment une représentation des fonctions de L^2 dans une base dont les fonctions sont bien localisées en espace et en fréquence. Elles conservent donc les avantages de la transformée de Fourier sans les inconvénients liés au manque de localisation.

L'idée de la transformée en ondelettes est due à J. Morlet (1981) qui, pour une étude haute résolution de signaux sismiques, a proposé une transformation à fenêtres de tailles variables.

L'analyse par ondelettes est ainsi une technique de fenêtrage avec des régions de taille variable. Elle permet l'étude :

- De longs intervalles de temps sur lesquels on souhaite avoir plus d'information sur les basses fréquences,
- De régions plus réduites sur lesquelles on souhaite avoir plus d'information sur les hautes fréquences.

Après avoir rappelé les avantages mais aussi les inconvénients de ce que l'on pourrait considérer comme l'ancêtre des transformées en ondelettes, l'analyse de Fourier à fenêtre glissante (ou transformée de Gabor), nous préciserons l'approche de J. Morlet avant de rappeler la définition et les propriétés des différents types de transformées en ondelettes utilisées à l'heure actuelle.

La transformée de Gabor

La transformée de Fourier classique (5.17) permet de passer de l'espace de représentation usuel d'une fonction temporelle à un espace fréquentiel.

$$T^{fourier}x(\omega) = \int_{-\infty}^{+\infty} x(t) \exp^{-i\omega t} dt \quad (5.17)$$

où $T^{Fourier}x$ est la transformée de Fourier de la fonction x . La pulsation $\omega = 2\pi f$, f étant la fréquence, et le temps t varie dans \mathbb{R} .

Un schéma numérique de calcul rapide de la transformée de Fourier a été développé pour des valeurs discrètes de temps n et de fréquence k : la transformée de Fourier Rapide ou FFT $T^{fft}x$ [BRI88] d'une fonction échantillonnée x est donnée par la relation (5.18).

$$T^{fft}x(k) = \sum_{n=0}^{N-1} x(n)W_N^{kn} \quad (5.18)$$

où $W_N = \exp^{-2i\pi/N}$ et N est le nombre total d'échantillons de la fonction x .

Le recours à l'analyse de Fourier suppose qu'un choix a été fait entre la variable temps et la variable fréquence, mais "nos expériences quotidiennes - notamment nos sensations auditives - imposent une description en termes de temps et de fréquence" a écrit Gabor. Pour analyser les signaux en temps et en fréquence, il a montré comment utiliser la transformée de Fourier à fenêtre glissante (ou transformée de Gabor).

L'idée de cette transformée consiste à décomposer le signal à analyser en fréquence sur chaque intervalle de temps, ce qui limite la plage de temps analysée. Cette transformation vise donc à améliorer la localisation en temps, en préservant au mieux la localisation en fréquence.

Cette transformation consiste à comparer un signal à des morceaux de courbes oscillantes de différentes fréquences, alors que la transformée de Fourier classique compare le signal entier à des sinusoides infinies de diverses fréquences. Quand un segment temporel a été analysé, la fenêtre est glissée de manière à analyser une autre portion (5.19).

$$T^{fen}x(\omega, b) = \int_{-\infty}^{+\infty} x(s)g(s-b)e^{-i\omega s} ds \quad (5.19)$$

où x est la fonction à analyser et g est une fonction porte appropriée centrée en b .

En notant $\Psi_{b,\omega}(s) = g(s-b)e^{i\omega s}$, on peut interpréter $T^{fen}x(\omega, b)$ comme la projection de x sur la base des fonctions fenêtres glissantes Ψ au sens du produit

scalaire hermitien usuel (5.20).

$$T^{fen}x(\omega, b) = \langle x, \Psi_{b,\omega} \rangle \quad (5.20)$$

La fonction g est généralement choisie telle qu'elle soit à support compact, suffisamment régulière et bien localisée en espace et en fréquence. Une fonction possible est la gaussienne $h(t) = Ce^{-\alpha t^2}$, C et α étant des réels positifs. Cette fonction est à support non compact mais à décroissance rapide.

La résolution dans le plan temps-fréquence de la transformation peut être estimée par la variance de la fonction analysante dans l'espace temporel σ_t (5.21) et dans l'espace fréquentiel σ_f (5.22).

$$\sigma_t = \int_{-\infty}^{+\infty} t^2 |\Psi(t)|^2 dt \quad (5.21)$$

$$\sigma_f = \int_{-\infty}^{+\infty} f^2 |\hat{\Psi}(f)|^2 df \quad (5.22)$$

Dans le cas de la transformée de Gabor, les résolutions temporelle et fréquentielle sont constantes, ce qui se traduit par un pavage temps-fréquence régulier (figure 5.1).

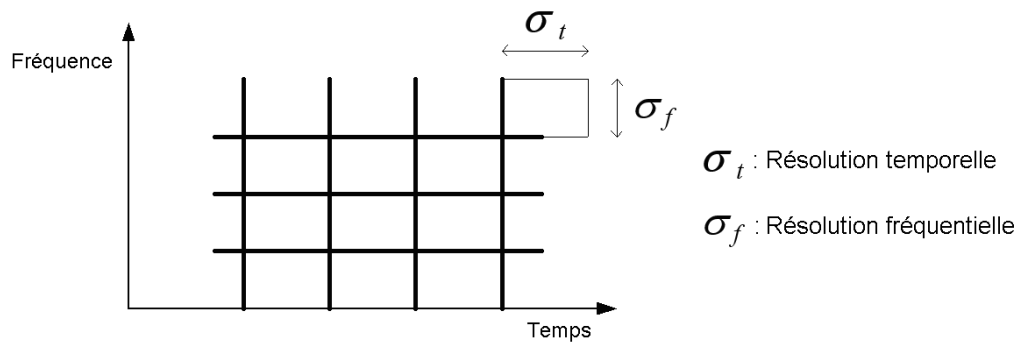


FIG. 5.1 – Pavage temps-fréquence pour la transformée à fenêtre glissante

La transformée de Fourier à fenêtre glissante permet donc une représentation temps-fréquence ou espace-fréquence. Cependant la taille fixe de la fenêtre n'est pas adaptée à la représentation de fonctions ou de signaux ayant des composantes de tailles très différentes de celle de la fenêtre. Ainsi lorsque la fenêtre est étroite, il est possible de localiser les changements soudains (pics, discontinuités, ...), mais les basses fréquences du signal, de période trop grande pour entrer dans la petite

fenêtre, ne peuvent pas être mises en évidence. À l'inverse lorsque la fenêtre est large, il n'est pas possible de préciser l'instant d'un pic ou d'une discontinuité, cette information étant noyée dans la totalité de l'information correspondant à l'intervalle de temps sélectionné par la fenêtre.

De plus, contrairement à la transformation de Fourier classique, cette transformation ne permet aucune méthode numérique de reconstruction du signal à partir de la transformée.

Les ondelettes de Morlet

Jean Morlet s'est inspiré de la transformée de Gabor, mais au lieu de garder fixe la taille de la fenêtre et de faire varier le nombre d'oscillations, il fait l'inverse, c'est-à-dire qu'il garde constant le nombre d'oscillations et fait varier la taille de la fenêtre (grâce à un paramètre d'échelle), l'étirant ou la comprimant :

- Étirer la fenêtre, étire les oscillations donc baisse leur fréquence,
- Comprimer la fenêtre, comprime les oscillations donc produit des fréquences plus élevées.

Ainsi, il est possible de localiser les hautes fréquences avec les petites fenêtres, et d'étudier les basses fréquences avec les fenêtres plus larges.

Jean Morlet et Alex Grossman ont montré que l'énergie (i.e. la valeur moyenne du carré de l'amplitude) d'un signal ne change pas quand on lui applique une transformation en ondelettes. On peut donc transformer un signal en ondelettes, puis reconstruire exactement le même signal à partir des ondelettes. De plus cette transformation est robuste au sens où un petit changement de la représentation en ondelettes induit un changement de taille comparable dans le signal.

La notion de représentation temps-échelle ou espace-échelle permet de voir la transformée en ondelettes comme un microscope mathématique, le zoom étant le facteur d'échelle et l'optique l'ondelette.

L'ondelette de Morlet est décrite par la relation (5.23).

$$\Psi(s) = \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} e^{-i\omega_0 s} \quad (5.23)$$

Généralités sur les ondelettes

Par définition, la fonction de deux variables décrite par (5.24) est une représentation temps-échelle appelée "transformation en ondelettes" de la fonction f .

$$T^{ond} f(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} f(t) \Psi \left(\frac{t-b}{a} \right) dt \quad (5.24)$$

où a est le facteur d'échelle ($a \neq 0$), b est le paramètre de translation, f la fonction transformée, et Ψ une fonction appelée ondelette mère.

a correspond à l'inverse de la fréquence : plus a est petit, moins l'ondelette (la fonction analysante) est étendue temporellement, donc plus la fréquence centrale de son spectre est élevée.

Il est possible d'interpréter l'équation (5.24) comme la projection du signal sur une famille de fonctions analysantes $\Psi_{a,b}$ construite à partir d'une fonction mère Ψ (5.25).

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \Psi \left(\frac{t-b}{a} \right) \quad (5.25)$$

Cette projection est décrite par la relation (5.26).

$$T^{ond} f(a, b) = \langle f, \Psi_{a,b} \rangle \quad (5.26)$$

De plus, la norme est invariante au changement d'échelle (5.27).

$$\|\Psi_{a,b}\|^2 = \int_{-\infty}^{+\infty} \frac{1}{|a|} \left| \Psi \left(\frac{t-b}{a} \right) \right|^2 dt = \frac{1}{|a|} \int_{-\infty}^{+\infty} |\Psi(x)|^2 |a| dx = \|\Psi\|^2 \quad (5.27)$$

Sous l'hypothèse $\int_{-\infty}^{+\infty} x^2 |\Psi(x)|^2 dx = 1$, la résolution temporelle σ_t est alors donnée par la relation (5.28).

$$\begin{aligned} \sigma_t^2 &= \int_{-\infty}^{+\infty} t^2 |\Psi_{a,0}(t)|^2 dt \\ &= \int_{-\infty}^{+\infty} t^2 \frac{1}{|a|} \left| \Psi \left(\frac{t}{a} \right) \right|^2 dt \\ &= \int_{-\infty}^{+\infty} a^2 x^2 \frac{1}{|a|} |\Psi(x)|^2 |a| dx \\ &= a^2 \int_{-\infty}^{+\infty} x^2 |\Psi(x)|^2 dx \\ &= a^2 \end{aligned} \quad (5.28)$$

L'occupation fréquentielle de l'ondelette est obtenue en calculant l'écart-type de la transformée de Fourier de $\Psi_{a,0}$, notée $\hat{\Psi}_{a,0}$, et en prenant comme unité l'écart-type de la transformée de Fourier de l'ondelette mère Ψ (5.29).

$$\sigma_{\omega}^2 = \int_{-\infty}^{+\infty} \omega^2 |\hat{\Psi}_{a,0}(\omega)|^2 d\omega = \int_{-\infty}^{+\infty} \omega^2 \frac{1}{|a|} |a\hat{\Psi}(a\omega)|^2 d\omega \quad (5.29)$$

De plus,

$$\hat{\Psi}_{a,0}(\omega) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{|a|}} \Psi\left(\frac{t}{a}\right) e^{-i\omega t} dt = \frac{a}{\sqrt{|a|}} \int_{-\infty}^{+\infty} \Psi(x) e^{-i\omega a x} dx = \sqrt{|a|} \hat{\Psi}(a\omega) \quad (5.30)$$

D'où finalement,

$$\sigma_{\omega}^2 = \int_{-\infty}^{+\infty} \frac{x^2}{a^2} \frac{1}{a} |a\hat{\Psi}(x)|^2 \frac{dx}{a} = \frac{1}{a^2} \quad (5.31)$$

Le pavé élémentaire dans l'espace temps-fréquence est donc de surface constante alors que la résolution temporelle est proportionnelle à a et que la résolution fréquentielle est proportionnelle à $\frac{1}{a}$ (figure 5.2).

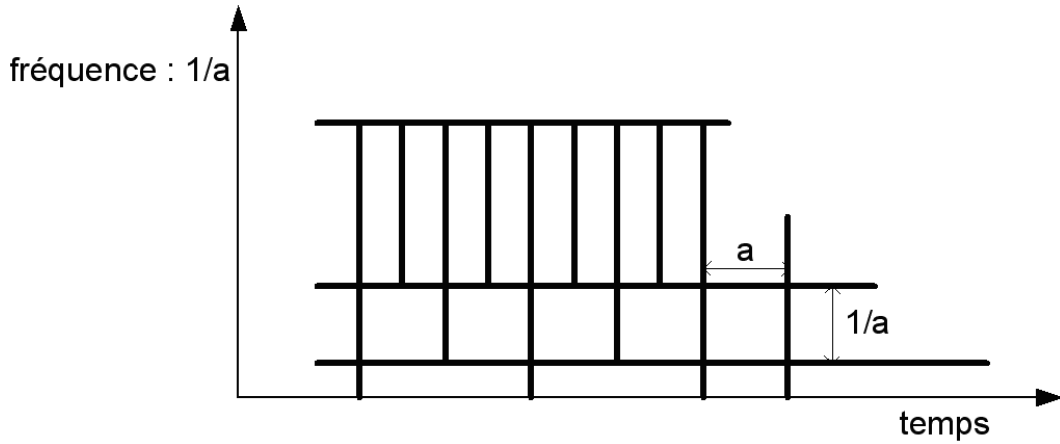


FIG. 5.2 – Pavage temps-fréquence pour la transformée en ondelettes discrète

Le "chapeau mexicain" (5.32) est un exemple d'ondelette fréquemment utilisé.

$$\Psi(x) = \frac{2}{\sqrt{3}} \pi^{-\frac{1}{4}} (1 - x^2) e^{-\frac{x^2}{2}} \quad (5.32)$$

La figure 5.3 représente cette ondelette pour les facteurs d'échelle $a = 1$ (la courbe la plus localisée), et $a = 2$.

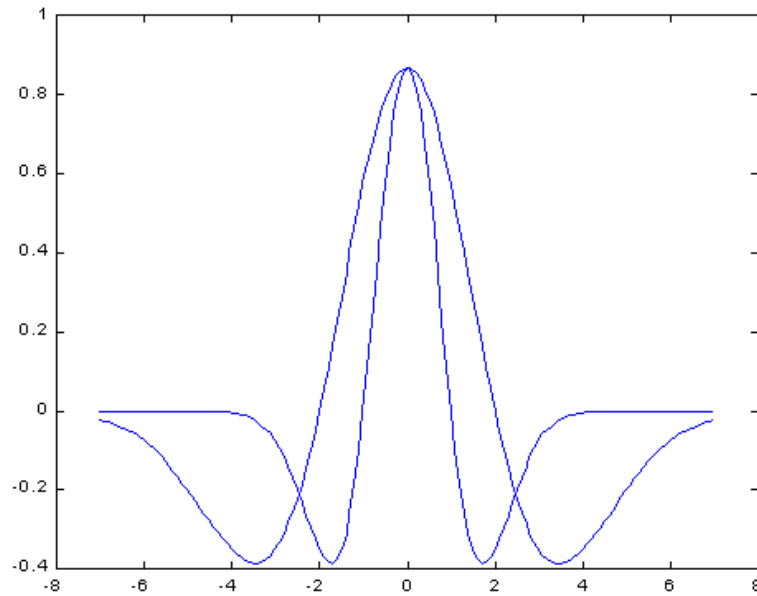


FIG. 5.3 – Ondelette "chapeau mexicain"

Inversion et admissibilité

Ingrid Daubechies [DAU92] a montré que si l'ondelette est convenablement choisie, la transformée en ondelettes est inversible, et il est alors possible de reconstruire la fonction après l'analyse (5.33). L'ondelette dans ce cas est dite *admissible*.

$$f(t) = C_{\Psi}^{-1} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{1}{a^2} \langle f(t), \Psi_{a,b}(t) \rangle \Psi_{a,b}(t) da db \quad (5.33)$$

Le coefficient C_{Ψ} est donné par l'équation (5.34).

$$C_{\Psi} = 2\pi \int_{-\infty}^{+\infty} |\hat{\Psi}(\omega)|^2 \frac{d\omega}{\omega} \quad (5.34)$$

La condition d'existence de ce coefficient, qui est aussi la condition d'admissibilité de l'ondelette analysante, correspond à la relation (5.35).

$$\int_0^{+\infty} |\hat{\Psi}(\omega)|^2 \frac{d\omega}{|\omega|} = \int_{-\infty}^0 |\hat{\Psi}(\omega)|^2 \frac{d\omega}{|\omega|} < \infty \quad (5.35)$$

Cette relation est le plus souvent remplacée par la contrainte (5.36).

$$\int_{-\infty}^{+\infty} \Psi(t) dt = 0 \quad (5.36)$$

Les différentes classes de transformées en ondelettes

Quatre types de transformées en ondelettes ont été développées. Il s'agit de la transformée en ondelettes continue, de la transformée en ondelettes discrète, de l'analyse multirésolution et des paquets d'ondelettes. Ces méthodes vont être présentées dans les sections suivantes.

La transformée en ondelettes continue

La transformation en ondelettes continue [MEY93] consiste à créer, à partir d'une fonction mère Ψ , une famille d'ondelettes $\Psi(ax+b)$, où a est un nombre réel servant à dilater (compresser ou étendre) la fonction Ψ , et b un nombre réel servant à la translater.

La définition générale d'une transformée en ondelettes (5.24) dépend des paramètres a et b , qui peuvent prendre une infinité de valeurs.

On parle de transformée en ondelettes continue lorsque l'on fait varier ces paramètres continuellement. Une transformée en ondelettes, ainsi définie, est redondante, c'est-à-dire que l'on obtient plus de coefficients d'ondelettes ($T^{ond}f(a,b)$) que nécessaire pour décrire le signal de manière exhaustive, car la base engendrée par l'ondelette est généralement non orthogonale. On peut noter que des variations de a ou de b vont entraîner des variations sur les valeurs $T^{ond}f(a,b)$.

Pour réduire cette redondance, les valeurs de a et b sont discrétisées, puisqu'en pratique, les signaux considérés sont le plus souvent discrets.

La transformée en ondelettes discrète

Dans ce type de transformation en ondelettes, l'ondelette est translaturée et dilatée seulement selon des valeurs discrètes. Généralement le facteur de dilatation a est une puissance de 2 (transformée en ondelettes dyadiques). Les ondelettes sont alors de la forme $\Psi(2^k t + l)$, où k et l sont des nombres entiers.

Pour la transformée en ondelettes continue, la famille d'ondelettes $\Psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \Psi\left(\frac{t-b}{a}\right)$ était considérée, avec a et b réels ($a \neq 0$) et Ψ admissible.

Dans le cas de la transformée en ondelettes discrète, on ne considère que des valeurs de a positives, pour plus de commodité. De plus, a et b sont restreints à

des valeurs discrètes. On choisit par exemple $a = a_0^m$ et $b = nb_0a_0^m$ où m et n sont des entiers relatifs, a_0 est le pas de dilatation ($a_0 \neq 0$) fixé et b_0 fixé. La famille d'ondelettes correspondante est indiquée par la relation (5.37).

$$\Psi_{m,n}(t) = a_0^{\frac{-m}{2}} \Psi\left(\frac{t - nb_0a_0^m}{a_0^m}\right) = a_0^{\frac{-m}{2}} \Psi(a_0^{-m}t - nb_0) \quad (5.37)$$

Pour une fonction mère Ψ raisonnable et des valeurs de a_0 et b_0 appropriées, il existe $\Psi_{m,n}$ telle que la relation (5.38) soit vérifiée.

$$f = \sum_{m,n} \langle f, \Psi_{m,n} \rangle \Psi_{m,n} \quad (5.38)$$

Les familles d'ondelettes discrètes ne forment que très rarement des bases. Ainsi, pour avoir un algorithme de reconstruction numérique stable de f , il faut que l'ensemble des $\Psi_{m,n}$ forme une trame d'ondelettes ou (*frames*) [DAU90].

Une famille d'ondelettes $\{\Psi_{m,n}; m, n \in \mathbb{Z}\}$ forme une trame d'ondelettes s'il existe deux réels A et B tels que l'inéquation (5.39) soit vérifiée pour toute fonction f de L^2 .

$$A\|f\|^2 \leq \sum_{m,n} |\langle f, \Psi_{m,n} \rangle|^2 \leq B\|f\|^2 \quad (5.39)$$

A et B sont les bornes de la trame et vérifient $0 < A \leq B < +\infty$.

Les ondelettes orthogonales sont des ondelettes discrètes particulières. Elles sont plus difficiles à construire, mais fournissent une représentation sans redondance.

Analyse multirésolution

Dans L^2 , une analyse à la résolution j d'une fonction f est obtenue par projection de f sur V_j (sous espace de L^2) : $A_j f \in V_j$, A_j étant un projecteur sur V_j .

L'analyse multirésolution [STR97, MAL89] se construit en considérant les sous espaces V_j emboîtés les uns dans les autres. Le passage de V_j à V_k représente un changement d'échelle. Il existe une fonction Φ de L^2 , appelée "fonction d'échelle", qui engendre une base orthonormée de V_j par dilatation et translation.

Les fonctions de base sont construites par la relation (5.40).

$$\Phi_{j,n}(t) = 2^{-\frac{j}{2}}\Phi(2^{-j}t - n), n \in \mathbb{Z} \quad (5.40)$$

La relation d'orthogonalité entre les fonctions de base pour une échelle donnée est indiquée par (5.41).

$$\forall j, k, n \in \mathbb{Z} \quad \langle \Phi_{j,n}, \Phi_{j,k} \rangle = \delta_{n,k} \quad (5.41)$$

où $\delta_{n,k}$ est le symbole de kroeneker.

Ces espaces vérifient la propriété (5.42).

$$\cdots \subset V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset \cdots \subset V_{j+1} \subset V_j \cdots \quad (5.42)$$

avec :

$$\bigcup_{j \in \mathbb{Z}} V_j = L^2 \quad (5.43)$$

$$\bigcap_{j \in \mathbb{Z}} V_j = \{0\} \quad (5.44)$$

Les sous espaces V_j sont appelés espaces d'approximation. On définit les espaces de détail W_j comme les compléments orthogonaux aux V_j dans V_{j-1} .

Les espaces de détail W_j forment une famille d'espaces orthogonaux.

Il existe alors une ondelette Ψ qui par dilatation et translation engendre une base orthonormée des W_j . On note Ψ cette ondelette de L^2 .

Les fonctions de base seront construites par (5.45) et vérifient la relation d'orthogonalité (5.46).

$$\Psi_{j,n}(t) = 2^{-\frac{j}{2}}\Psi(2^{-j}t - n), n \in \mathbb{Z} \quad (5.45)$$

$$\forall i, j, k, n \in \mathbb{Z} \quad \langle \Psi_{j,n}, \Psi_{i,k} \rangle = \delta_{j,i}\delta_{n,k} \quad (5.46)$$

Il est alors possible de reconstruire l'approximation à l'échelle immédiatement plus fine en utilisant les détails du signal fournis par sa projection sur la base de W_j .

La formule de reconstruction s'écrit :

$$A_{j-1}f = A_j f + D_j f \quad (5.47)$$

où : A_j est un projecteur sur V_j et D_j est un projecteur sur W_j :

$$A_j f = \sum_n \langle f, \Phi_{j,n} \rangle \Phi_{j,n} \quad (5.48)$$

$$D_j f = \sum_n \langle f, \Psi_{j,n} \rangle \Psi_{j,n} \quad (5.49)$$

$$(5.50)$$

Dans le cadre du traitement du signal, l'analyse multirésolution nous permet de décomposer toute suite temporelle $(x(t))_t$ en un ensemble de signaux d'approximation et de détail (5.51), les signaux d'approximation représentant les composantes basses fréquences de la série, et les signaux de détail représentant les composantes hautes fréquences.

$$x(t) = \sum_{j=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} c_{j,n} \Phi_{j,n}(t) + \sum_{j=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} d_{j,n} \Psi_{j,n}(t) \quad (5.51)$$

Les approximations correspondent à la première partie de la somme, alors que les détails correspondent à la seconde partie.

Les coefficients discrets d'approximation et de détail sont donnés respectivement par les relations (5.52) et (5.53).

$$c_{j,n} = \langle x, \Phi_{j,n} \rangle = \int_{-\infty}^{+\infty} x(t) \Phi_{j,n}(t) dt, \quad j, n \in Z \quad (5.52)$$

$$d_{j,n} = \langle x, \Psi_{j,n} \rangle = \int_{-\infty}^{+\infty} x(t) \Psi_{j,n}(t) dt, \quad j, n \in Z \quad (5.53)$$

Le schéma de décomposition correspondant est présenté figure 5.4.

Les paquets d'ondelettes

Il s'agit d'une généralisation de l'analyse multirésolution, en appliquant le schéma présenté figure 5.4 également aux sous espaces W_j .

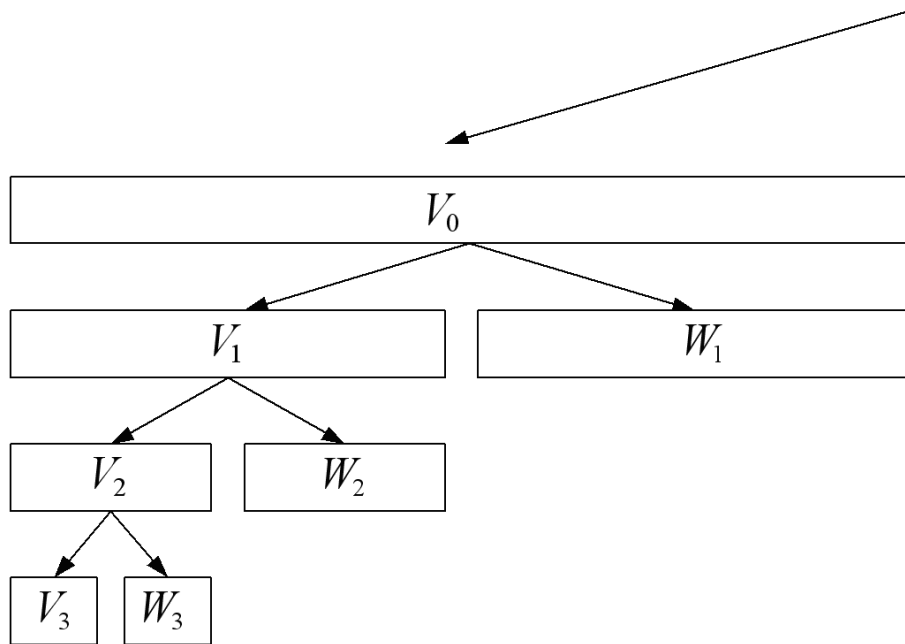


FIG. 5.4 – Analyse multirésolution

Il est possible d'utiliser des bases de fonctions d'échelle différentes de celles des fonctions d'ondelettes. De plus, les fonctions de base peuvent être différentes à chaque échelle. La reconstruction est faite, pour une résolution donnée, en réutilisant la base utilisée lors de l'analyse pour cette résolution.

Cet algorithme conduit à une décomposition en sous-bandes de fréquence du signal, et grâce à sa souplesse, par le choix des différentes bases, permet une bonne adaptation aux signaux à étudier.

Le processus de décomposition est présenté figure 5.5.

Les différents types de transformées en ondelettes présentés précédemment permettent donc une représentation des signaux dans un espace temps-échelle, ou temps-fréquence. A partir de cette représentation, un certain nombre de primitives peuvent donc être extraites afin de caractériser les signaux. L'extraction des primitives fait l'objet de la section suivante.

5.5.2 Extraction des primitives

Les signaux sonores qui nous intéressent sont caractérisés par un comportement non stationnaire, ce qui est le cas de nombreux types de signaux, et particulièrement des signaux médicaux [GUT01, HUL02]. Ainsi, l'utilisation des ondelettes comme

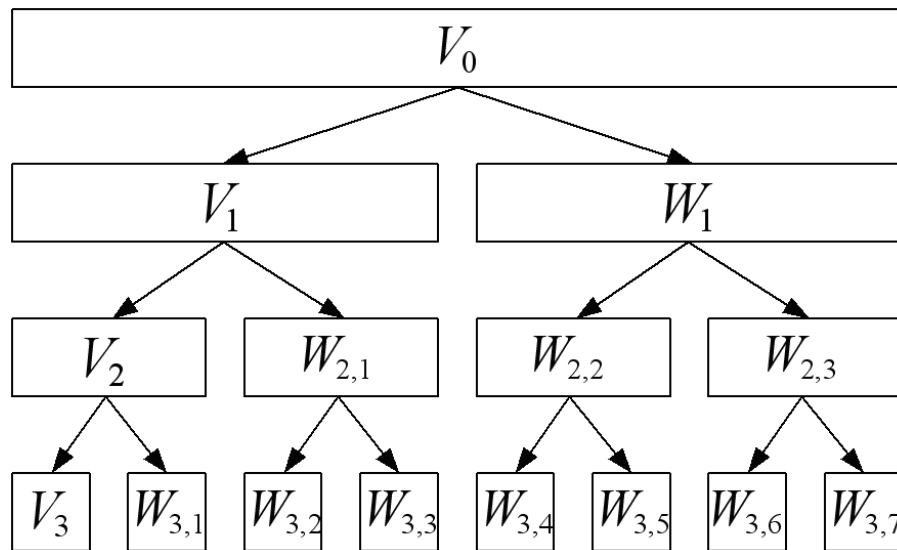


FIG. 5.5 – Analyse en paquets d'ondelettes

outil d'analyse nous a paru intéressante, ces transformées étant particulièrement bien adaptées aux signaux non stationnaires présentant de fortes discontinuités.

Les quatre types de transformation en ondelettes présentés précédemment sont donc envisageables afin d'obtenir une représentation des signaux dans un espace temps-fréquence. Cependant, notre choix s'est porté sur l'analyse multirésolution puisqu'elle représente le meilleur compromis entre rapidité d'exécution et précision. En effet, l'analyse multirésolution correspond à une technique de calcul rapide de la transformée en ondelettes discrète, cette dernière présentant elle-même un coût calculatoire très inférieur à celui de la transformée en ondelettes continue qui comporte une redondance importante de l'information. La méthode par paquets d'ondelettes a l'avantage d'être plus précise que l'analyse multirésolution. Cependant le gain obtenu par cette précision ne justifie pas toujours la perte liée à un temps de calcul supérieur, l'analyse multirésolution étant souvent suffisante pour décrire le comportement non stationnaire des signaux.

Les signaux sonores sont donc décomposés grâce à une analyse multirésolution par ondelettes présentée à la section 5.5.1. La décomposition étant réalisée jusqu'au niveau T , T signaux d'approximation, et autant de signaux de détail sont obtenus. Un exemple de décomposition d'un signal de déglutition est donné figure 5.6, T étant

égal à 10.

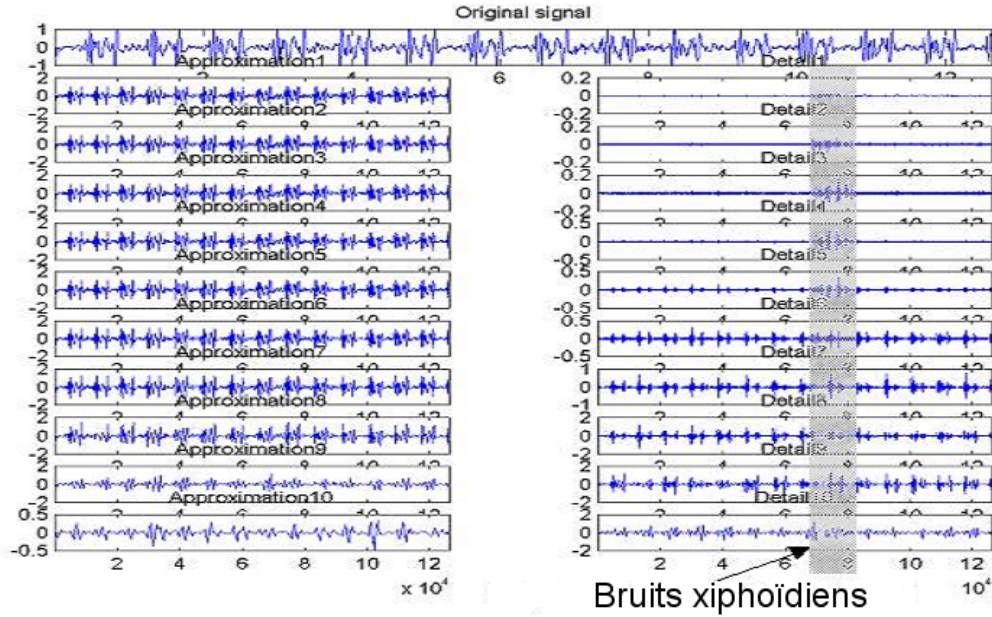


FIG. 5.6 – Décomposition d'un signal de déglutition jusqu'au niveau 10. Le signal du haut correspond au signal d'origine, alors que la colonne de gauche présente les signaux d'approximation successifs, et la colonne de droite, les signaux de détail successifs.

Les variations brutales sont plus particulièrement mises en évidence dans les signaux de détail qui représentent les composantes hautes fréquences du signal d'origine. Ce sont donc ces signaux de détail qui nous intéressent.

Ainsi, pour chacun d'eux, trois primitives sont calculées : l'énergie En , l'amplitude maximale A , et l'entropie Ent .

L'énergie $En(j)$ et l'entropie $Ent(j)$ du signal de détail Sd_j de niveau j sont respectivement données par (5.54) et (5.55).

$$En(j) = \frac{1}{N} \sum_{k=0}^{N-1} Sd_j^2(k) \quad (5.54)$$

$$Ent(j) = - \sum_{k=0}^{N-1} Sd_k^2(k) \log Sd_j^2(k) \quad (5.55)$$

Le vecteur de primitives d'ondelettes que nous considérons est donc composé de $3T$ éléments, T étant le niveau maximal de décomposition.

Les quatre ensembles de primitives présentés dans cette section et dans les sections précédentes sont issus de méthodes fréquemment utilisées dans le domaine du traitement du signal, et de ces applications à la reconnaissance de la parole, de la musique ou encore de signaux médicaux. La section suivante est consacrée à la présentation d'une technique permettant la caractérisation de la structure de phénomènes jusqu'alors absente du domaine du traitement du signal sonore. Nous allons donc montrer comment nous avons adapté à l'étude de signaux sonores, l'analyse par les lois de Zipf, développée initialement pour l'analyse linguistique statistique.

5.6 Etude des signaux par les lois de Zipf

Afin de caractériser la structure des signaux sonores, nous avons élaboré un ensemble de primitives liées aux lois de *Zipf* et *Zipf Inverse* traditionnellement utilisées pour l'analyse statistique de la structure de textes littéraires. Ces lois vont être présentées dans un premier temps. Puis leur adaptation à l'étude des signaux sonores sera décrite ainsi que les primitives extraites.

5.6.1 Les lois de Zipf pour l'étude statistique de phénomènes

G. K. Zipf proposa dans [ZIP49] une loi empirique connue sous le nom de loi de Zipf, qui a été observée dans de nombreux domaines [LI99]. Cette loi énonce que la fréquence $f(p)$ d'un événement p et son rang $r(p)$ par rapport à sa fréquence d'apparition (du plus fréquent au moins fréquent) sont liés par une loi puissance (5.56).

$$f(p) = A_1 r(p)^{-\beta} \quad (5.56)$$

où A_1 et β sont des constantes réelles.

L'application la plus célèbre de la loi de Zipf est la linguistique statistique. En effet, il a été montré [ZIP49] que cette loi est valide dans le cas de textes écrits en langage naturel, β étant proche de 1. Le même phénomène a été observé pour la population des villes [ZIP49, MAR98] : la population des villes et leur rang (de la plus peuplée à la moins peuplée) sont liés par la même loi puissance, avec β également proche de 1. En fait, il est apparu que cette régularité concerne de très

nombreux phénomènes dans divers domaines. Citons encore par exemple la distribution du revenu des employés, la distribution de la taille des entreprises aux Etats-Unis [AXT01, STA95], les séquences d'ADN [PEN95], les battements cardiaques [KAL01], les accès internet [BRE99], et même les tirs bloqués des joueurs de basket de la NBA [POO95].

Diverses approches ont été envisagées afin d'expliquer les fondements théoriques de la loi de Zipf [POO95]. Cependant, elles reposent sur des hypothèses concernant les propriétés du système sous-jacent, qui ne peuvent pas toujours être vérifiées. Ainsi, de tels systèmes sont généralement connus pour obéir à la loi de Zipf par une étude empirique d'observation.

Grâce à la loi de Zipf, une caractérisation de certains phénomènes peut donc être réalisée. Tout d'abord, il est possible de distinguer ceux qui obéissent à la loi de Zipf, et les autres. Puis, lorsque la loi de Zipf est vérifiée, les valeurs des paramètres définissant la loi appropriée, tels que A_1 et β peuvent apporter une information intéressante. Par exemple, dans [COH97] la loi de Zipf est utilisée afin de discriminer les textes écrits en langage naturel de ceux écrits en langage artificiel. De plus, [HAV95] a montré que la distance entre les courbes de Zipf associées à différents livres permet la caractérisation de l'auteur, la distance étant plus faible lorsque les livres sont écrits par le même auteur, que lorsqu'ils sont écrits par des auteurs différents.

Un autre comportement observé pour les textes écrits en langage naturel a été modélisé par une loi, appelée loi de Zipf Inverse [COH97]. Elle s'intéresse à la distribution de la fréquence des mots d'un texte. Zipf a découvert une loi puissance qui est vérifiée uniquement pour les mots de faible fréquence : le nombre de mots distincts $I(f)$ de fréquence d'apparition f et la fréquence f sont liés par la relation (5.57).

$$I(f) = A_2 f^{-\gamma} \quad (5.57)$$

où A_2 et γ sont des constantes réelles.

Les perspectives très intéressantes offertes par les lois de Zipf nous ont amené à évaluer leur pertinence et leur efficacité pour le problème de classification de signaux sonores. Ainsi, nous avons choisi d'utiliser une analogie entre les signaux sonores et les textes, considérés comme des séquences ordonnées de motifs. La première étape du processus est donc le codage des signaux sonores afin de les transformer en textes.

La pertinence de tels codages est discutée dans [WAT02]. Les lois de Zipf pourront alors être utilisées pour l'analyse de ces signaux à valeurs discrètes.

5.6.2 Codage des signaux sonores

Le codage des signaux sonores est une étape très importante car il doit mettre en évidence l'information recherchée dans les signaux.

Un signal sonore contient au moins deux types d'informations : une information temporelle et une information fréquentielle. C'est la raison pour laquelle trois types de codages ont été élaborés afin d'exploiter l'information pertinente contenue dans les signaux. Le premier type repose sur la représentation temporelle des signaux sonores, le second repose sur leur représentation fréquentielle, et le troisième sur leur représentation temps-fréquence.

Codages temporels

Les codages temporels des signaux sonores exploitent l'information apportée par l'évolution temporelle de l'amplitude des signaux. Trois codages TC (Temporal coding) ont été développés : les codages TC1, TC2, et TC3.

Le codage TC1 : codage d'évolution

Comme pour tous les codages proposés, le principe du codage TC1, décrit figure 5.7 est de permettre la construction d'une séquence de motifs basée sur le codage du signal sonore d'origine.

Ici, les motifs sont eux-mêmes des séquences de trois caractères et représentent l'évolution temporelle locale sur quatre échantillons. Ainsi, la séquence de motifs est liée à l'enveloppe globale du signal temporel. Les caractères, pour chaque échantillon sonore, peuvent prendre trois valeurs : le caractère *U* (Up), *F* (Flat) et *D* (Down), qui remplacent les valeurs du signal.

La lettre *U* est utilisée lorsque la différence d'amplitude entre deux échantillons successifs du signal est positive. Si cette différence est proche de 0, la lettre *F* est utilisée, et sinon la lettre *D* est utilisée.

Le motif suivant est obtenu en translatant d'un échantillon vers la droite, la fenêtre d'analyse de taille 3 échantillons.

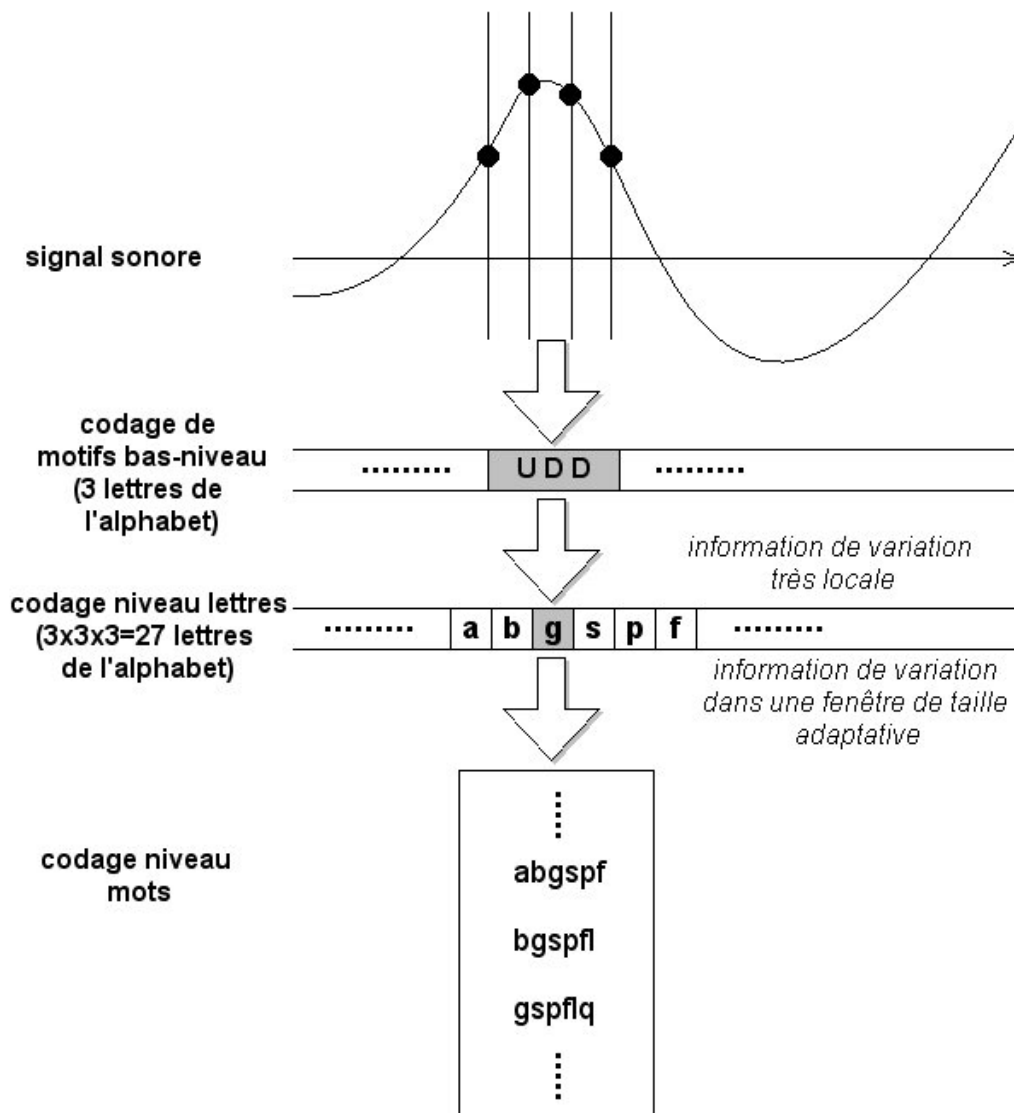


FIG. 5.7 – Principe du codage TC1.

Une séquence de motifs est finalement construite à partir du signal sonore. Il y a $3^3 = 27$ possibilités différentes de motifs lorsqu'ils sont de taille 3. Chacun d'eux est associé à une lettre de l'alphabet. Le nombre de lettres nécessaires est égal au cardinal de l'ensemble des motifs.

A ce stade, une séquence de lettres est obtenue, ces dernières prenant en compte une information de variation d'amplitude très locale. Afin d'adapter la longueur de la fenêtre d'observation à des échelles plus larges, les lettres sont regroupées afin de créer des mots de longueur n qui vont constituer un texte. n est un paramètre qui

doit être fixé et qui dépend de la longueur de la fenêtre d'observation à utiliser.

L'intérêt du codage TC1 est que chaque lettre correspond à la forme locale de l'enveloppe du signal temporel. Ainsi, les mots représentent les variations d'amplitude du signal.

Cependant, l'information concernant l'importance de ces variations n'est pas utilisée dans le but de limiter le nombre de motifs possibles. En effet, seulement des évolutions positive, négative ou nulle du signal sont considérées.

Le codage TC2 : codage d'évolution adaptatif

L'objectif du codage TC2 est d'utiliser l'information relative à l'importance des variations d'amplitude du signal sonore, contrairement au codage précédent. La notion de grande ou petite variation est bien sûr liée au signal à étudier car pouvant dépendre notamment des conditions d'acquisition du signal. Une méthode de définition adaptative de l'importance des signaux est donc utilisée. Pour cela, le signal doit être lu entièrement avant le début du processus de codage.

Le principe du codage TC2, représenté figure 5.8, est de calculer la différence d'amplitude entre les échantillons sonores successifs du signal. Afin de réduire la dimension des données, ces différences d'amplitude doivent être discrétisées selon K valeurs. Ainsi, à partir de ces données initiales, une classification par l'algorithme des K -means [HAR75] est réalisée, permettant d'obtenir K classes. Chacune d'elles correspond à un intervalle de variations d'amplitude. Ce processus de classification permet de s'assurer que le code n'est pas absolu mais plus sensible aux variations internes des signaux. Une lettre de l'alphabet est ensuite associée à chaque classe.

A partir du signal sonore, une séquence de lettres est obtenue, celles-ci correspondant à la classe de chaque échantillon.

Enfin, les lettres sont regroupées afin de construire un texte constitué de mots de longueur n .

Le codage TC3 : codage d'amplitude adaptatif

Son principe, expliqué figure 5.9 est de considérer la valeur d'amplitude des échantillons sonores, au lieu de la valeur de différence d'amplitude entre deux échantillons successifs comme c'était le cas pour le codage TC2.

Une classification par l'algorithme des K -means est utilisée pour les mêmes raisons que pour le codage TC2, afin d'obtenir K classes correspondant à des intervalles d'amplitude, et que l'on peut associer à une lettre de l'alphabet. Les lettres obtenues

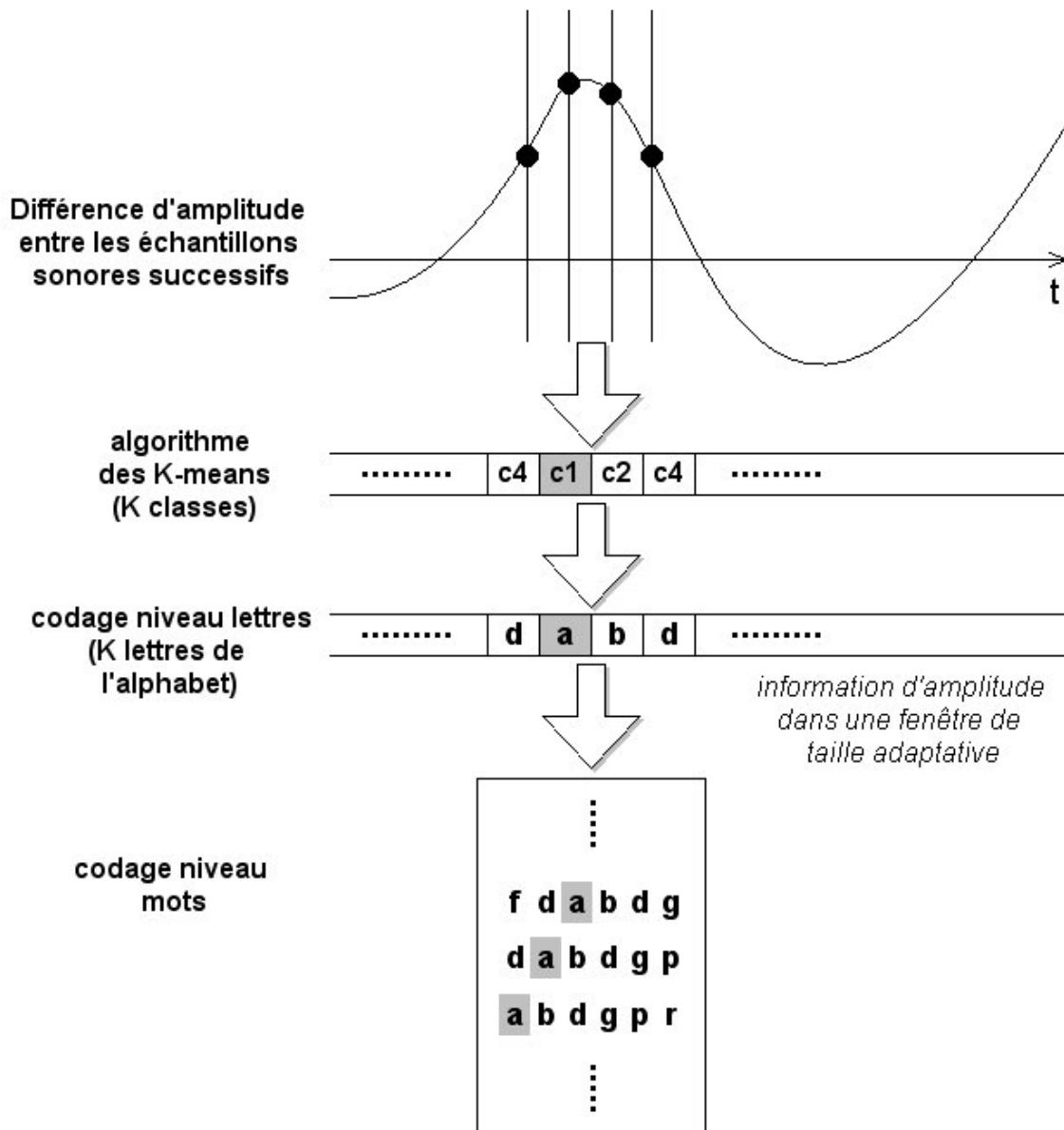


FIG. 5.8 – Principe du codage TC2.

sont ensuite regroupées afin de former une séquence de mots de longueur n .

Codage Fréquentiel

Les codages présentés dans la section précédente permettent de construire une séquence de mots à partir de la représentation temporelle des signaux sonores. L'ob-

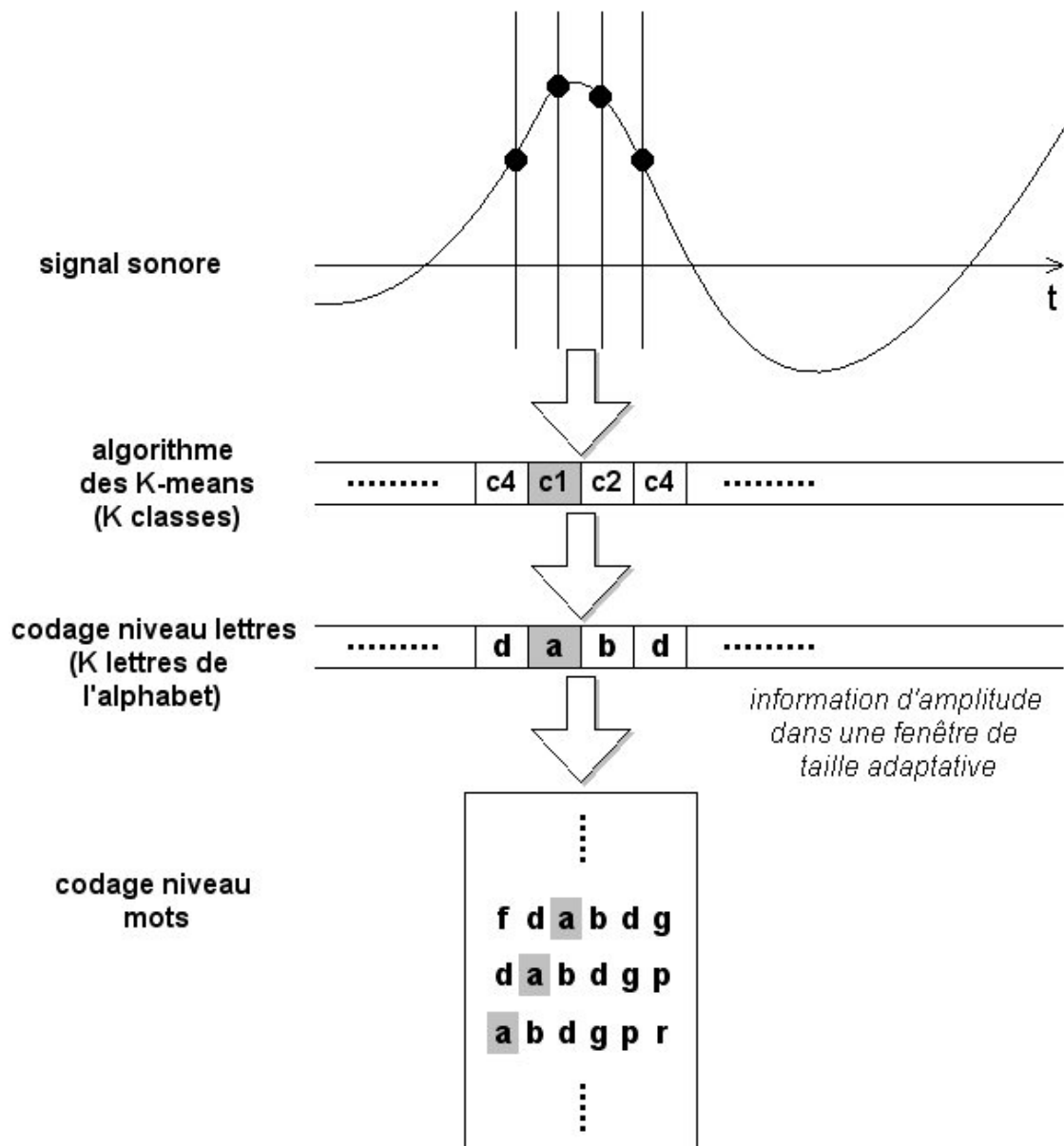


FIG. 5.9 – Principe du codage TC3.

jectif du codage FC (Frequential Coding) présenté dans cette section, et décrit figure 5.10, est d'exploiter l'information fréquentielle contenue dans les signaux, grâce au codage de leur représentation fréquentielle obtenue par la transformée de Fourier décrite par les relations (5.17,5.18) de la section 5.5.1.

Le module de la transformée de Fourier $|X(k)|$ est ensuite calculé. A partir de ces

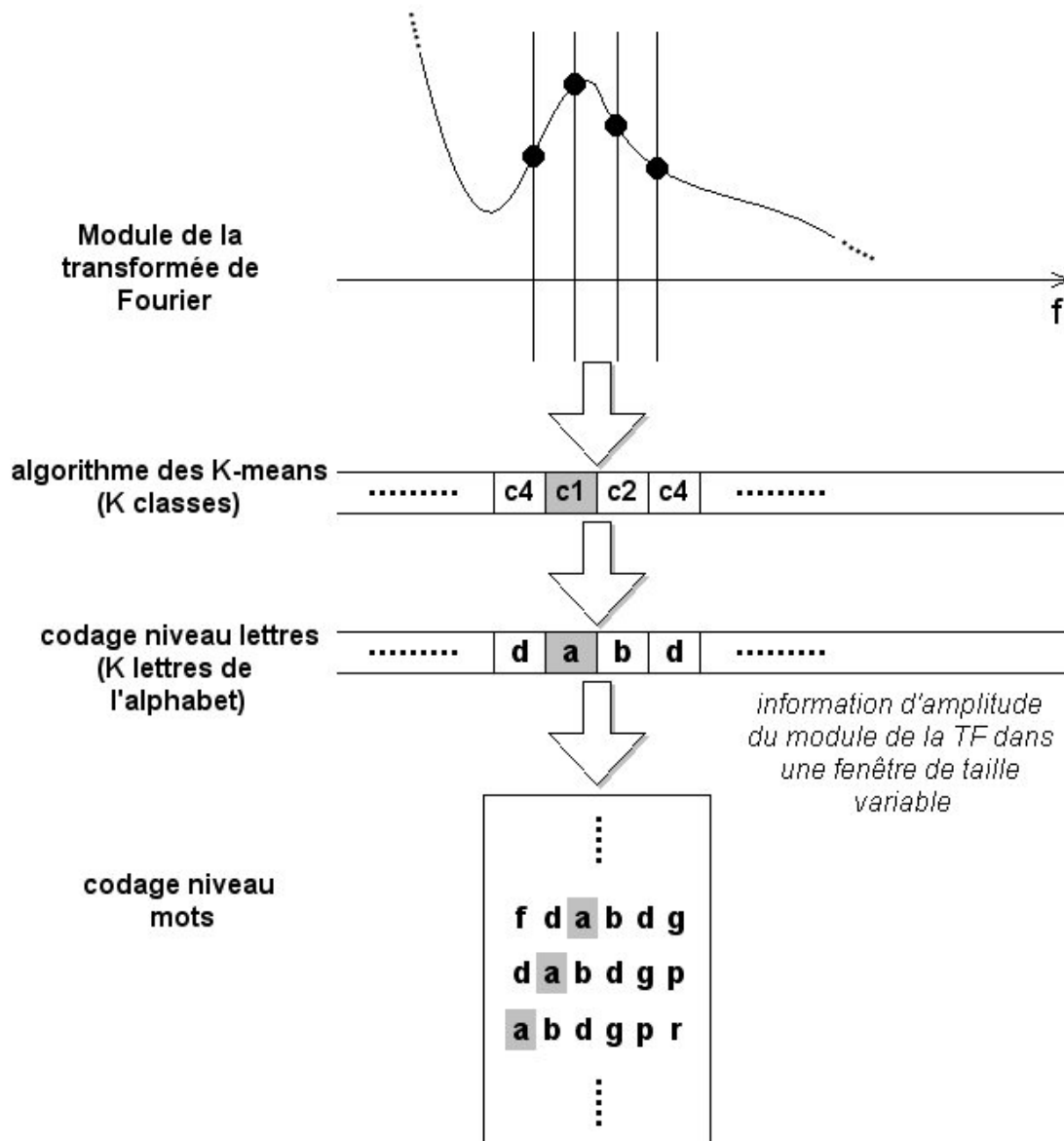


FIG. 5.10 – Principe du codage FC.

données, une classification par l'algorithme des K -means est réalisée afin d'obtenir K classes correspondant aux intervalles de valeurs du module de la transformée de Fourier. Chaque classe est ensuite associée à une lettre d'un alphabet de cardinal K . Une séquence de lettres est ensuite obtenue en remplaçant la valeur du module de la transformée de Fourier des échantillons du signal sonore par la lettre associée à leur classe. Ces lettres sont finalement regroupées afin de former une séquence de mots de longueur n . Ici, le module de la transformée de Fourier est traité de la même

manière que le signal sonore dans le cas du codage TC3.

Codages temps-fréquence

Le premier type de codage présenté repose sur la représentation temporelle des signaux sonores alors que le deuxième type de codage repose sur leur représentation fréquentielle.

Ainsi, les informations temporelle et fréquentielle ont été exploitées, mais de manière indépendante. Dans cette section, deux codages basés sur la représentation temps-échelle, ou temps-fréquence, des signaux sonores sont présentés, permettant de prendre en compte l'information temporelle et fréquentielle simultanément.

Ces deux codages sont basés sur des transformées en ondelettes (section 5.5.1) permettant d'obtenir une représentation temps-échelle des signaux. Le premier codage repose sur la transformée en ondelettes continue et le second codage repose sur une analyse multirésolution des signaux.

Codage TSC1 : transformée en ondelettes continue

Le codage TSC1 (Time Scale Coding), dont le principe est décrit figure 5.11, repose sur une représentation temps-échelle des signaux sonores, obtenus grâce à la transformée en ondelettes continue. Cette transformée est présentée dans la section 5.5.1 et repose sur l'équation (5.24).

Le résultat de cette transformée correspond aux coefficients d'ondelettes qui sont fonction de l'échelle a liée à la fréquence, et de la position b . Les positions correspondent à l'indice des échantillons du signal, et les valeurs d'échelle doivent être fixées.

La transformée en ondelettes continue des signaux sonores est réalisée en utilisant k différentes valeurs d'échelle. Ainsi, chaque position b est associée à un ensemble de k coefficients appelé *profil* C_b et organisé dans le vecteur décrit dans (5.58).

$$C_b = (c_1^b, c_2^b, \dots, c_k^b)^T \quad (5.58)$$

où M^T désigne la transposée de la matrice M et c_i^b est le coefficient d'ondelette d'indice i calculé à la position b .

L'objectif, à ce stade où la nature de l'information à considérer à chaque instant a été choisie, est de créer un code en utilisant un nombre limité de symboles qui

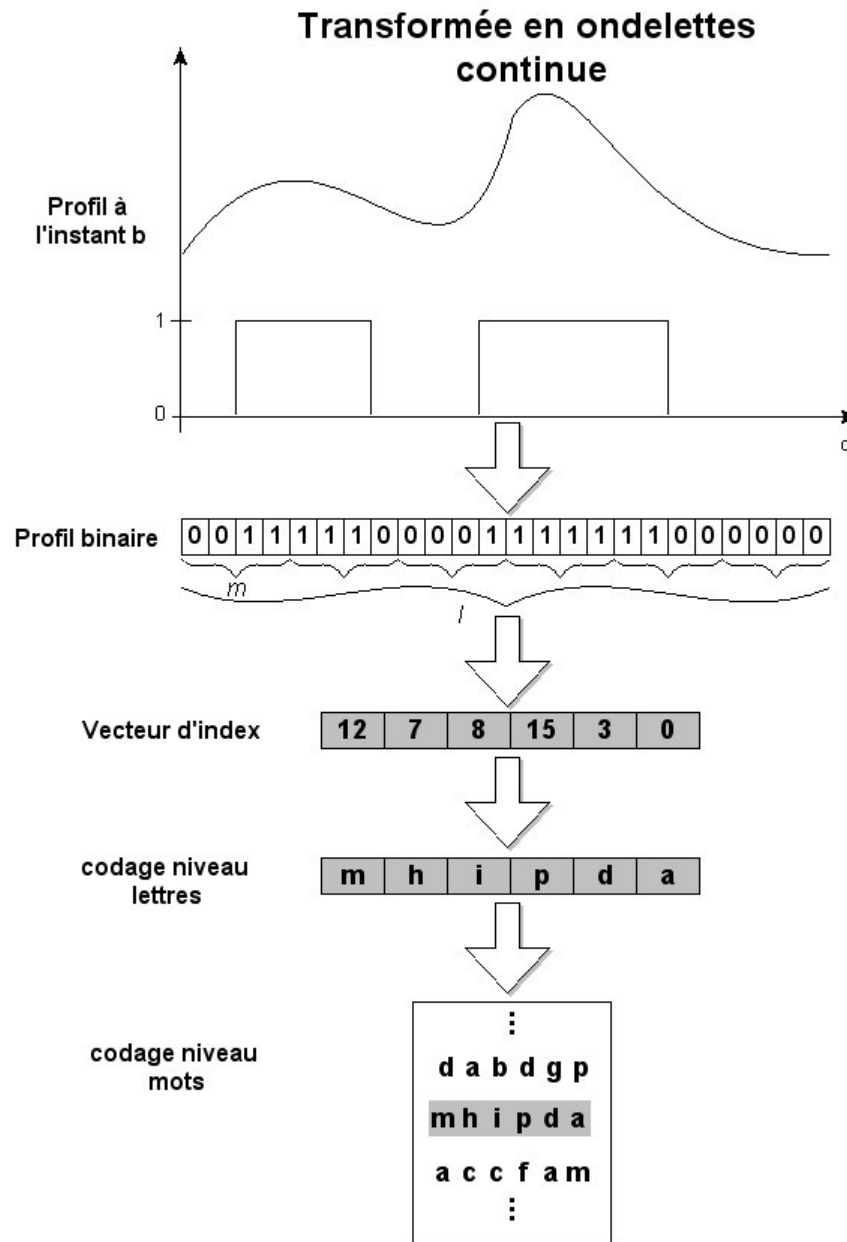


FIG. 5.11 – Principe du codage TSC1.

seront regroupés afin de former de courtes séquences significatives correspondant à des "mots".

Ainsi, les profils C_b calculés à chaque instant b sont constitués de k coefficients numériques tels que les premiers soient liés aux hautes fréquences du signal alors que les suivants soient liés aux fréquences plus basses. Cette variété de l'information doit être préservée dans le codage, et les mots vont donc représenter cette

unité d'observation. Afin de permettre l'étude de l'évolution des paramètres dans le temps, chaque valeur de coefficient est comparée à la valeur moyenne des coefficients correspondants calculée sur l'ensemble du signal. Seule une information binaire est alors conservée, relative à l'importance du coefficient considéré par rapport à la valeur moyenne. Cette information est contenue dans une représentation binaire de k coefficients. Néanmoins, afin d'obtenir un codage plus compact, ces coefficients sont regroupés de manière à former une séquence de longueur m interprétée comme le code d'un nombre exprimé dans une base binaire. Un symbole explicite est ainsi défini parmi 2^m symboles possibles.

Les différentes étapes de ce processus sont exprimées par les formules suivantes.

Soit le vecteur \bar{C} égal à la valeur moyenne des vecteurs C_b , défini par (5.59).

$$\bar{C} = \frac{1}{n} \sum_{\tau=1}^n C_b = (\bar{c}_1, \bar{c}_2, \dots, \bar{c}_k)^T \quad (5.59)$$

où n est égal au nombre d'échantillons considérés.

Pour chaque position b , un profil binaire est ensuite calculé (5.60).

$$B_b = (b_1^b, b_2^b, \dots, b_k^b)^T \quad (5.60)$$

avec :

$$b_i^b = \begin{cases} 1 & \text{si } c_i^b > \bar{c}_i \\ 0 & \text{sinon} \end{cases} \quad \forall i \in [1, k] \quad (5.61)$$

Les coefficients b_i^b sont ensuite regroupés en l ensembles de m coefficients tels que l et m soient des entiers strictement positifs et $l * m = k$.

Le vecteur S_b est finalement calculé (5.62).

$$S_b = (s_1^b, s_2^b, \dots, s_l^b)^T \quad (5.62)$$

avec :

$$s_j^b = \sum_{i=1}^m 2^{i-1} b_{(j-1)*m+i}^b \quad \forall j \in [1, l] \quad (5.63)$$

Les entiers s_j^b sont interprétés comme l'index de caractères d'un alphabet contenant 2^m caractères. Ainsi, chaque vecteur S_b , et donc chaque position b , est codé par un mot de longueur l lettres, sélectionné dans un ensemble de mots de cardinal 2^{lm} .

A partir de l'ensemble de profils constitués des coefficients de la transformée en ondelettes continue du signal, un texte composé d'une séquence ordonnée de mots est donc obtenu.

Le codage TSC2 : multirésolution par ondelettes

Le codage TSC2 (Time Scale Coding) est basé sur une analyse multirésolution par ondelettes des signaux sonores.

Cette analyse, présentée dans la section 5.5.1 permet de décomposer un signal d'origine en un ensemble de signaux d'approximation comportant les informations basse fréquence du signal, et de signaux de détail comportant les informations haute fréquence.

Puisque les signaux auxquels nous nous intéressons ont un comportement non stationnaire avec des variations brutales, notre étude porte sur les signaux de détail permettant de mettre en évidence ces informations.

Ainsi, à partir de la représentation du signal obtenue par une analyse multirésolution par ondelettes, des profils C_b sont calculés, tout comme dans le cas du codage TSC1. Ces profils sont obtenus en considérant k signaux de détail de niveaux successifs. Pour chaque position b , les profils C_b sont donc constitués de k coefficients numériques correspondant aux valeurs des signaux de détail à la position b .

Un texte est ensuite obtenu de la même manière que pour le codage TSC1.

Cette phase de codage peut être considérée comme la transformation d'un signal dans un nouvel espace de représentation, et dans chaque cas, est obtenue une séquence de mots construits sur un alphabet relativement réduit. Chaque codage a pour rôle de mettre en valeur différentes caractéristiques du signal sonore. Néanmoins la quantité de données est toujours importante et un nombre limité de primitives doit être extrait afin de permettre la classification des signaux.

5.6.3 Extraction des primitives

Dans cette section vont être décrites les primitives extraites en considérant les lois de Zipf et Zipf Inverse présentées dans la section 5.6.1, à partir des signaux sonores en utilisant les codages présentés dans la section précédente. Cet ensemble de primitives constituera la représentation des échantillons sonores, qui sera utilisée pour la classification.

Comme indiqué dans la section 5.6.1, la loi de Zipf, mathématiquement représentée par la relation (5.56), implique le nombre d'occurrences d'un événement et permet de modéliser la relation entre le nombre d'occurrences de ces événements et leur rang relatif à leur fréquence d'apparition (du plus fréquent au moins fréquent). Cette relation s'exprime sous la forme d'une loi puissance qui, lorsqu'elle est représentée dans un espace à échelles logarithmiques (courbe de Zipf) devient linéaire. Un exemple est donné figure 5.12 dans le cas des codages TC3 et TSC1 d'un signal sonore.

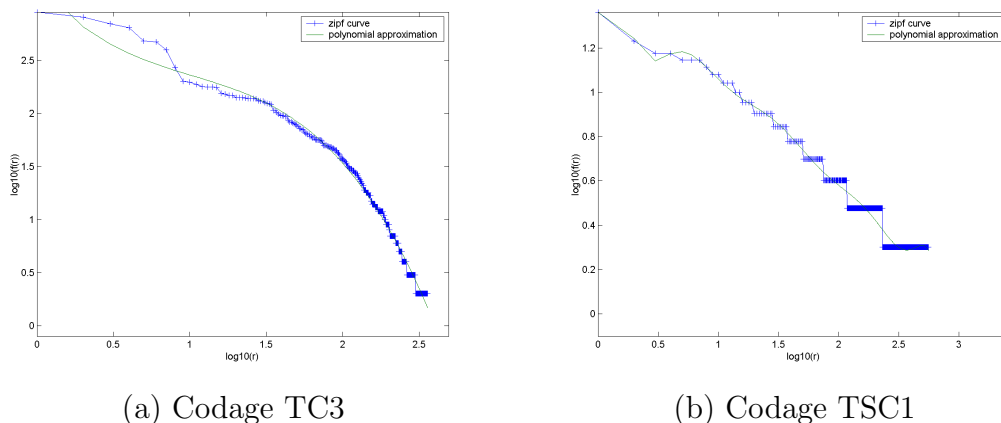


FIG. 5.12 – Exemple de courbes de Zipf dans le cas des codages TC3 (a) et TSC1 (b).

La forme de cette courbe est liée à la structure du texte correspondant. Afin de l'étudier, la loi de Zipf n'étant pas toujours parfaitement vérifiée, la fonction correspondante est approximée par une fonction polynômiale dont les coefficients feront partie de l'ensemble des primitives associées au signal.

Le polynôme est choisi de manière à minimiser l'erreur d'approximation de la fonction associée à la courbe de Zipf, par la méthode des moindres carrés. La difficulté consiste donc à déterminer le degré optimal du polynôme. La méthode retenue est de faire varier le degré du polynôme d'une valeur minimale à une valeur maximale,

et de déterminer dans chaque cas un polynôme par une méthode de minimisation des erreurs. Pour chaque degré l'erreur d'approximation E_d est ainsi calculée. La moyenne des erreurs E_d et leur l'écart-type sont alors calculés. La somme constitue la valeur seuil utilisée pour déterminer le degré du polynôme. De manière à choisir une meilleure approximation globale de la courbe nous favorisons les polynômes de faible degré. Ainsi, le degré retenu est le plus petit tel que la valeur d'erreur associée soit inférieure à ce seuil. Si cette condition n'est vérifiée dans aucun cas, le degré maximum est sélectionné. L'approximation obtenue ne reflète pas l'allure de la courbe, mais l'ensemble des points calculés.

Du fait de la transposition de la relation de Zipf à un espace logarithmique, la distribution des projections des points sur l'axe des abscisses n'est pas homogène. Un rééchantillonnage à pas constant de la courbe est donc réalisé permettant d'obtenir la courbe de Zipf rééchantillonnée qui sera approximée par un polynôme de la même manière que précédemment, qui sera plus fidèle à la forme globale de la courbe. Un exemple est donné figure 5.13 dans le cas des codages TC3 et TSC1 d'un signal sonore.

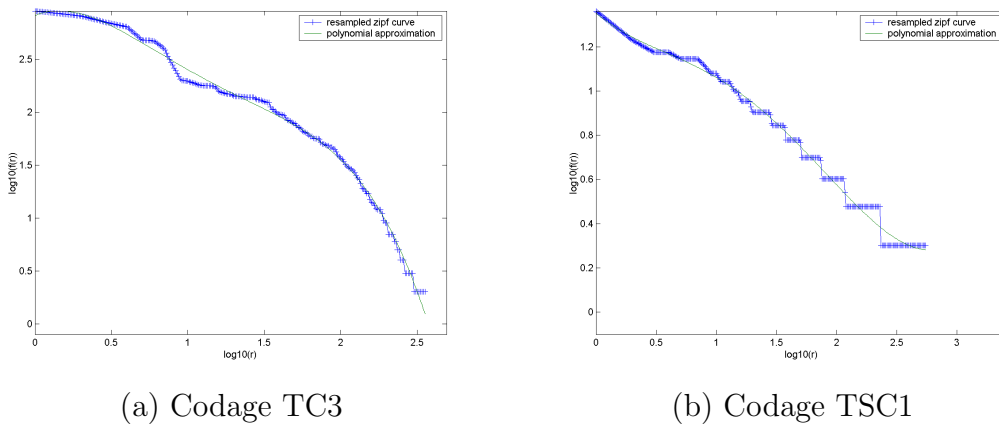


FIG. 5.13 – Exemple de courbes de Zipf rééchantillonnées dans le cas des codages TC3 (a) et TSC1 (b).

L'étude de la courbe de Zipf non rééchantillonnée permet d'évaluer de manière plus importante l'influence des fréquences faibles, pour lesquelles le nombre de mots associés est plus élevé. Du fait du rééchantillonnage à pas constant, la courbe de Zipf rééchantillonnée quant à elle permet une étude plus générale, chaque fréquence ayant la même influence.

La loi de Zipf Inverse, représentée par la relation (5.57), correspond à l'étude de la distribution de la fréquence des mots d'un texte. Cette loi étant également une

loi puissance, elle est représentée dans un espace logarithmique, la courbe correspondante étant appelée courbe de Zipf Inverse. Un exemple est présenté figure 5.14 dans le cas des codages TC3 et TSC1. Seules les fréquences les plus faibles sont considérées (par exemple les 10 plus faibles). Il apparaît dans les cas étudiés que cette courbe est généralement correctement approximée par une droite. Une seule valeur est donc retenue afin de caractériser la courbe de Zipf Inverse : le coefficient directeur de la droite de régression linéaire.

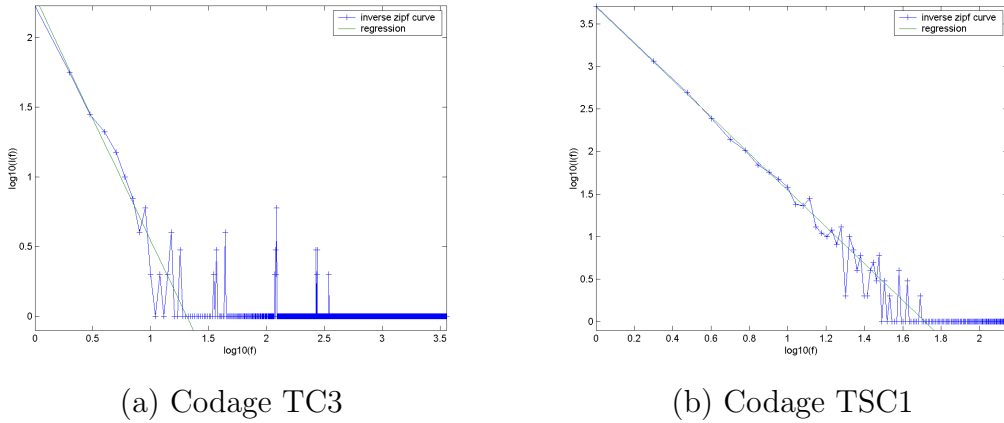


FIG. 5.14 – Exemple de courbes de Zipf Inverse dans le cas des codages TC3 (a) et TSC1 (b).

D'autres primitives sont également calculées. L'entropie liée à la loi de Zipf E_z en fait partie, et s'exprime par la relation (5.64).

$$E_z = - \sum_{r=1}^R \frac{f(r)}{T} \log_R \frac{f(r)}{T} \quad (5.64)$$

où R est le nombre de mots distincts dans le texte contenant un total de T mots. Un logarithme de base R est utilisé afin de normaliser les valeurs, permettant la comparaison de textes.

L'entropie liée à la loi de Zipf Inverse E_i est quant à elle donnée par la relation (5.65).

$$E_i = - \sum_{f=1}^F \frac{I(f)}{R} \log_F \frac{I(f)}{R} \quad (5.65)$$

où F est le nombre total des différentes occurrences. Un logarithme de base F est également utilisé dans un souci de normalisation.

Enfin, les paramètres suivants sont calculés : l'aire sous les courbes de Zipf A_z et de Zipf rééchantillonnée A_{rz} , le nombre de mots distincts R , le nombre total de mots T , le nombre de mots dont la fréquence d'apparition est de 1 $I(1)$, le nombre total d'occurrences différentes F et la fréquence la plus élevée f_m .

Ces primitives vont ensuite être assemblées dans un vecteur entité associé à chaque signal sonore.

L'étape suivante correspond à la classification des signaux afin de déterminer les classes auxquelles les signaux ont la plus grande probabilité d'appartenir.

5.7 Classification des signaux

Quatre méthodes de classification ont été envisagées afin de permettre une représentation des données dans des espaces différents et complémentaires : l'analyse discriminante linéaire (AD) [HAR75], un réseau de neurones de type perceptron (RN) [BRA95, WEN93], un algorithme des k -means (KMEANS) [HAR75], ainsi qu'une méthode de distance aux classes utilisant la distance de Mahalanobis (MAHAL).

Ces espaces de représentation définis par et pour les différents classifieurs permettent une meilleure prise en compte des données du problème grâce à plusieurs points de vue. C'est la raison pour laquelle nous avons également utilisé une méthode de fusion des classifieurs basée sur leur matrice de confusion [LX92] afin de combiner ces méthodes.

Le processus de classification est réalisé sur un sous-ensemble des données, appelé ensemble d'apprentissage. Le taux de reconnaissance réel pour chaque méthode citée précédemment est ensuite obtenu en utilisant le sous-ensemble complémentaire, appelé ensemble de test.

Les quatre méthodes de classification choisies sont caractérisées par le fait que les fonctions de décision sont linéaires pour AD, et non-linéaires pour RN. De plus, la distance utilisée pour KMEANS est euclidienne, ce qui n'est pas le cas pour MAHAL.

Le principe de l'analyse discriminante linéaire est de construire une frontière de décision linéaire entre les classes par la maximisation de la distance inter-classes simultanément à la minimisation de la distance intra-classe.

Le réseau de neurones que nous avons utilisé est un réseau de type perceptron composé d'une couche d'entrée, d'une couche cachée, et d'une couche de sortie. Le nombre d'entrées est égal au nombre de primitives utilisées pour caractériser les éléments. Le nombre de neurones cachés doit être fixé expérimentalement car il dépend des données et du problème. Enfin, le nombre de neurones de sortie est égal au nombre de classes à considérer, chaque classe étant associée à un neurone. L'algorithme d'apprentissage est de type rétro-propagation du gradient, et les fonctions de transfert sont des fonctions de type tan-sigmoïde pour les neurones cachés et linéaire pour les neurones de la couche de sortie.

Le principe de l'algorithme des k -means est de minimiser itérativement jusqu'à convergence la distance euclidienne moyenne entre chaque élément et le centre de la classe la plus proche. La phase d'initialisation qui consiste en l'attribution d'une classe à chaque élément est réalisée ici en utilisant l'ensemble d'apprentissage.

Le principe du classifieur MAHAL est de déterminer le centre de chaque classe et d'affecter les nouveaux éléments à la classe dont le centre est le plus proche en considérant la distance de Mahalanobis (5.66).

$$d_{Mahal}(x, c_i) = (x - c_i)^t \cdot \Sigma_i^{-1} \cdot (x - c_i) \quad (5.66)$$

où c_i est le centre de la classe i dont la matrice de covariance est Σ_i .

Plusieurs techniques permettant la combinaison de classifieurs sont envisageables. Celle que nous avons retenue permet la fusion des classifieurs en considérant leur matrice de confusion [LX92]. Ces matrices représentent les erreurs de classification associées à chaque classifieur. Cette information est utilisée afin de déterminer la classe à laquelle chaque élément a la plus grande probabilité d'appartenir.

5.8 Résultats

Nous avons évalué les différents ensembles de primitives ainsi que les différentes méthodes de classification sur des signaux médicaux qui sont des signaux de bruits xiphoidiens extraits de signaux de déglutition, présentés à la section 1.2.

Pour cette étude, 111 signaux de bruits xiphoidiens ont été traités. Ils ont été enregistrés à partir de 11 patients souffrant de reflux gastro-œsophagien. Pour chacun d'eux, les signaux ont été enregistrés avant et après opération. La classification a été

réalisée par rapport aux ensembles de classes fournis par les médecins et présentés à la section 1.2.2. Les effectifs des différentes classes sont présentés dans le tableau 5.1. De plus, dans chaque cas, 2/3 des signaux ont été utilisés pour l'ensemble d'apprentissage, et 1/3 pour l'ensemble de test.

Ensembles de classes	Primitives 1,4 et 5			Primitives 2 et 3		
	c1	c2	total	c1	c2	total
ba	57	54	111	86	92	178
hiat1	29	23	52	49	21	70
hiat2	29	23	52	42	28	70
oo	30	29	59	62	46	108
vh	34	25	59	72	36	108
od	27	32	59	39	69	108
ao	23	36	59	45	63	108

TAB. 5.1 – Effectifs des différents ensembles constitués des classes c1 et c2, en fonction des types de primitives considérées : bas-niveau (1), cepstre (2), arma (3), ondelettes (4) et Zipf (5).

5.8.1 Analyse bas-niveau des signaux

Les primitives liées à l'analyse bas-niveau des signaux sont extraites à partir des signaux de bruits xiphoidiens considérés dans leur globalité. Ainsi, chaque signal est associé aux 18 primitives qui ont été présentées à la section 5.2. Les résultats obtenus sont donnés TAB. 5.2. Les taux de reconnaissance obtenus par AD, RN, KMEANS et MAHAL ainsi que la fusion de ces classifieurs sont indiqués, en pourcentage, pour les sept ensembles de classes. Le taux maximum de reconnaissance est également présenté.

Ensemble de classes	AD	RN	KMEANS	MAHAL	FUSION	MAX
ba	73,0	64,9	51,4	59,5	73,0	73,0
hiat1	81,8	63,6	54,6	63,6	63,6	81,8
hiat2	63,6	81,8	81,8	45,5	81,8	81,8
oo	66,7	66,7	66,7	66,7	73,3	73,3
vh	53,3	73,3	73,3	66,7	73,3	73,3
od	66,7	80,0	60,0	53,3	86,7	86,7
ao	60,0	86,7	53,3	26,7	80,0	86,7

TAB. 5.2 – Résultat de la classification pour l'ensemble de primitives bas-niveau.

Ces résultats montrent que l'identification des signaux reposant sur les primi-

tives bas-niveau n'est globalement pas très efficace dans le cas de signaux de bruits xiphoïdiens, le taux de reconnaissance moyen étant de 79,6 %. Cependant l'identification des signaux relative aux ensembles de classes od et ao est satisfaisante, avec un taux de 86,7 %.

5.8.2 Analyse cepstrale des signaux

Les primitives issues de l'analyse cepstrale présentée à la section 5.3, ont été extraites en considérant des fenêtres chevauchantes de 0,5 seconde contenant 22050 échantillons, incluses dans les signaux de bruits xiphoïdiens. En effet, il est apparu nécessaire de limiter l'horizon considéré pour l'analyse. Les primitives que nous avons choisies pour représenter chacune de ces fenêtres sont les 20 premiers coefficients cepstraux. Les résultats de la classification sont donnés TAB. 5.3.

Ensemble de classes	AD	RN	KMEANS	MAHAL	FUSION	MAX
ba	62,3	58,0	59,4	58,0	59,4	62,3
hiat1	86,4	81,8	54,6	90,9	90,9	90,9
hiat2	72,7	59,1	50,0	72,7	72,7	72,7
oo	60,5	57,9	57,9	57,9	57,9	60,5
vh	89,5	86,8	65,8	89,5	89,5	89,5
od	68,4	63,2	57,9	65,8	63,2	68,4
ao	76,3	60,5	47,4	84,2	79,0	84,2

TAB. 5.3 – Résultat de la classification pour l'ensemble de primitives liées à une analyse cepstrale.

Le taux de reconnaissance moyen dans ce cas est de 75,5 %, avec cependant une reconnaissance satisfaisante concernant les ensembles de classes vh (89,5 %) et hiat1 (90,9 %).

5.8.3 Modélisation ARMA des signaux

La modélisation ARMA supposant les signaux stationnaires, nous avons réalisé l'hypothèse de stationnarité des signaux des bruits xiphoïdiens sur des fenêtres temporelles de 20 ms, l'ordre de grandeur généralement considéré dans le cas de l'étude de la parole étant de 10 à 30 ms. Chaque fenêtre est représentée par un ensemble de 20 primitives correspondant aux coefficients du modèle ARMA associé. De plus, afin de limiter le nombre de données et de permettre la comparaison des méthodes,

la classification a été réalisée en considérant des zones temporelles de 0,5 seconde, comme dans le cas de l'étude cepstrale, représentées par la moyenne des coefficients des fenêtres incluses dans ces zones. Les résultats de la classification sont donnés TAB. 5.4.

Ensemble de classes	AD	RN	KMEANS	MAHAL	FUSION	MAX
ba	73,9	76,8	60,9	87,0	87,0	87,0
hiat1	86,4	81,8	72,7	81,8	86,4	86,4
hiat2	73,9	56,5	43,5	56,5	73,9	73,9
oo	81,6	47,4	92,1	79,0	92,1	92,1
vh	82,9	82,9	71,4	71,4	71,4	82,9
od	81,6	73,7	65,8	84,2	92,1	92,1
ao	84,2	60,5	86,8	89,5	86,8	89,5

TAB. 5.4 – Résultat de la classification pour l'ensemble de primitives liées à une modélisation ARMA des signaux.

Les résultats obtenus pour la classification à partir des primitives liées à une modélisation ARMA montrent que cette technique est particulièrement bien adaptée aux signaux de bruits xiphoïdiens, puisque le taux de reconnaissance moyen est de 86,3 %, ce qui est supérieur à celui des deux méthodes précédentes. La reconnaissance concernant les ensembles de classes oo et od est particulièrement efficace, avec un taux égal à 92,1 %.

5.8.4 Analyse en ondelettes des signaux

Les primitives issues de l'analyse multirésolution en ondelettes, présentée à la section 5.5, sont calculées sur les signaux de bruits xiphoïdiens dans leur globalité. Chaque signal est donc représenté par 18 primitives, 6 niveaux de décomposition ayant été considérés. Les résultats de la classification sont donnés TAB. 5.5.

Ces résultats montrent que la classification à partir de primitives liées à une analyse par ondelettes est globalement équivalente à celle liée à la structure bas-niveau des signaux, avec un taux de reconnaissance moyen de 78,4 %. Cette méthode est cependant bien adaptée à l'identification de signaux relative à l'ensemble de classes ao, le taux de reconnaissance étant de 93,3 %.

Ensemble de classes	AD	RN	KMEANS	MAHAL	FUSION	MAX
ba	64,9	67,6	48,7	67,6	67,6	67,6
hiat1	63,6	63,6	72,7	63,6	63,6	72,7
hiat2	81,8	63,6	72,7	45,5	72,7	81,8
oo	80,0	73,3	60,0	66,7	73,3	80,0
vh	66,7	73,3	33,3	66,7	66,7	73,3
od	66,7	80,0	46,7	40,0	80,0	80,0
ao	93,3	93,3	40,0	53,3	93,3	93,3

TAB. 5.5 – Résultat de la classification pour l'ensemble de primitives issues de l'analyse en ondelettes.

5.8.5 Analyse des signaux par les lois de Zipf

Nous allons maintenant présenter les résultats que nous avons obtenus en utilisant les primitives issues de l'adaptation des lois de Zipf et Zipf Inverse aux signaux de bruits xiphœdiens à partir des différents codages présentés à la section 5.6. Les résultats de la classification sont indiqués TAB. 5.6, 5.7, 5.8, 5.9, 5.10, 5.11 respectivement pour les codages TC1, TC2, TC3, FC, TSC1 et TSC2. Chaque signal de bruits xiphœdiens est représenté par un ensemble de 18 primitives, puisque seuls les 3 premiers coefficients d'approximation de la courbe de Zipf originale et rééchantillonnée se sont révélés être pertinents.

Ensemble de classes	AD	RN	KMEANS	MAHAL	FUSION	MAX
ba	80,0	83,3	73,3	73,3	90,0	90,0
hiat1	80,0	90,0	80,0	60,0	90,0	90,0
hiat2	70,0	60,0	60,0	30,0	80,0	80,0
oo	92,9	92,9	50,0	57,1	92,9	92,9
vh	57,1	78,6	57,1	64,3	64,3	78,6
od	92,9	85,7	50,0	64,3	92,9	92,9
ao	71,4	85,7	42,9	50,0	85,7	85,7

TAB. 5.6 – Résultat de la classification pour le codage TC1.

Les taux maximums de reconnaissance pour les différents ensembles de classes à partir des différents codages sont présentés TAB.5.12. Cela montre que généralement, un bon taux de reconnaissance est obtenu. En effet, le taux maximum est 92,9% pour les ensembles oo, od et ao, en utilisant soit TC1, soit TC2, soit TSC2. Le taux minimum quant à lui est de 81,8% pour l'ensemble hiat2 avec soit TC2 soit TSC1. Le taux de reconnaissance moyen est de 89,6%, ce qui est supérieur à celui des quatre méthodes présentées précédemment. Ces résultats indiquent que les codages sont

Ensemble de classes	AD	RN	KMEANS	MAHAL	FUSION	MAX
ba	65,6	71,9	59,4	71,9	71,9	71,9
hiat1	54,6	81,8	54,6	63,6	81,8	81,8
hiat2	63,6	63,6	72,7	63,6	81,8	81,8
oo	71,4	50,0	42,9	64,3	64,3	71,4
vh	57,1	85,7	50,0	64,3	64,3	85,7
od	92,9	78,6	78,6	64,3	92,9	92,9
ao	71,4	92,9	71,4	42,9	92,9	92,9

TAB. 5.7 – Résultat de la classification pour le codage TC2.

Ensemble de classes	AD	RN	KMEANS	MAHAL	FUSION	MAX
ba	78,1	75,0	56,3	65,6	75,0	78,1
hiat1	81,8	81,8	72,7	63,6	90,9	90,9
hiat2	54,6	63,6	54,6	45,5	72,7	72,7
oo	85,7	64,3	71,4	57,1	85,7	85,7
vh	71,4	71,4	50,0	64,3	64,3	71,4
od	78,6	85,7	64,3	57,1	85,7	85,7
ao	78,6	85,7	57,1	42,9	85,7	85,7

TAB. 5.8 – Résultat de la classification pour le codage TC3.

Ensemble de classes	AD	RN	KMEANS	MAHAL	FUSION	MAX
ba	59,4	50,0	53,1	53,1	46,9	59,4
hiat1	81,8	81,8	63,6	63,6	81,8	81,8
hiat2	63,6	54,6	54,6	36,4	72,7	72,7
oo	35,7	21,4	57,1	57,1	78,6	78,6
vh	71,4	50,0	57,1	57,1	71,4	71,4
od	64,3	71,4	28,6	42,9	85,7	85,7
ao	64,3	64,3	71,4	50,0	64,3	71,4

TAB. 5.9 – Résultat de la classification pour le codage FC.

Ensemble de classes	AD	RN	KMEANS	MAHAL	FUSION	MAX
ba	84,4	62,5	65,6	56,3	81,3	84,4
hiat1	54,6	72,7	63,6	63,6	72,7	72,7
hiat2	63,6	72,7	63,6	27,3	81,8	81,8
oo	64,3	64,3	57,1	57,1	64,3	64,3
vh	50,0	78,6	42,9	64,3	64,3	78,6
od	57,1	71,4	71,4	42,9	71,4	71,4
ao	64,3	78,6	64,3	21,4	71,4	78,6

TAB. 5.10 – Résultat de la classification pour le codage TSC1.

Ensemble de classes	AD	RN	KMEANS	MAHAL	FUSION	MAX
ba	53,1	43,8	68,8	43,8	68,8	68,8
hiat1	81,8	72,7	63,6	72,7	90,9	90,9
hiat2	72,7	63,6	63,6	45,5	63,6	72,7
oo	71,4	57,1	50,0	57,1	57,1	71,4
vh	71,4	78,6	71,4	64,3	78,6	78,6
od	71,4	64,3	64,3	42,9	64,3	71,4
ao	78,6	92,9	78,6	28,6	92,9	92,9

TAB. 5.11 – Résultat de la classification pour le codage TSC2.

Ensemble	Tmr	Codage(s)
ba	90,0%	TC1
hiat1	90,9%	TC3, TSC2
hiat2	81,8%	TC2, TSC1
oo	92,9%	TC1
vh	85,7%	TC2
od	92,9%	TC1, TC2
ao	92,9%	TC2, TSC2
Taux de reconnaissance moyen	89,6%	

TAB. 5.12 – Taux maximum de reconnaissance (Tmr) pour les différents ensembles de classes.

complémentaires, et que leur efficacité dépend de l'information à considérer.

La bonne classification des bruits xiphoïdiens prouve que, d'une part l'information sur la pathologie obtenue à partir des données médicales peut être extraite des signaux sonores, et que d'autre part les primitives de Zipf sont bien adaptées à ces signaux.

Nous avons également évalué la capacité des primitives liées aux différents codages, à discriminer les bruits xiphoïdiens du bruit de fond, afin de permettre la segmentation des bruits xiphoïdiens dans des signaux de déglutition. Pour cela, deux classes ont été considérées : la classe des bruits xiphoïdiens et la classe du bruit de fond des signaux de déglutition. Les résultats pour les différents codages sont présentés TAB.5.13.

Le bon taux de reconnaissance de l'ensemble de classes *bruits de fond - bruits xiphoïdiens* en considérant le codage TSC1 (89,7 %) suggère que la méthode peut également être envisagée pour la détection de bruits xiphoïdiens dans des signaux bruités de déglutition.

Codage	AD	RN	KMEANS	MAHAL	FUSION	MAX
TC1	56,9	53,9	53,9	49,2	58,5	58,5
TC2	75	71,7	66,7	71,7	76,7	76,7
TC3	57,1	52,4	50,8	63,5	52,4	63,5
FC	52,4	50,8	47,6	52,4	54,0	54,0
TSC1	77,9	89,7	45,6	64,7	89,7	89,7
TSC2	81,0	79,4	49,2	85,7	79,4	85,7

TAB. 5.13 – Discrimination *bruits de fond* - *bruits xiphoïdiens*.

Comme nous l'avons vu précédemment, les différents codages sont complémentaires. Aussi, il peut être intéressant, selon la nature des signaux à étudier, de prendre en considération simultanément l'information issue de plusieurs codages.

Cependant, le nombre de primitives considérées augmente de manière importante, certaines étant de plus inutiles et leur utilisation pouvant éventuellement perturber le processus d'identification. Il peut donc être judicieux de sélectionner les primitives les plus pertinentes, ce qui permet également de réduire la dimension de l'espace de représentation.

Pour cela, plusieurs méthodes de sélection de primitives peuvent être envisagées. Une méthode fréquemment utilisée pour ce problème est l'analyse en composantes principales (ACP) [ESC88].

Son principe est de construire un nouvel espace de représentation à partir de l'espace d'origine. Les composantes du nouvel espace sont obtenues par combinaison linéaire des composantes de l'espace d'origine. Elles doivent satisfaire deux contraintes : maximiser l'inertie de la projection des points dans le nouvel espace, et être orthogonales entre elles. L'avantage de l'ACP est qu'elle permet de réduire la dimension de l'espace de représentation grâce aux nouvelles composantes moins nombreuses et portant plus d'information que les composantes d'origine. Le principal inconvénient est que le résultat peut être perturbé par quelques points extrêmes ayant un comportement atypique.

C'est pourquoi nous nous sommes intéressé aux algorithmes génétiques (AG) [VAF92]. Les AG ont été adaptés à la sélection de primitives depuis quelques années. Il a été montré que dans la plupart des domaines, les AG permettent une réduction significative du nombre de primitives simultanément à une amélioration des taux de reconnaissance.

Le principe des AG est de maintenir pendant plusieurs générations une popula-

tion de taille constante d'individus caractérisés par leur chromosome. Chaque chromosome est représenté par une chaîne de longueur l de gènes binaires. Chaque gène est associé à une primitive de l'espace de représentation d'origine. La valeur de ce gène indique l'inclusion (si sa valeur est 1) ou l'exclusion (si sa valeur est 0) dans le nouvel espace, de la primitive correspondante.

L'évolution de la population est gouvernée par trois règles : la sélection, le croisement et la mutation. La sélection permet l'élimination des individus les moins performants par rapport à un objectif donné. Le croisement permet la création de nouveaux individus par combinaison des chromosomes des deux parents. La mutation est utilisée afin d'éviter une stagnation qui se produirait au cours du processus de recherche, et d'augmenter le domaine d'exploration. Son principe est de modifier de manière aléatoire les gènes d'un individu sélectionné.

La performance d'un individu est donnée par une fonction d'évaluation. Dans le cas d'un problème de classification, il est naturel que la fonction d'évaluation représente le taux de reconnaissance du classifieur associé à l'espace représenté par le chromosome.

Les méthodes de classification envisagées sont celles présentées à la section précédente : l'analyse discriminante linéaire, un réseau de neurones de type perceptron, un algorithme des k -means, ainsi qu'une méthode de distance aux classes utilisant la distance de Mahalanobis.

Au terme de l'évolution de la population, l'individu le plus performant pour chaque classifieur est sélectionné, et les valeurs de ses gènes indiquent les primitives retenues dans les nouveaux espaces de représentation.

Ces espaces de représentation définis par et pour les différents classifieurs permettent une meilleure prise en compte des données du problème grâce à plusieurs points de vue. C'est la raison pour laquelle nous avons également utilisé une méthode de fusion des classifieurs basée sur leur matrice de confusion [LX92] afin de combiner ces méthodes.

Le processus de sélection de primitives est réalisé sur un sous-ensemble de l'ensemble d'apprentissage, appelé ensemble de validation.

Afin d'illustrer l'utilisation de la sélection de primitives par algorithme génétique, nous l'avons appliqué aux signaux présentés précédemment en utilisant simultanément les primitives issues de quelques codages, permettant de considérer des informations de nature différente. Les résultats sont donnés TAB.5.14. np correspond au

nombre de primitives, et *tr* correspond au taux de reconnaissance.

Codages	AD		RN		KMEANS		MAHAL		FUSION	MAX
	np	tr	np	tr	np	tr	np	tr	tr	tr
TC3+TSC1 (vh)	12	60,0	12	90,0	13	40,0	14	60,0	90,00	90,00
TC1+TC2+TC3 (ba)	22	75,0	22	85,0	21	70,0	15	85,0	92,5	92,5

TAB. 5.14 – Résultat de la classification en utilisant un algorithme génétique pour la sélection de primitives.

Ces résultats montrent que l'utilisation d'un algorithme génétique dans le but de permettre la sélection des primitives les plus pertinentes dans le cas de la prise en compte simultanée de plusieurs codages, peut se révéler être intéressante. En effet, en considérant simultanément les primitives issues des codages TC3 et TSC1, le taux de reconnaissance lié à l'ensemble de classes *vh* atteint 90 %, le maximum étant de 85,7 % dans le cas de l'utilisation d'un codage unique. De plus, un taux de reconnaissance de 92,5 % est obtenu pour l'ensemble de classes *ba*, en considérant simultanément les trois codages temporels TC1, TC2 et TC3, contre 90 % lorsqu'un seul codage est considéré.

Cette technique permet donc d'améliorer dans certains cas de manière significative l'identification des signaux, la contrepartie étant un temps de calcul accru du fait de la nécessité de réaliser l'ensemble des codages utilisés.

5.9 Conclusion

Les recherches dans le domaine de l'identification de signaux sonores incluent la définition de primitives qui doivent être bien adaptées au phénomène à étudier, pour lequel nous avons très souvent peu de connaissances.

Dans le cadre de l'analyse de signaux médicaux correspondant à des signaux de bruits xiphoïdiens, nous avons considéré quatre types de primitives reposant sur des informations de nature différente : l'étude de la structure bas-niveau des signaux, l'analyse cepstrale, la modélisation par modèles ARMA et l'analyse en ondelettes. Les résultats ont montré que la modélisation par modèles ARMA est la technique la plus adaptée à ces signaux, avec un taux de reconnaissance moyen de 86,3 % pour les différents ensembles de classes considérés.

Nous avons également élaboré un ensemble de primitives reposant sur les lois de Zipf et Zipf Inverse, et permettant une bonne modélisation de la structure interne

des signaux sonores. Les trois types d'information que nous avons considérés (temporel, fréquentiel et temps-fréquence) contribuent à étendre les propriétés des signaux pouvant être mises en évidence par l'utilisation de plusieurs codages. Les résultats se sont révélés être particulièrement intéressants, puisque le taux de reconnaissance moyen est 89,6 % pour les différents ensembles de classes. De plus, en considérant simultanément les primitives issues des lois de Zipf à partir de plusieurs codages, et en sélectionnant les plus pertinentes par un algorithme génétique, il est possible d'améliorer l'identification.

Le succès de l'application de ces différents codages à un problème difficile d'analyse de signaux sonores médicaux montre que la diversité de l'information considérée permet de prendre en compte les différents aspects d'un même phénomène.

Chapitre 6

Détection d'événements dans des signaux bruités

6.1 Introduction

Nous avons présenté dans le chapitre précédent un ensemble de méthodes ayant pour objectif l'identification du contenu de signaux sonores, la démarche consistant en l'extraction de primitives caractérisant les signaux, puis en la comparaison avec les éléments enregistrés dans une base de données.

Le problème que nous nous posons dans ce chapitre est celui de la détection d'événements particuliers dans des signaux sonores non stationnaires et fortement bruités. La démarche précédente, bien qu'applicable dans ce cas, n'est pas la plus adaptée, ni la plus rapide. En effet, il serait alors nécessaire de découper le signal en fenêtres chevauchantes de faible longueur pour réaliser l'hypothèse de stationnarité et permettre l'invariance par translation assurant une plus grande robustesse. L'extraction des primitives devant être le moins possible sensibles au bruit, sur chaque fenêtre d'observation serait ensuite réalisée et précéderait la phase de classification. Puisque seule la détection et non l'interprétation est nécessaire ici, nous avons développé une méthode rapide ne nécessitant pas l'ensemble de ces étapes, mais plutôt une analyse globale du signal sonore.

Un certain nombre de recherches a déjà été réalisé sur le sujet [DEC96], notamment dans le domaine médical [AKA98, DEL03b], et plus particulièrement sur la détection de spikes dans des signaux électriques afin d'étudier l'activité neu-

ronale [GUT01, HUL02]. Parmi les méthodes souvent utilisées, on peut citer la modélisation ARMA et l'analyse de Fourier présentées respectivement sections 5.4 et 5.5.1. Cependant, ces méthodes présentent des limites importantes lorsque les signaux étudiés ont un comportement hautement non stationnaire, ce qui se révèle être très fréquent. C'est pourquoi, à l'heure actuelle, les méthodes de détection les plus efficaces sont basées sur des méthodes multirésolution par transformée en ondelettes [SZU96] ou par paquets d'ondelettes [GUT01, HUL02]. En effet, ces méthodes permettent, de manière naturelle, l'analyse de signaux localisés en temps et en fréquence. Le problème lié à ces méthodes est de pouvoir déterminer le niveau de résolution le plus pertinent mettant le mieux en évidence l'information recherchée. Les algorithmes proposés sont en général basés sur l'entropie, ce qui est le cas des algorithmes *best basis* et *local discriminant basis* [HUL02]. La détection est ensuite réalisée en sélectionnant les signaux satisfaisant des critères fixés, tels que la durée temporelle, l'énergie, l'amplitude, . . . [GUT01].

Ces méthodes ont l'inconvénient d'être rigides du fait de l'utilisation de seuils, et de ne pas pouvoir s'adapter à tous les types de signaux, l'utilisation de l'entropie n'étant pas toujours pertinente.

La méthode de détection que nous avons développée repose sur l'analyse en ondelettes multirésolution présentée à la section 5.5.1. Nous présenterons dans un premier temps le principe de cette méthode, puis les résultats obtenus dans le cas de l'étude de signaux sonores de déglutition.

6.2 Méthode de détection multirésolution

L'étude de signaux, et particulièrement des signaux médicaux, est connue pour être un difficile problème du fait de leur nature très bruitée pouvant être due notamment aux conditions d'acquisition (bruits de matériels) ou aux conditions physiologiques (parole, respiration, battements cardiaques, . . .). Un exemple de signal sonore bruité présentant un faible rapport signal sur bruit est donné figure 1.3. Il s'agit d'un signal de déglutition contenant des bruits xiphoïdiens produits lors du passage d'un bolus alimentaire dans l'œsophage et le sphincter œsophagien inférieur. L'objectif est alors de détecter les bruits xiphoïdiens au sein des signaux de déglutition.

Afin d'améliorer le rapport signal sur bruit des signaux, des méthodes de filtrage sont généralement utilisées [POR97]. Elles sont souvent efficaces pour diminuer

le bruit et mettre en évidence l'information pertinente contenue dans les signaux. Cependant, elles supposent que certaines propriétés des signaux considérés soient connues, ce qui est rarement le cas.

Ici, nous nous intéressons à l'étude de signaux non stationnaires, ce qui caractérise la majorité des signaux médicaux [GUT01,HUL02]. Les méthodes conventionnelles telles que la Transformée de Fourier [BRI88], les modèles paramétriques [POR97], ou même la Transformée de Fourier à fenêtre glissante [POR97], sont alors inadaptées.

L'analyse multirésolution par ondelettes constitue un puissant outil de traitement du signal, particulièrement bien adapté aux signaux caractérisés par un comportement transitoire et des discontinuités.

6.2.1 Extraction de l'information par analyse multirésolution

L'analyse multirésolution de signaux non stationnaires permet la mise en évidence de discontinuités grâce aux niveaux de détail. Cela est illustré figure 6.1 dans le cas de la décomposition jusqu'au niveau 10 d'un signal de déglutition, qui est le type de signaux sur lesquels est centrée notre étude. Cependant le niveau de détail apportant le plus d'information varie d'un signal à l'autre en fonction du patient mais également de la déglutition, ce qui interdit de fixer le niveau d'analyse. La méthode de détection doit donc être capable de déterminer de manière automatique les niveaux les plus pertinents.

Le principe de la méthode de détection repose sur l'étude des variations temporelles de l'amplitude des niveaux de détail successifs. Pour cela, le signal d'origine est décomposé jusqu'à un niveau L , et chaque signal de détail fait l'objet d'un traitement. Afin de réduire les éléments parasites, les signaux de détail sont décomposés en T fenêtres chevauchantes de longueur N . L'énergie E de chacune de ces fenêtres est ensuite calculée (6.1), permettant de mettre en évidence les variations temporelles de l'amplitude. Un exemple est donné figure 6.2.

$$E(n) = \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} s_n^2(i)}, \quad \forall n \in [1, T] \cap \mathbb{N} \quad (6.1)$$

où $s_n(i)$ est l'amplitude de l'échantillon i de la fenêtre n de taille N .

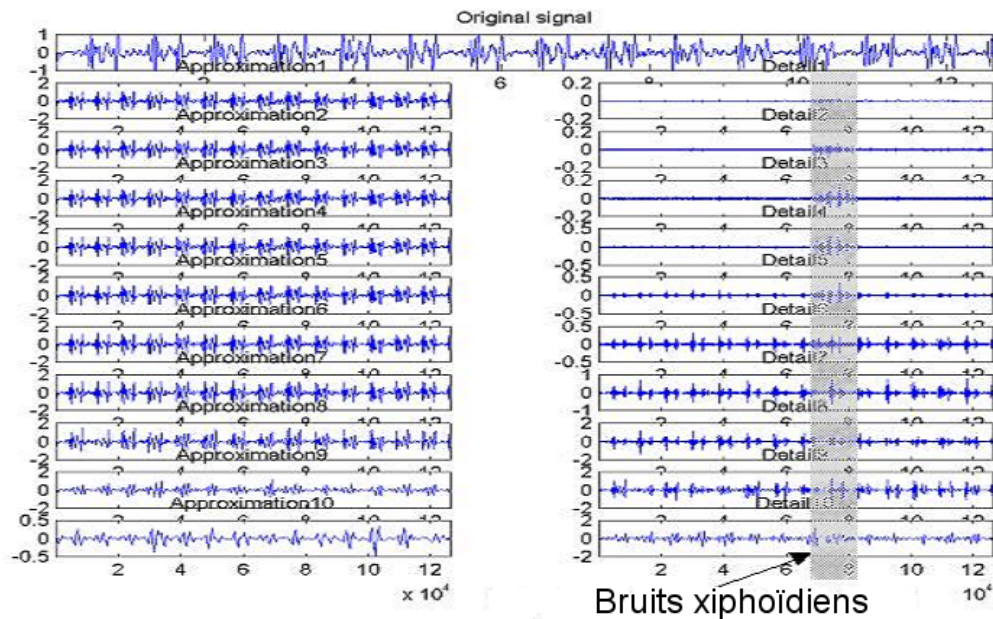


FIG. 6.1 – Décomposition jusqu'au niveau 10 d'un signal de déglutition.

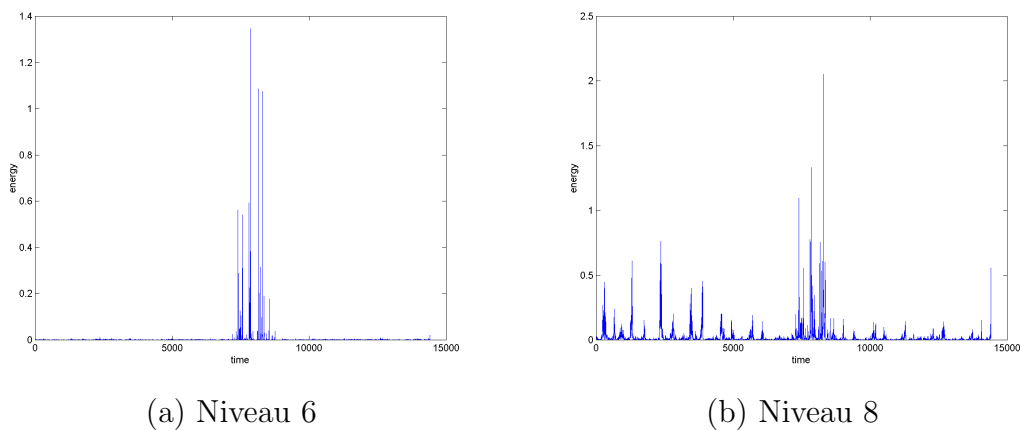


FIG. 6.2 – Exemple de l'énergie d'un signal aux niveaux de détail 6 (a) et 8 (b).

Le problème est alors de sélectionner les niveaux de détail les plus intéressants, c'est-à-dire ceux pour lesquels le bruit est le plus faible et l'information recherchée la mieux mise en valeur. Différents types d'algorithmes ont été développés dans ce but, souvent basés sur le calcul de l'entropie [HUL02]. Cependant, il apparaît que l'évolution de l'entropie des niveaux de détail successifs n'est pas suffisante pour permettre, de manière robuste et systématique, la détection des niveaux pour lesquels les événements sont les plus visibles, comme l'illustre la figure 6.3.

Pour résoudre ce problème, il faudrait pouvoir sélectionner les signaux de détail

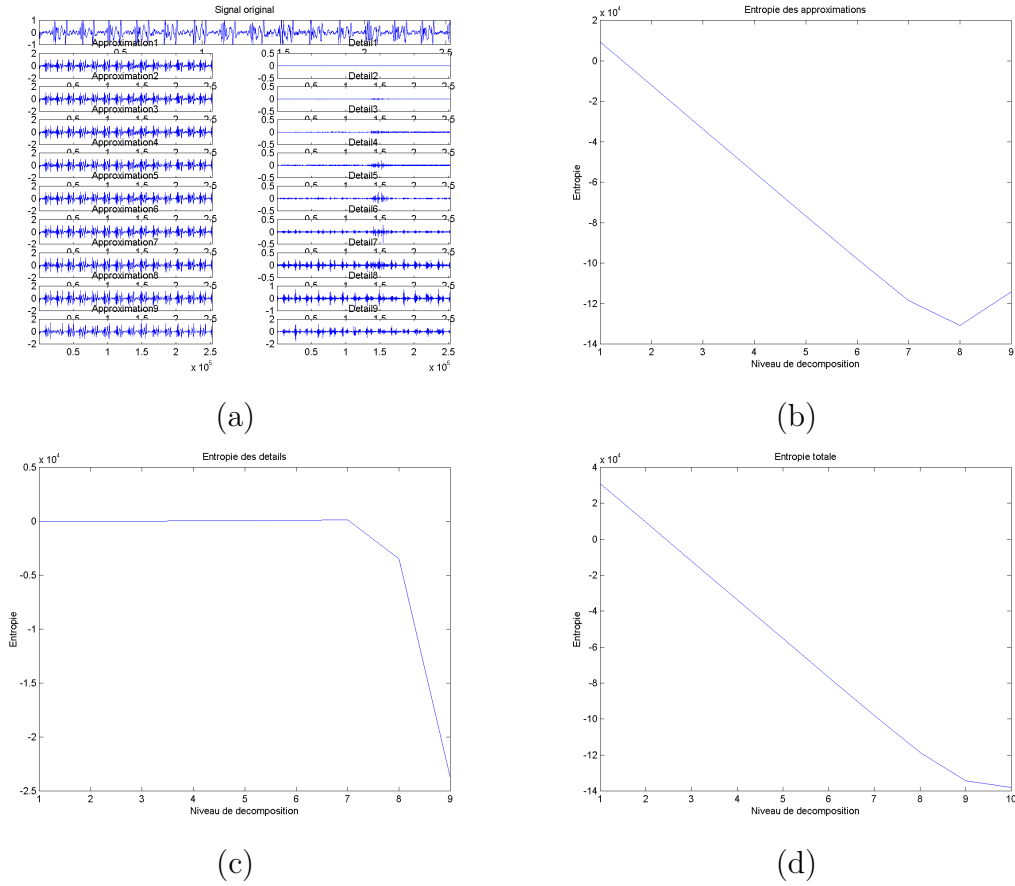


FIG. 6.3 – Décomposition d’un signal (a) et évolution de l’entropie en fonction du niveau de décomposition pour les signaux d’approximation (b), de détail (c) et pour l’entropie totale (d).

caractérisés par de faibles variations de l’énergie dans le temps, à l’exception de courts intervalles temporels, associés aux événements à détecter, que nous appellerons des *pics*.

Ainsi, afin de détecter des pics dans les signaux de détail, la fonction caractéristique B (6.2) est définie. Aux échantillons du signal de détail considéré, correspondant à des pics est associée la valeur 1, et aux autres, la valeur 0.

$$B : [1, T] \rightarrow \{0, 1\}$$

$$B(n) = \begin{cases} 1 & \text{si } E(n) \geq thres \\ 0 & \text{sinon} \end{cases} \quad (6.2)$$

où $thres$ est une valeur de seuil pouvant par exemple être égale à $moyenne(E) + \text{ecart} - type(E)$.

B est la fonction caractéristique d'un ensemble X . L'homogénéisation du signal, correspondant au regroupement des pics proches les uns des autres, est réalisée grâce à une fermeture morphologique Y de X . Finalement, les pics p_i sont les sous-ensembles convexes de Y .

Deux exemples de fermeture morphologique d'un signal sont présentés figure 6.4 par leur fonction caractéristique. Le premier exemple correspond à un niveau de décomposition permettant une bonne mise en évidence de l'événement recherché, alors que dans le deuxième exemple, l'information est noyée dans le bruit.

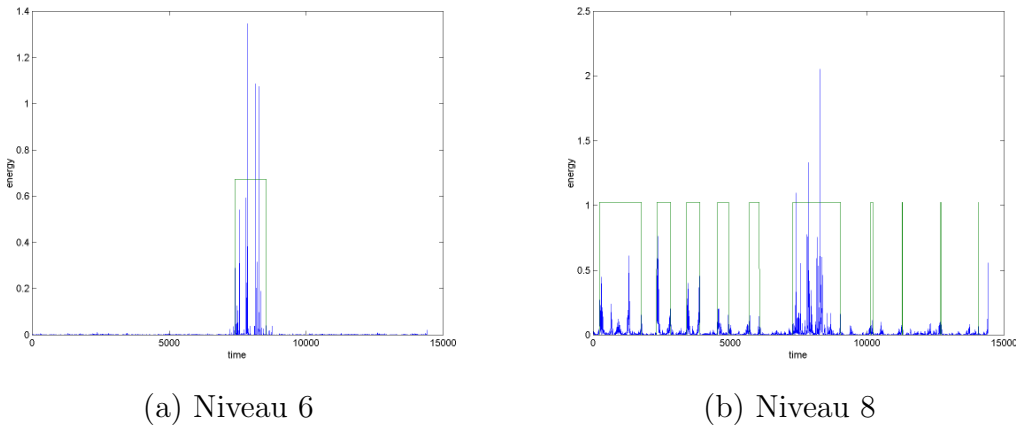


FIG. 6.4 – Exemple de l'énergie des détails de niveaux 6 (a) et 8 (b), ainsi que leur fermeture morphologique.

Afin de sélectionner uniquement les niveaux de détail intéressants, l'activité des signaux de détail est calculée (6.3).

$$A = \frac{\text{cardinal}(Y)}{T} \quad (6.3)$$

Plus l'activité est proche de 1 et plus le signal de détail est bruité. Ainsi, seuls les niveaux de décomposition présentant une faible valeur d'activité sont retenus.

Parmi les pics détectés dans les signaux de détail sélectionnés, il est ensuite nécessaire de déterminer celui (ou ceux) qui correspondent à l'événement recherché. Cela est réalisé grâce à une fonction de décision multicritères.

6.2.2 Fonction de décision multicritères

L'étape précédente a permis de sélectionner les pics potentiels pouvant représenter l'événement à détecter. Afin de déterminer le plus probable, une fonction de décision multicritère est calculée. Ainsi, pour chaque pic, trois critères sont calculés et utilisés par la fonction de décision. Deux critères sont liés aux caractéristiques de la forme du pic sélectionné, et le troisième permet l'introduction d'une connaissance a priori sur l'événement, la chronologie des événements.

Critères

Le premier critère correspond à la valeur d'énergie normalisée NE_i des pics p_i , qui est calculée en considérant la valeur d'énergie la plus élevée des éléments de p_i , normalisée par l'écart-type de l'énergie sur p_i (6.4).

$$NE_i = \frac{\max(E_i(n))}{\text{std}(E_i)} \quad (6.4)$$

où $E_i(n)$ est la valeur d'énergie de l'élément de p_i d'indice n .

Le second critère porte sur l'énergie PE_i des pics p_i , qui est définie par la valeur moyenne de l'énergie des éléments de p_i (6.5).

$$PE_i = \frac{1}{N_i} \sum_{n=0}^{N_i-1} E_i^2(n) \quad (6.5)$$

où N_i est le cardinal du pic p_i .

Le dernier critère permet de prendre en compte une information a priori sur l'événement. La solution retenue ici est de considérer un intervalle temporel, représentant un délai théorique $d_{\text{théorique}}$ entre l'événement précédent connu s'étant déroulé à l'instant tc , et l'événement à détecter.

Ainsi, pour chacun des pics p_i , la différence D_i (6.6) entre le délai observé entre les deux événements et la valeur théorique est calculée. Plus cette différence sera proche de 0, et plus grande sera la probabilité que le pic considéré corresponde à l'événement à détecter.

$$D_i = ||tc - t_{p_i}| - d_{\text{théorique}}| \quad (6.6)$$

où t_{p_i} est l'instant auquel s'est produit l'événement correspondant au pic p_i . Les instants considérés correspondent au moment de début des événements.

Fonction de décision

L'objectif de la fonction de décision est de sélectionner le pic ayant la plus grande probabilité de correspondre à l'événement à détecter. Pour cela, un tri est réalisé afin de classer les pics du plus probable au moins probable pour chacun des critères et de leur attribuer un score (s_{NE} , s_{PE} , s_D). La fonction de décision multicritère est choisie comme la somme pondérée des scores obtenus par les pics pour les trois critères :

$$score(p_i) = w1 * s_{NE}(p_i) + w2 * s_{PE}(p_i) + w3 * s_D(p_i) \quad (6.7)$$

où $w1$, $w2$, $w3$ sont les poids associés aux trois critères, permettant de donner plus d'importance à certains critères.

Finalement, le pic obtenant le score le plus élevé sera considéré comme celui correspondant à l'événement à détecter.

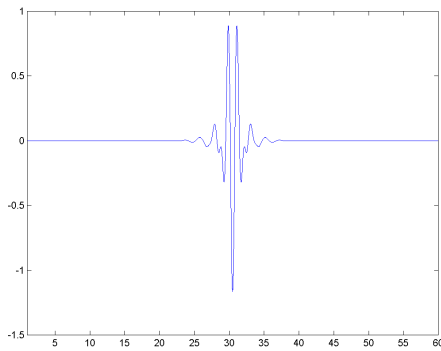
6.3 Résultats

Nous avons appliqué cette méthode de détection à l'analyse de signaux sonores médicaux, décrits dans la section 1.2. L'objectif est de détecter dans des signaux de déglutition les bruits xiphœidiens produits lors du passage d'un bolus baryté dans l'œsophage et le sphincter inférieur de l'œsophage. Un exemple de signal de déglutition contenant des bruits xiphœidiens a été présenté figure 1.3, ainsi que sa décomposition par ondelettes figure 6.1.

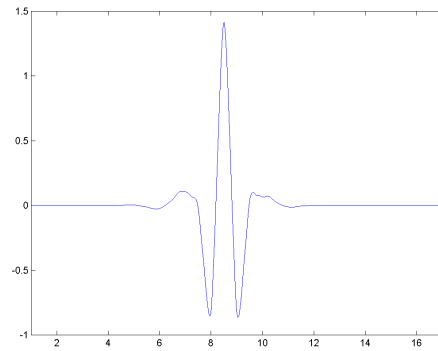
Différents types d'ondelettes ont été évalués, les quatre présentées ici étant les ondelettes de *Meyer*, *Coiflet 3* et *Daubechies 2 et 4*, présentées figure 6.5.

41 signaux enregistrés à partir de 5 individus ont été analysés. La fréquence d'échantillonnage des signaux est de 44100 Hz, et leur durée approximative est de 16 secondes, ce qui correspond approximativement à 700000 échantillons.

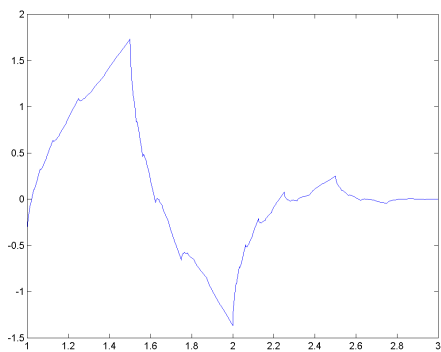
Afin d'évaluer la méthode de détection, les intervalles de bruits xiphœidiens détectés i_m ont été comparés avec ceux fournis par les experts i_e . Pour cela, une



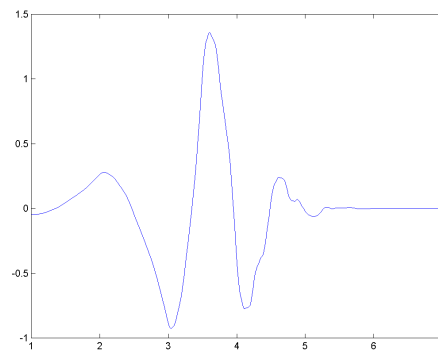
(a) Ondelette de Meyer



(b) Ondelette Coiflet 3



(c) Ondelette de Daubechies 2



(d) Ondelette de Daubechies 4

FIG. 6.5 – Représentation des ondelettes utilisées pour la détection.

mesure de similarité a été définie (6.8). Son principe est d'évaluer l'intersection des deux intervalles. Ainsi, plus les bornes des deux intervalles seront proches, et plus la valeur de la mesure sera proche de 1.

$$d(i_m, i_e) = \frac{\min(\sup(i_e), \sup(i_m)) - \max(\inf(i_e), \inf(i_m))}{\sup(i_e) - \inf(i_e)} \quad (6.8)$$

où i_m et i_e sont respectivement les intervalles donnés par les médecins et par notre méthode. $\sup(i)$ représente la borne supérieure de l'intervalle i , et $\inf(i)$ sa borne inférieure.

La détection est considérée comme réussie lorsque la mesure est supérieure à un seuil. La valeur 0 a été retenue pour nos expérimentations.

Les résultats obtenus pour les différents types d'ondelettes sont donnés TAB. 6.1. Les poids associés aux critères de la fonction de décision ont été choisis de manière à privilégier l'information a priori, puisque les experts considèrent que le temps de

transit moyen, c'est-à-dire le délai entre le bruit cricoïdien correspondant au passage d'un bolus dans le sphincter supérieur de l'œsophage, et le bruit xiphoïdien est en moyenne de 6 secondes. Les valeurs des poids sont donc les suivantes : $w_1 = 1$, $w_2 = 1$, $w_3 = 2$. Il apparaît que l'ondelette de Meyer est la plus adaptée à nos signaux, permettant 80,5% de réussite.

Ondelette	Pourcentage de détection
Daubechies 2	56,1
Daubechies 4	70,7
Coiflet 3	75,6
Meyer	80,5

TAB. 6.1 – Pourcentage de détection pour les ondelettes Daubechies 2 et 4, Coiflet 3 et Meyer.

6.4 Conclusion

La méthode de détection d'événements proposée permet de détecter de manière satisfaisante les bruits xiphoïdiens dans des signaux de déglutition très bruités du fait des conditions d'acquisition. En effet, l'acquisition est perturbée par des facteurs extérieurs tels que les bruits de la caméra à rayons X utilisée simultanément à l'acquisition sonore, des bruits de paroles, de portes, ainsi que par des facteurs physiologiques internes à l'individu tels que les battements cardiaques, la respiration.

L'ondelette de Meyer semble la plus adaptée à ces signaux, permettant de détecter avec succès (80,5 %) les bruits xiphoïdiens dans des signaux de déglutition. Ce résultat est plus faible que celui obtenu au chapitre précédent dans le cadre de la discrimination *bruits de fond - bruits xiphoïdiens*, qui est de 89,7 % pour le codage TSC1. Cependant la méthode présentée ici permet une détection beaucoup plus rapide, de l'ordre de quelques secondes.

Conclusion et perspectives

La problématique à laquelle nous nous sommes intéressé dans le cadre de cette thèse concerne de manière générale l'analyse de séquences multimédia, et plus précisément l'analyse du contenu de séquences multimédia constituées d'images et de données sonores. Notre objectif était d'atteindre un certain niveau de compréhension du contenu global de la séquence par l'étude de la seule information sonore. L'information relative à la séquence d'images est néanmoins extraite dans un premier temps afin de permettre la validation des résultats. Notre travail a ainsi un double objectif : l'analyse de séquences d'images, et l'analyse de données sonores.

Bien que nos travaux tirent leur motivation d'un problème concret auquel se sont trouvés confrontés les experts de la motricité digestive, nous avons tenté de développer des méthodes pouvant s'appliquer à des problématiques moins particulières.

L'analyse des séquences d'images que nous avons retenue consiste à suivre les objets en mouvement afin d'étudier l'évolution de leurs caractéristiques au cours du temps. L'étude des données sonores, quant à elle, correspond à un problème d'identification de classes de signaux sonores.

Nous avons proposé une méthode de suivi d'objets dans une séquence d'images, permettant de suivre des objets déformables de forme et de texture quelconques soumis à un mouvement quelconque pouvant être de forte amplitude, dans des séquences d'images en 256 niveaux de gris. La méthode est constituée d'une phase de prédiction du mouvement de l'objet par filtrage de Kalman afin d'assurer une bonne initialisation, et d'une combinaison de suivi de région par un modèle de régions déformables, et de suivi de contours par un modèle de contours actifs à minimisation locale, dans le but d'optimiser la détection de l'objet dans l'image.

Cette méthode permet un suivi efficace d'objets déformables dans des séquences d'images bruitées mais présente trois limitations : un nombre important de pa-

ramètres doit être réglé empêchant une réelle automatisation du suivi, la gestion du changement de topologie des objets n'est pas assurée, et enfin le coût calculatoire de l'algorithme est élevé du fait des nombreux modules et de l'utilisation de la méthode du recuit simulé comme méthode d'optimisation.

Nous avons donc proposé une méthode de suivi d'objets déformables permettant un suivi rapide d'objets pouvant être soumis à un changement de topologie en réduisant autant que possible le paramétrage de l'algorithme afin de permettre une automatisation du suivi. Cependant, cet algorithme repose sur l'hypothèse que les objets à suivre ont une texture uniforme. La méthode est basée sur un modèle de contours actifs soumis uniquement aux forces de l'image, et dont l'évolution repose sur la théorie des ensembles de niveaux.

L'utilisation de l'une ou l'autre de ces méthodes de suivi est conditionnée par les contraintes du problème considéré. Ces méthodes ont été appliquées avec succès au suivi d'un bolus baryté dans des séquences d'images radiologiques en niveaux de gris. L'objectif était de suivre ce bolus lors de son déplacement à travers l'œsophage et le sphincter inférieur de l'œsophage afin de permettre la caractérisation de ces organes de manière indirecte. Pour cela, un certain nombre de paramètres a été considéré dans le but d'identifier les caractéristiques de ces organes pouvant intervenir dans la pathologie du reflux gastro-œsophagien.

L'analyse des données sonores a consisté en l'élaboration d'un système d'identification de signaux sonores. Deux difficultés principales dans la réalisation de ces systèmes peuvent être notées : d'une part le choix de l'espace de représentation, et d'autre part le choix du processus de classification et de reconnaissance. L'extraction de l'information pertinente doit conduire à un bon choix de primitives, à partir des signaux. Ainsi, dans un premier temps, trois approches ont été considérées en raison de la diversité de l'information possible à mettre en valeur, reposant sur des techniques fréquemment utilisées dans le domaine : l'étude des propriétés bas-niveau des signaux, l'analyse cepstrale et la modélisation ARMA. Afin de compléter ces méthodes, nous avons proposé un ensemble de primitives liées à une analyse multirésolution par ondelettes. De plus, une approche novatrice dans le domaine de l'analyse sonore a été développée. Elle repose sur l'adaptation de l'analyse suivant les lois de Zipf et de Zipf Inverse aux signaux sonores, ces lois étant connues pour modéliser de manière efficace la structure de nombreux phénomènes. Pour cela, trois types de codages des signaux sonores ont été proposés reposant sur leur information temporelle, fréquentielle et temps-fréquence. La phase de classification a été réalisée

en considérant quatre classifieurs différents, ce qui permet une meilleure prise en compte des données du problème grâce à plusieurs points de vue. C'est la raison pour laquelle nous avons également utilisé une méthode de fusion des classifieurs afin de combiner ces méthodes.

Ces différentes méthodes ont été évaluées sur des signaux de bruits xiphoïdiens produits par le passage d'un bolus alimentaire dans l'œsophage et le SIO, enregistrés chez des patients souffrant de la pathologie du reflux gastro-œsophagien. L'objectif était de déterminer si l'information concernant la pathologie obtenue à partir des données médicales pouvait être extraite des signaux sonores. Les résultats ont montré que tel est le cas, et que les primitives issues de l'analyse des signaux sonores suivant les lois de Zipf et de Zipf Inverse sont particulièrement bien adaptées à ce problème, permettant une bonne identification des signaux.

Enfin, nous avons proposé une méthode de détection rapide d'événements dans des signaux bruités, reposant sur une analyse en ondelettes multirésolution, que nous avons appliquée à la détection de bruits xiphoïdiens dans des signaux de déglutition.

Le système d'analyse de séquences multimédia constituées d'images et de données sonores proposé repose sur l'identification du contenu sonore, et peut être validé par des caractéristiques extraites des séquences d'images.

Cette méthode a été appliquée à l'analyse de séquences multimédia médicales dans le but d'extraire l'information relative aux propriétés de la pathologie du reflux gastro-œsophagien afin de fournir aux médecins et aux physiologues experts un outil d'analyse et d'aide à l'interprétation et à la décision dans le cas de l'établissement d'un diagnostic. Les résultats obtenus ont montré la faisabilité de cette approche. Une perspective est donc de travailler à la mise en œuvre à grande échelle du système de manière à produire un outil convivial et utilisable dans un cabinet médical. De plus, une étude approfondie des corrélations entre les phénomènes mis en évidence par l'analyse des images et des données sonores pourra être réalisée afin d'aider à une meilleure compréhension de la pathologie. Enfin, nous comptons valider notre approche dans d'autres domaines ; les vidéos sportives en particulier sont envisagées. Nous pourrions montrer ainsi la capacité de généralisation de notre méthode.

Annexe : Environnement matériel et logiciel

Les séquences multimédia qui ont été le support de notre étude, correspondent à des séquences acoustico-radiologiques acquises en salle radiologie au sein de l'hôpital Trousseau à Tours, dont quelques photographies sont présentées dans la figure 6. Notre objectif était d'acquérir des séquences d'images radiologiques centrées sur la jonction œso-gastrique permettant de visualiser le déplacement d'un bolus baryté, synchronisées avec un signal sonore stéréo contenant les bruits produits par ce bolus lors de son passage aux niveaux des sphincters supérieur et inférieur de l'œsophage.



FIG. 6 – Contexte d'acquisition en salle radiologie de l'hôpital Trousseau (Tours).

Le signal vidéo issu de la caméra radiologique reposant sur un standard propriétaire non conventionnel (constitué d'images de dimension 1112x1024 pixels cadencées à la vitesse de 25 images par seconde) ne pouvait être acquis avec un système d'acquisition grand public. Nous avons donc utilisé une carte d'acquisition vidéo professionnelle de marque Imasys. L'acquisition du signal sonore est, quant-à-elle, réalisée par l'intermédiaire d'une carte son de marque Marian de fréquence d'échantillonnage de chaque canal du signal stéréo de 44,1 KHz. L'ordinateur dédié

à cette tâche d'acquisition devait en outre posséder une quantité de mémoire vive importante (512 Mo), ainsi qu'un disque dur permettant un taux de transfert élevé, à interface SCSI afin de permettre une acquisition continue sans perte d'information.

Afin d'exploiter ces deux cartes, nous avons développé un logiciel d'acquisition permettant l'enregistrement simultané et synchronisé des signaux vidéos et sonores. L'interface logicielle d'acquisition est présentée figure 7.

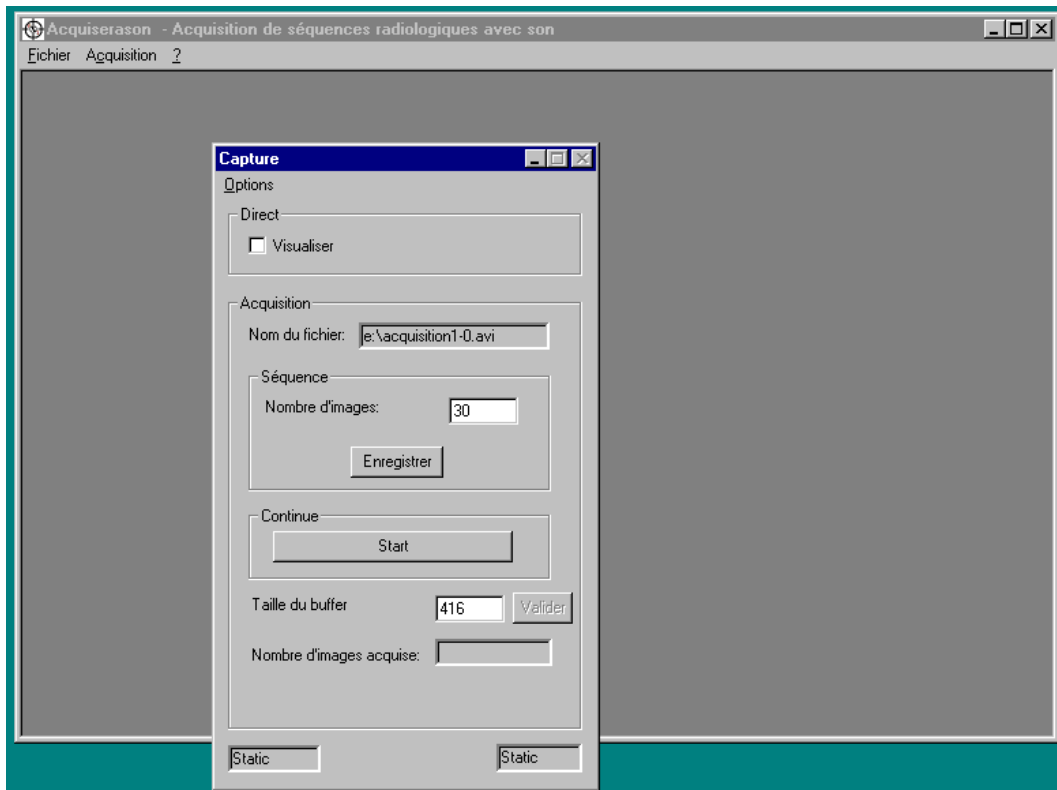


FIG. 7 – Interface logicielle d'acquisition des signaux multimédia.

Dans le but d'analyser ces signaux, les experts médicaux nous ont sollicité afin de développer un logiciel permettant la visualisation simultanée des images radiologiques et des signaux sonores. Cette interface, présentée figure 8, est d'un grand intérêt puisqu'elle permet d'étudier les corrélations entre les phénomènes visibles soit à partir du signal vidéo, soit à partir du signal sonore.

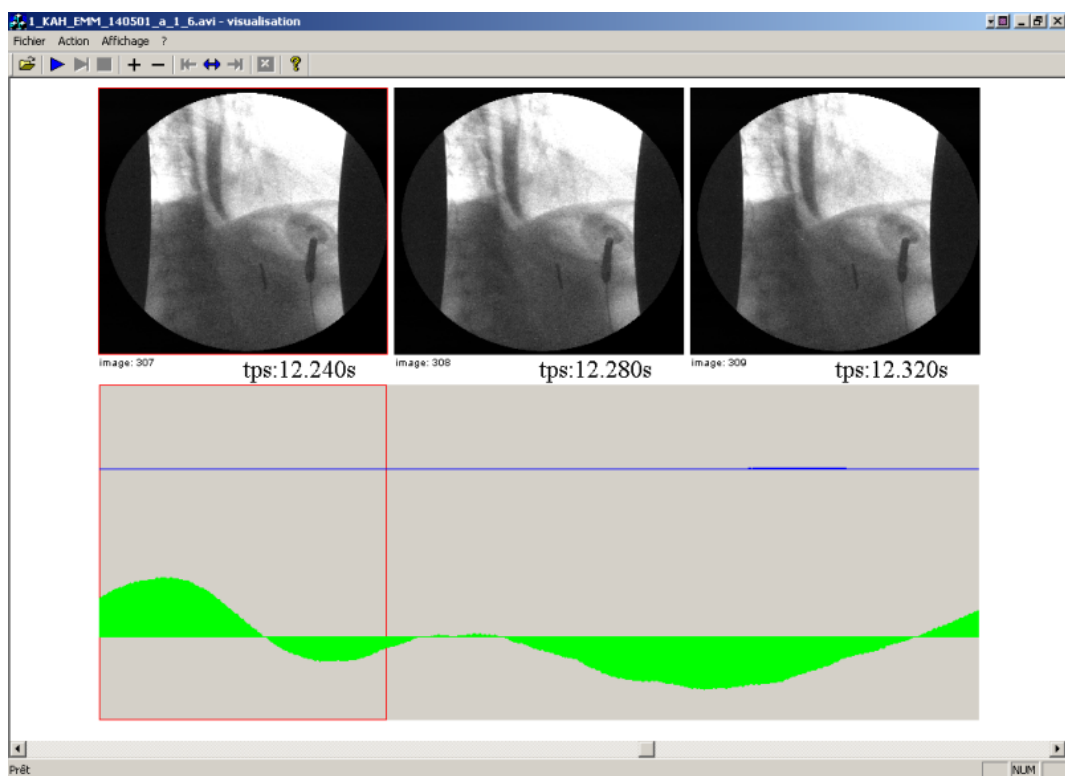


FIG. 8 – Interface logicielle de visualisation des signaux multimédia

Publications

Articles dans des revues internationales

E. Dellandréa, P. Makris, N. Vincent, *Zipf Analysis of Audio Signals*, *Fractals*, World Scientific Publishing Company, à paraître.

Communications dans des congrès internationaux avec comité de lecture

E. Dellandréa, P. Makris, N. Vincent, *Inner structure computation for audio signal analysis*, 3rd International Symposium on Image and Signal Processing and Analysis (ISPA 2003), Rome (Italie), pp. 140-145, Septembre, 2003.

E. Dellandréa, P. Makris, N. Vincent, *Wavelets and Zipf Law for Audio Signal Analysis*, 7th International Symposium on Signal Processing and its Applications (ISSPA 2003), Paris (France), vol. 2, pp. 483-486, Juillet, 2003.

E. Dellandréa, P. Makris, M. Boiron, N. Vincent, *Multiresolution for the detection of xiphoidal sounds in noisy medical audio signals*, 4th EURASIP Conference focused on Video/Image Processing and Multimedia Communication (EC-VIP-MC 2003), Zagreb (Croatie), pp. 619-624, Juillet, 2003.

E. Dellandréa, P. Makris, M. Boiron, N. Vincent, *Xiphoidal Sounds Analysis by the way of Audio Signal Codings and Global Parameters Extraction*, 4th International IEEE EMBS Special Topic Conference on Information Technology Applications in Biomedicine, Birmingham (UK), pp. 9-12, Avril, 2003.

E. Dellandréa, P. Makris, M. Boiron, N. Vincent, *Comparison of Audio Signal Codings for Zipf Analysis*, WSEAS Int. Conference ICAMSL, Tenerife (Espagne),

Décembre, 2002.

E. Dellandréa, P. Makris, M. Boiron, N. Vincent, *A medical acoustic signal analysis method based on Zipf law*, IEEE International Conference on Digital Signal Processing (DSP 2002), Santorin (Grèce), Vol. 2, pp. 615-618, Juillet, 2002.

E. Dellandréa, P. Makris, M. Boiron, N. Vincent, *Active Contours for Bolus Tracking in X-Ray Images Sequences*, IAPR Workshop on Machine Vision Applications (MVA 2000), Tokyo (Japon), pp. 303-306, 2000.

E. Dellandréa, P. Makris, C. Melin, M. Boiron, N. Vincent, *On the Contribution of Zipf Analysis for the Characterization of Medical Audio Signals*, International Conference on Image and Signal Processing (ICISP), Agadir (Maroc), pp. 439-445, Juin, 2003.

Communications dans des congrès nationaux avec comité de lecture

E. Dellandréa, P. Makris, M. Boiron, N. Vincent, *Analyse de signaux sonores par les lois de Zipf et Zipf Inverse*, 19e colloque GRETSI sur le traitement du signal et des images, Paris (France), pp. 387-390, Septembre, 2003.

E. Dellandréa, P. Makris, M. Boiron, N. Vincent, *Suivi d'objets déformables pour l'analyse du sphincter œsophagien inférieur*, Journées Francophones des Jeunes Chercheurs en Vision par Ordinateur (ORASIS), Gérardmer (France), pp. 23-31, Mai, 2003.

Communications dans des groupes nationaux, séminaires, forums

E. Dellandréa, P. Makris, M. Boiron, N. Vincent, *Fusion de données radiologiques et sonores dans l'étude du traitement chirurgical du reflux gastro-œsophagien*, 10ème Forum des jeunes chercheurs Génie Biomédical, Biophysique et traitement d'images, Tours (France), pp. 62-63, Mai, 2000.

Séminaire dans un groupe de travail de l'Université de Tours, mai 2003, *L'utilisation de lois de statistique linguistique pour l'étude de signaux sonores*.

Forum Jeunes Chercheurs de GDR-ISIS, octobre 2002, *L'image ou le son comme aide au diagnostic médical.*

Forum de l'Ecole Doctorale SST de l'Université de Tours, mai 2002, *Analyse de signaux sonores par la loi de Zipf.*

GT4 GDR-ISIS, mars 2001, *Contours actifs appliqués au suivi d'un bolus baryté dans des images radiologiques.*

Forum de l'Ecole Doctorale SST de l'Université de Tours, mai 2001, *Analyse de signaux sonores et radiologiques, appliquée à la compréhension du phénomène de reflux gastro-œsophagien.*

Bibliographie

- [ADA95] D. ADALSTEINSSON et J. SETHIAN. A fast level set method for propagating interfaces. *Journal of Computational Physics*, 118 :269–277, 1995.
- [AKA98] M. AKAY. *Time Frequency and wavelets in biomedical signal processing*. IEEE Press Series in Biomedical Engineering, 1998.
- [AMI90] A.A. AMINI, T.E. WEYMOUTH, et R.C. JAIN. Using dynamic programming for solving variational problems in vision. *IEEE Transactions on pattern analysis and machine intelligence*, 12(9) :855–867, 1990.
- [AO97] R. ANDRE-OBRECHT et B. JACOB. Direct identification vs. correlated models to process acoustic and articulatory informations in automatic speech recognition. *Proc. of ICASSP '97, Munich*, pages 989–992, 1997.
- [ATT95] D. ATTALI. Squelettes et graphes de voronoï 2d et 3d. *Thèse de Doctorat de l'Université Joseph Fourier - Grenoble I*, Octobre 1995.
- [AUB98] G. AUBERT et L. BLANC-FERAUD. An elementary proof of the equivalence between 2d and 3d classical snakes and geodesic active contours. *Rapport de Recherche 3340 - INRIA Sophia Antipolis*, Janvier 1998.
- [AXT01] R.L. AXTELL. Zipf distribution of u.s. firm size. *SCIENCE*, 293 :1818–1820, September 2001.
- [BAR96] F. BARBARESCO, J. LAMBERT, et C. LAMBERT. Segmentation spatio-temporelle et suivi dynamique des fouillis radar. *Traitement du Signal*, 13(5) :473–487, 1996.
- [BAS94a] B. BASCLE, P. BOUTHEMY, R. DERICHE, et F. MEYER. Suivi de primitives complexes sur une séquence d'images. *Rapport de recherche 2428, INRIA-Sophia Antipolis*, Décembre 1994. 22 pages.
- [BAS94b] B. BASCLE et R. DERICHE. Region tracking through image sequences. *Rapport de recherche 2439, INRIA-Sophia Antipolis*, 21 pages, Décembre 1994.
- [BER91] M.O. BERGER. Les contours actifs : modélisation, comportement et convergence. *Thèse de Doctorat de l'Institut National Polytechnique de Lorraine*, Février 1991.
- [BIT98] E. BITTAR. Modèles déformables surfaciques, implicites et volumiques, pour l'imagerie médicale. *Thèse de Doctorat de l'Université Joseph Fourier - Grenoble I*, Mars 1998.

- [BLA98] G. BLANCHET et M. CHARBIT. *Traitement numérique du signal*. Hermes, Paris, 1998.
- [BOI97a] M. BOIRON, F. BAULIEU, P. ROULEAU, L. PICON, et E.H. METMAN. Application de la technique d'enregistrement des bruits de déglutition à l'étude de l'achalasie après dilatation pneumatique progressive. *Arch. Physiol. Bioch.*, 105 :250, 1997.
- [BOI97b] M. BOIRON et P. ROULEAU. Etude préliminaire d'une technique d'enregistrement des bruits de déglutition pour l'exploration du transit pharyngo-œsophagien. *Arch. Physiol. Bioch.*, 105 :249–250, 1997.
- [BOI97c] M. BOIRON, P. ROULEAU, et E.H. METMAN. Exploration of pharyngeal swallowing by audiosignal recording. *Dysphagia*, pages 86–92, 1997.
- [BOI98] M. BOIRON, Z. BENCHELLAL, N. HUTEN, E.H. METMAN, et P. ROULEAU. Nissen fundoplication for gastroesophageal reflux disease (gerd) patients : Interest of swallowing sounds recordings. *Reprints of the 8th World Congress of the International Gastro-Surgical Club, Bismuth, Galmiche, Huguier, Eds, Monduzzi ed (Italy)*, pages 49–52, 1998.
- [BON94] L. BONNAUD et C. LABIT. Etude d'algorithmes de suivi temporel de segmentation basée mouvement pour la compression de séquences d'images. *Publication interne 793, IRISA-Université de Rennes, 43 pages*, Janvier 1994.
- [BON98] L. BONNAUD. Schémas de suivi d'objets vidéo dans une séquence animée : application à l'interpolation d'images intermédiaires. *Thèse de doctorat, IRISA-Université de Rennes*, 1998.
- [BRA95] P.J. BRASPENNING, F. THUIJSMAN, et A.J.M.M. WEIJTERS. *Artificial Neural Networks*. Springer, 1995.
- [BRE99] L. BRESLAU, P. CAO, L. FAN, G. PHILLIPS, et S. SHENKER. Web caching and zipf-like distributions : evidence and implications. *INFO-COM (IEEE Press)*, pages 126–134, 1999.
- [BRI88] E.O. BRIGHAM. *The Fast fourier Transform and its applications*. Prentice-Hall International Editions, 1988.
- [BRI91] F.Y. BRIAND. Conception d'outils de traitement d'images pour l'analyse et l'interprétation de radiographies industrielles. *Thèse de doctorat, Université de Nantes*, 1991.
- [BRO86] T. BROIDA et R. CHELLAPPA. Estimation of object motion parameters from noisy images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-8(1)* :90–99, January 1986.
- [CAM97] J. CAMPBELL. Speaker recognition : A tutorial. *Proceedings of the IEEE*, 85(9) :1437–1462, 1997.
- [CAS93] V. CASELLES, F. CATTE, T. COLL, et F. DIBOS. A geometric model for active contours in image processing. *Numerische Mathematik*, 66 :1–31, 1993.

-
- [CAS97] V. CASELLES, R. KIMMEL, et G. SAPIRO. Geodesic active contours. *International Journal of Computer Vision*, 22(1) :61–79, 1997.
- [CHA97] D. CHARLET et D. JOUVET. Optimizing feature set for speaker verification. *Pattern Recognition Letters*, 18 :873–879, 1997.
- [CHA99] C.P. CHAN, Y.W. WONG, T. LEE, et P.C. CHING. Two-dimensional multi-resolution analysis of speech signals and its application to speech recognition. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 1 :405–408, March 1999.
- [CHA03] A. CHARNOZ, D. LINGRAND, et J. MONTAGNAT. Segmentation du coeur dans des séquences d’images 3d temp par ensemble de niveaux. *Journées Francophones des Jeunes Chercheurs en Vision par Ordinateur (ORASIS)*, pages 15–22, Mai 2003.
- [CHE00] C. CHESNAUD. Techniques statistiques de segmentation par contour actif et mise en œuvre rapide. *Thèse de Doctorat de l’Université de d’Aix-Marseille*, Février 2000.
- [COC95] J.P. COCQUEREZ et S. PHILIPP. Analyse d’images : filtrage et segmentation. *Paris, Masson, 457 pages*, 1995.
- [COH91] L.D. COHEN. On active contour models and balloons. *Computer Vision, Graphics, and Image Processing (CVGIP)*, 53(2) :211–218, 1991.
- [COH97] A. COHEN, R.N. MANTEGNA, et S. HAVLIN. Numerical analysis of word frequencies in artificial and natural language texts. *Fractals*, 5(1) :95–104, 1997.
- [COR87] A. CORANA, M. MARCHESI, C. MARTINI, et S. RIDELLA. Minimizing multimodal variables with the “simulated annealing” algorithm. *ACM Transactions on Mathematical Software*, 13(3) :262–280, Septembre 1987.
- [DAH74] G. DAHLQUIST et A. BJORCK. Numerical methods. *Prentice-Hall, Inc., Englewood Cliffs, NJ*, 1974.
- [DAU90] I. DAUBECHIES. The wavelet transform, time-frequency localization and signal analysis. *IEEE Transactions on Information Theory*, 36(5) :961–1005, 1990.
- [DAU92] I. DAUBECHIES. *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 1992.
- [DEC96] D. DECLERCQ et A. QUINQUIS. *Détection et estimation des signaux*. Hermes, Paris, 1996.
- [DEL00] E. DELLANDREA, P. MAKRIS, M. BOIRON, et N. VINCENT. Active contours for bolus tracking in x-ray images sequences. *IAPR Workshop on Machine Vision Applications (MVA 2000), Tokyo (Japon)*, pages 303–306, Novembre 2000.
- [DEL02a] E. DELLANDREA, P. MAKRIS, M. BOIRON, et N. VINCENT. Comparison of audio signal codings for zipf analysis. *WSEAS Int. Conference ICAMSL, Tenerife (Espagne)*, Décembre 2002.

- [DEL02b] E. DELLANDREA, P. MAKRIS, M. BOIRON, et N. VINCENT. A medical acoustic signal analysis method based on zipf law. *IEEE International Conference on Digital Signal Processing (DSP 2002), Santorin (Grèce)*, 2 :615–618, Juillet 2002.
- [DEL03a] E. DELLANDREA, P. MAKRIS, M. BOIRON, et N. VINCENT. Analyse de signaux sonores par les lois de zipf et zipf inverse. *19e colloque GRETSI sur le traitement du signal et des images, Paris (France)*, pages 387–390, Septembre 2003.
- [DEL03b] E. DELLANDREA, P. MAKRIS, M. BOIRON, et N. VINCENT. Multiresolution for the detection of xiphoidal sounds in noisy medical audio signals. *4th EURASIP Conference focused on Video/Image Processing and Multimedia Communication (EC-VIP-MC 2003), Zagreb (Croatie)*, pages 619–624, Juillet 2003.
- [DEL03c] E. DELLANDREA, P. MAKRIS, M. BOIRON, et N. VINCENT. Suivi d’objets déformables pour l’analyse du sphincter œsophagien inférieur. *Journées Francophones des Jeunes Chercheurs en Vision par Ordinateur (ORASIS), Gérardmer (France)*, pages 23–31, Mai 2003.
- [DEL03d] E. DELLANDREA, P. MAKRIS, M. BOIRON, et N. VINCENT. Xiphoidal sounds analysis by the way of audio signal codings and global parameters extraction. *4th International IEEE EMBS Special Topic Conference on Information Technology Applications in Biomedicine, Birmingham (UK)*, pages 9–12, Avril 2003.
- [DEL03e] E. DELLANDREA, P. MAKRIS, C. MELIN, M. BOIRON, et N. VINCENT. On the contribution of zipf analysis for the characterization of medical audio signals. *International Conference on Image and Signal Processing (ICISP), Agadir (Maroc)*, pages 439–445, Juin 2003.
- [DEL03f] E. DELLANDREA, P. MAKRIS, et N. VINCENT. Inner structure computation for audio signal analysis. *3rd International Symposium on Image and Signal Processing and Analysis (ISPA 2003), Rome (Italie)*, pages 140–145, Septembre 2003.
- [DEL03g] E. DELLANDREA, P. MAKRIS, et N. VINCENT. Wavelets and zipf law for audio signal analysis. *7th International Symposium on Signal Processing and its Applications (ISSPA 2003), Paris (France)*, 2 :483–486, Juillet 2003.
- [DEL03h] E. DELLANDREA, P. MAKRIS, et N. VINCENT. Zipf analysis of audio signals. *Fractals, World Scientific Publishing Company, à paraître*, 2003.
- [DIX00] S. DIXON. On the computer recognition of solo piano music. *Australian Computer Music Conference, Brisbane, Australia*, pages 31–37, 2000.
- [ELL03] C.M. ELLIOTT et V. STYLES. Computations of bidirectional grain boundary dynamics in thin metallic films. *Journal of Computational Physics*, 187(2) :524–543, May 2003.

-
- [ESC88] B. ESCOPIER et J. PAGES. *Analyses factorielles simples et multiples : objectifs, méthodes et interprétation*. Dunod, Paris, 1988.
- [EZZ91] A. EZZAHAR. Estimation et détection d'un signal contaminé par un bruit autorégressif. *Thèse de Doctorat de l'Université Joseph Fourier - Grenoble I*, Octobre 1991.
- [FIS95] S. FISCHER, R. LIENHART, et W. EFFELSBURG. Automatic recognition of film genres. *ACM Multimedia*, pages 295–304, 1995.
- [FOO97] J. FOOTE. Content-based retrieval of music and audio. *Multimedia Storage and Archiving Systems II, Proceedings of SPIE*, pages 138–147, 1997.
- [FOO99] J. FOOTE. An overview of audio information retrieval. *Multimedia Systems*, 7(1) :2–10, 1999.
- [FRA91] E. FRANCOIS. Interprétation qualitative du mouvement à partir d'une séquence d'images. *Thèse de Doctorat de l'Université de Rennes I*, Juin 1991.
- [FV96] S. FERNANDEZ-VIDAL et G. MALANDAIN. Squelettes euclidiens d'objets discrets n-dimensionnels. *Rapport de recherche 2771, INRIA-Sophia Antipolis*, 1996.
- [GAN99] Y. GANDON. Les sources d'images numériques. *Cours de DEA SIBM, U335 & Département d'Imagerie Médicale, Rennes, 7 pages*, 1999.
- [GEL98] M. GELGON. Segmentation spatio-temporelle et suivi dans une séquence d'images : application à la structuration et à l'indexation de vidéo. *Thèse de doctorat, Université de Rennes 1*, 1998.
- [GER98] D. GERHARD. Automatic interval naming using relative pitch. *Bridges : Mathematical Connections in Art, Music and Science. Winfield, Kansas*, pages 37–48, 1998.
- [GER00a] D. GERHARD. Audio signal classification. *Ph.D Thesis of the School of Computing Science, Simon Fraser University*, February 2000.
- [GER00b] D. GERHARD. Audio signal classification : an overview. *Canadian Artificial Intelligence*, 45 :4–6, 2000.
- [GG93] V. GARCIA-GARDUNO et C. LABIT. Suivi de segmentation spatio-temporelle pour la compression de séquences animées. *Quatorzième Colloque GRETSI, Juan-les-Pins*, Septembre 1993.
- [GG95] V. GARCIA-GARDUNO. Une approche de compression orientée-objets par suivi de segmentation basée mouvement pour le codage de séquences d'images numériques. *Thèse de doctorat, Université de Rennes 1*, 1995.
- [GOL01] R. GOLDENBERG, R. KIMMEL, E. RIVLIN, et M. RUDZSKY. Fast geodesic active contours. *IEEE Transactions on Image Processing*, 10(10) :1467–1475, 2001.
- [GUT01] J. GUTIERREZ, R. ALCANTARA, et V. MEDINA. Analysis and localization of epileptic events using wavelet packets. *Medical Engineering & Physics*, 23 :623–631, 2001.

- [HAR75] J.A. HARTIGAN. *Clustering algorithms*. John Wiley & Sons, Inc, 1975.
- [HAS03] H. HASSAN et A.A. FARAG. Cerebrovascular segmentation for mra data using level sets. *International Congress Series*, 1256 :246–252, June 2003.
- [HAV95] S. HAVLIN. The distance between zipf plots. *Physica A216*, pages 148–150, 1995.
- [HEI02] T. HEITTOLA et A. KLAPURI. Locating segments with drums in music signals. *3rd International Conference on Music Information Retrieval (IRCAM)*, October 2002.
- [HUL02] E. HULATA, R. SEGEV, et E. BEN-JACOB. A method for spike sorting and detection based on wavelet packets and shannon’s mutual information. *Journal of Neuroscience Methods*, 117 :1–12, 2002.
- [JAI88] A.K. JAIN et R.C. DUBES. *Algorithms for clustering data*. Prentice Hall Inc., New Jersey, 1988.
- [JAR02] R. JARINA, N. O’CONNOR, et S. MARLOW. Rhythm detection for speech-music discrimination in mpeg compressed domain. *IEEE 14th International Conference on Digital Signal Processing*, July 2002.
- [JB03] S. JEHAN-BESSON. Modèles de contours actifs basés régions pour la segmentation d’images et de vidéos. *Thèse de Doctorat de l’Université de Nice - Sophia Antipolis*, Janvier 2003.
- [KAL60] R.E. KALMAN. A new approach to linear filtering and prediction problems. *Transactions of the ASME- Journal of Basic Engineering*, pages 35–45, March 1960.
- [KAL01] J. KALDA, M. SAKKI, M. VAINU, et M. LAAN. Zipf’s law in human heartbeat dynamics. <http://arxiv.org/abs/physics/0110075>, 2001.
- [KAS98] M. KASS, A. WITKIN, et D. TERZOPOULOS. Snakes : Active contour models. *International Journal of Computer Vision*, 1 :321–331, 1998.
- [KER95] C. KERVRANN. Modèles statistiques pour la segmentation et le suivi de structures déformables bidimensionnelles dans une séquence d’images. *Thèse de Doctorat de l’Université de Rennes I*, Novembre 1995.
- [KLA98] A. KLAPURI. Automatic transcription of music. *Master’s thesis, Tampere University of Technology, Department of Information Technology*, 1998.
- [KOR02] B. KOREN, M.R. LEWIS, E.H. VAN BRUMMELEN, et B. VAN LEER. Riemann-problem and level-set approaches for homentropic two-fluid flow computations. *Journal of Computational Physics*, 181(2) :654–674, September 2002.
- [LI99] W. LI. Zipf’s law. <http://linkage.rockefeller.edu/wli/zipf/>, 1999.
- [LI01] D. LI, I.K. SETHI, N. DIMITROVA, et T. MCGEE. Classification of general audio data for content-based retrieval. *Pattern Recognition Letters*, 22 :533–544, 2001.

-
- [LIU97] Z. LIU, J. HUANG, Y. WANG, et T. CHEN. Audio feature extraction and analysis for scene classification. *IEEE Signal Processing Society, Workshop on Multimedia Signal Processing*, 1997.
- [LOG00] B. LOGAN. Mel frequency cepstral coefficients for music modeling. *first International Symposium on Music Information Retrieval*, 2000.
- [LOW92] D.G. LOWE. Robust model-based motion tracking through the integration of search and estimation. *International Journal of Computer Vision*, 8(2) :113–122, 1992.
- [LU01] L. LU, H. JIANG, et H. ZHANG. A robust audio classification and segmentation method. *ACM Multimedia*, pages 203–211, 2001.
- [LX92] C.Y. SUEN L. XU, A. KRYZAK. Methods of combining multiple classifiers and their application to handwriting recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(3) :418–435, June 1992.
- [MAL89] S. MALLAT. A theory for multiresolution signal decomposition : the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7) :674–693, July 1989.
- [MAL95] R. MALLADI, J.A. SETHIAN, et B.C. VEMURI. Shape modelling with front propagation : A level set approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(2) :158–175, February 1995.
- [MAL96] R. MALLADI et J.A. SETHIAN. A unified approach to noise removal, image enhancement, and shape recovery. *IEEE Transactions on Image Processing*, 5(11) :1554–1568, November 1996.
- [MAL98a] G. MALANDAIN et S. FERNANDEZ-VIDAL. Euclidean skeletons. *Image and Vision Computing*, 16(5) :317–327, april 1998.
- [MAL98b] R. MALLADI et J.A. SETHIAN. A real-time algorithm for medical shape recovery. *International Conference on Computer Vision*, 5(11) :304–310, January 1998.
- [MAL03] N. MALAYATH et H. HERMANSKY. Data-driven spectral basis functions for automatic speech recognition. *Speech Communication*, 40 :449–466, 2003.
- [MAN02] A.R. MANSOURI. Region tracking via level set pdes without motion computation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7) :947–961, July 2002.
- [MAR96] K. D. MARTIN. Automatic transcription of simple polyphonic music : Robust front end processing. *M.I.T. Media Lab Perceptual Computing Technical Report 399*, November 1996.
- [MAR98] M. MARSILI et Y.C. ZHANG. Interacting individuals leading to zipf law. *PHYSICAL REVIEW LETTERS*, 80(12) :2741–2744, July 1998.
- [MEY93] Y. MEYER. *Wavelets : algorithms and applications*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 1993.

- [MEY94] F.G. MEYER et P. BOUTHEMY. Region-based tracking using affine motion models in long image sequences. *CVGIP :Image Understanding*, 60(2) :119–140, September 1994.
- [MOK96] C. MOKBEL, D. JOUVET, et J. MONNE. Deconvolution of telephone line effects for speech recognition. *Speech Communication*, 19 :185–196, 1996.
- [NAS93] C. NASTAR et N. AYACHE. Fast segmentation, tracking, and analysis of deformable objects. *Proc. of 4th Int. Conf. on Computer Vision (ICCV'93), Berlin, Germany*, pages 275–279, May 1993.
- [NIC92] H. NICOLAS. Hiérarchie de modèles de mouvement et méthodes d'estimation associées. application au codage de séquences d'images. *Thèse de doctorat, IRISA-Université de Rennes 1*, 1992.
- [NIC93] H. NICOLAS et C. LABIT. Motion and illumination variation estimation using a hierarchy of models : application to image sequence coding. *Publication interne 742, IRISA*, 1993.
- [NIL03] B. NILSSON et A. HEYDEN. A fast algorithm for level set-like active contours. *Pattern Recognition Letters*, 24(9-10) :1331–1337, June 2003.
- [OSH88] S. OSHER et J.A. SETHIAN. Fronts propagating with curvature-dependent speed : Algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics*, 79 :12–49, 1988.
- [OSH01a] S. OSHER et R.P. FEDKIW. Level set methods : An overview and some recent results. *IPAM GBM Tutorials*, March-April 2001.
- [OSH01b] S. OSHER et R.P. FEDKIW. Level set methods : An overview and some recent results. *Journal of Computational Physics*, 169 :463–502, 2001.
- [PAP95] A.P. PAPLINSKI et J.F. BOYCE. An implementation of the active contour method for noisy images using a local minimization algorithm. *Technical report 95-1, Faculty of Computing and Information Technology*, 1995.
- [PAR98] N.K. PARAGIOS et R. DERICHE. A pde-based level set approach for detection and tracking of moving objects. *IEEE International Conference on Computer Vision*, pages 1139–1145, 1998.
- [PAR99] N. PARAGIOS et R. DERICHE. Geodesic active regions for motion estimation and tracking. *Rapport de recherche 3631, INRIA-Sophia Antipolis*, 1999.
- [PAR00] N.K. PARAGIOS. Geodesic active region and level set methods : Contribution and applications in artificial vision. *Thèse de doctorat, Sophia Antipolis*, Janvier 2000.
- [PAT96] N.V. PATERL et I.K. SETHI. Audio characterization for video indexing. *Storage and Retrieval for Image and Video Databases (SPIE)*, pages 373–384, 1996.

-
- [PAT97] N.V. PATEL et I.K. SETHI. Video classification using speaker identification. *Storage and Retrieval for Image and Video Databases (SPIE)*, pages 218–225, 1997.
- [PAT99] C.S. PATTICHIS et A.G. ELIA. Autoregressive and cepstral analyses of motor unit action potentials. *Medical Engineering & Physics*, 21 :405–419, 1999.
- [PEL02] V. PELTONEN, J. TUOMI, A. KLAPURI, J. HUOPANIEMI, et T. SORSA. Computational auditory scene recognition. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'2002)*, May 2002.
- [PEN95] C.K. PENG et AL. Statistical properties of dna sequences. *Physica A* 221, pages 180–192, 1995.
- [PEN99] D. PENG, B. MERRIMAN, S. OSHER, H. ZHAO, et M. KANG. A pde-based fast local level set method. *Journal of Computational Physics*, 155 :410–438, 1999.
- [PFE96] S. PFEIFFER, S. FISCHER, et W. EFFELSBERG. Automatic audio content analysis. *Tech. Rep. TR-96-008, University of Mannheim, D-68131 Mannheim, Germany*, 1996.
- [POO95] V. POOSALA. Zipf's law. *Technical report, University of Wisconsin, Madison*, 1995.
- [POR97] B. PORAT. *A course in Digital Signal Processing*. John Wiley & Sons, Inc, 1997.
- [RIC97] Y. RICQUEBOURG. Analyse de mouvements articulés : mesure et suivi 2d ; application à la télésurveillance. *Thèse de Doctorat de l'Université de Rennes I*, Janvier 1997.
- [ROA01] M. ROACH et J. MASON. Classification of video genre using audio, 2001.
- [ROS95] P.L. ROSIN et G.A.W. WEST. Saliency distance transforms. *Graphical Models and Image Processing*, 57(6) :483–521, November 1995.
- [SCH97] E. SCHEIRER et M. SLANEY. Construction and evaluation of a robust multifeature speech/music discriminator. *ICASSP, Munich, Germany*, pages 1331–1334, 1997.
- [SCH00] S. SCHUPP. Prétraitement et segmentation d'images par mise en œuvre de techniques basées sur les équations aux dérivées partielles : application en imagerie microscopique biomédicale. *Thèse de Doctorat de l'Université de Caen*, Décembre 2000.
- [SET96] J.A. SETHIAN. A fast marching level set method for monotonically advancing fronts. *Proc. Nat. Acad. Sci.*, 93(4) :1591–1595, 1996.
- [SET98] J.A. SETHIAN. Fast marching methods and level set methods for propagating interfaces. *Von Karman Institute Lecture Series, Computational Fluid Mechanics*, 1998.

- [SOR70] H.W. SORENSON. Least-squares estimation : from gauss to kalman. *IEEE Spectrum*, 7 :63–68, July 1970.
- [STA95] M.H.R. STANLEY, S.V. BULDYREV, S. HAVLIN, R.N. MANTEGNA, M.A. SALINGER, et H.E. STANLEY. Zipf plots and the size distribution of firms. *Economics Letters*, 49 :453–457, 1995.
- [STR97] G. STRANG et T. NGUYEN. *Wavelets and Filter Banks*. Wellesley-Cambridge Press, 1997.
- [SUR01] J. SURI, K. LIU, S. SINGH, S. LAXMINARAYANA, et L. REDEN. Shape recovery algorithms using level sets in 2-d and 3-d medical imagery : A state-of-the-art review. *IEEE Trans. in Information Technology in Biomedicine (ITB)*, 2001.
- [SUS94] M. SUSSMAN et P. SMEREKA adn S. OSHER. A level set approach for computing solutions to incompressible two-phase flow. *Journal of Computational Physics*, 114 :146–159, 1994.
- [SUS03] M. SUSSMAN. A second order coupled level set and volume-of-fluid method for computing growth and collapse of vapor bubbles. *Journal of Computational Physics*, 187(1) :110–136, May 2003.
- [SZU96] H. SZU, B. TELFER, et J. GARCIA. Wavelet transforms and neural networks for compression and recognition. *Neural Networks*, 9(4) :695–708, 1996.
- [THO02] D.L. THOMSON et R. CHENGALVARAYAN. Use of voicing features in hmm-based speech recognition. *Speech Communication*, 37 :197–211, 2002.
- [TRU98] F. TRUCHETET. *Ondelettes pour le signal numérique*. Hermes, Paris, 1998.
- [TZA01] G. TZANETAKIS, G. ESSL, et P. COOK. Automatic musical genre classification of audio signals. *Internatinal Symposium on Music Information Retrieval (ISMIR)*, pages 205–210, October 2001.
- [TZA03] G. TZANETAKIS, A. ERMOLINSKYI, et P. COOK. Pitch histograms in symbolic and audio music information retrieval. *Journal of New Music Research*, to appear, 2003.
- [VAF92] H. VAFAIE et K. DE JONG. Genetic algorithms as a tool for feature selection in machine learning. *4th International Conference on Tools with Artificial Intelligence*, pages 200–204, 1992.
- [VER99] J. VERMAAK, C. ANDRIEU, A. DOUCET, et S.J. GODSILL. Non-stationary bayesian modelling and enhancement of speech signals. *Technical Report CUED/F-INFENG/TR.351, Signal Processing Group, Cambridge University Engineering Department*, 1999.
- [VM01] S. VENEGAS-MARTINEZ. Analyse et segmentation de séquences d’images en vue d’une reconnaissance de formes efficace. *Thèse de Doctorat de l’Université René Descartes - Paris V*, 2001.

-
- [WAT02] T. WATANABE, K. SUGAWARA, et H. SUGIHARA. A new pattern representation scheme using data compression. *IEEE Transactions On Pattern Analysis and Machine Intelligence*, 24(5) :579–590, May 2002.
- [WEL97] G. WELCH et G. BISHOP. An introduction to the kalman filter. http://www.cs.unc.edu/~welch/kalman/kalman_filter/kalman.html, 1997.
- [WEN93] D. WENZEK. Construction de réseaux de neurones. *Thèse de doctorat de l'Institut National Polytechnique de Grenoble*, Septembre 1993.
- [WIL92] D.J. WILLIAMS et M. SHAH. A fast algorithm for active contours and curvature estimation. *CVGIP : Image Understanding*, 55(1) :14–26, January 1992.
- [XU91] C. XU et J.L. PRINCE. Gradient vector flow : A new external force for snakes. *Computer Vision and Pattern Recognition(CVPR '97)*, pages 211–218, June 1991.
- [XU00] C. XU, D.L. PHAM, et J.L. PRINCE. Image segmentation using deformable models. *SPIE Handbook on Medical Imaging – Volume III : Medical Image Analysis*, edited by J.M. Fitzpatrick and M. Sonka, pages 129–174, May 2000.
- [YAS99] A. YASMIN. Speech enhancement using voice source models. *PhD Thesis of the University of Waterloo, Ontario, Canada*, 1999.
- [ZHA96] H.K. ZHAO, T. CHAN, B. MERRIMAN, et S. OSHER. A variational level set approach to multiphase motion. *Journal of Computational Physics*, 127(1) :179–195, August 1996.
- [ZHA98a] T. ZHANG et C. KUO. Content-based classification and retrieval of audio. *SPIE's 43rd Annual Meeting - Conference on Advanced Signal Processing Algorithms, Architectures, and Implementations VIII, San Diego*, July 1998.
- [ZHA98b] T. ZHANG et C.C.J. KUO. Hierarchical system for content-based audio classification and retrieval. *Multimedia Storage and Archiving Systems III, Proceedings of SPIE*, 3527 :398–409, 1998.
- [ZHA98c] H.K. ZHAO, B. MERRIMAN, S. OSHER, et L. WANG. Capturing the behavior of bubbles and drops using the variational level set approach. *Journal of Computational Physics*, 143(2) :495–518, July 1998.
- [ZIP49] G.K. ZIPF. *Human Behavior and the Principle of Least Effort*. Addison-Wesley Press, 1949.

Analyse de signaux vidéos et sonores : application à l'étude de signaux médicaux

Résumé : La problématique considérée concerne l'étude de séquences multimédia constituées d'images et de sons dont il s'agit d'étudier les corrélations de manière à aider à la compréhension de l'origine des bruits. L'analyse des séquences d'images consiste à suivre les objets en mouvement de manière à permettre leur étude. Une méthode générique, reposant sur une combinaison de suivi de régions et de contours, et une méthode adaptée aux objets homogènes, reposant sur la théorie des ensembles de niveaux, sont proposées. L'analyse des données sonores consiste en l'élaboration d'un système d'identification reposant sur l'étude de la structure des signaux grâce à des codages adaptés et à leur modélisation par les lois de Zipf. Ces méthodes ont été évaluées sur des séquences acoustico-radiologiques dans le cadre de l'étude de la pathologie du reflux gastro-oesophagien, en collaboration avec l'équipe *Acoustique et Motricité Digestive* de l'Université de Tours.

Mots-clés : Séquences multimédia, acquisitions acoustico-radiologiques, suivi d'objets, modèle de régions déformables, modèle de contours actifs, ensembles de niveaux, ondelettes, codages, extraction de primitives, classification, modélisation par lois puissance, détection d'événements, reconnaissance d'événements sonores.

Video and audio signal analysis : application to the study of medical signals

Abstract : The work deals with the study of multimedia sequences containing images and sounds. The analysis of images sequences consists in the tracking of moving objects in order to allow the study of their properties. The investigations have to enable the understanding of sounds when correlated to events in the image sequence. One generic method, based on the combination of regions and contours tracking, and one method adapted to homogeneous objects, based on level set theory, are proposed. The analysis of audio data consists in the development of an identification system based on the study of the structure of signals thanks to their coding and Zipf laws modeling. These methods have been evaluated on medical sequences within the framework of the gastro-oesophageal reflux pathology study, in collaboration with the *Acoustique et Motricité Digestive* research team of the University of Tours.

Keywords : Multimedia sequences, objects tracking, deformable regions model, active contours model, level set, wavelets, codings, feature extraction, classification, power law modeling, events detection, audio events recognition.

Laboratoire d'Informatique (EA2101) - Université de Tours
Equipe Reconnaissance des Formes et Analyse d'Images
64, avenue Jean Portalis, 37200 Tours
<http://www.li.univ-tours.fr>

