Du mouvement dans les images,
ou de l'algorithme à la cadence vidéo.

Mémoire d'Habilitation à Diriger des Recherches.

Franck Luthon

Laboratoire des Images et des Signaux
LIS, INPG, 46 av. Félix-Viallet, 38031 Grenoble Cedex, France
Courriel: Franck.Luthon@inpg.fr

Jury

Pr. Michel Baribaud
Patrick Bouthémy, rapporteur
Pr. Alain Chéhikian, rapporteur
Pr. Jean-Louis Lacoume
Pr. Henri Maître, rapporteur
Jean-Luc Schwartz
Patrice Senn

19 Janvier 1999

Et, pour tout dire l -dessus,
il n'y a point d'images,
il n'y a que des objets imaginaires.

Alain ([Ala41], p. 65)

A mes fils.

Sommaire

Liste des Figures

Liste des Tableaux

Avant-Propos

Je remercie dans leur ensemble tous les membres permanents ainsi que les thésards et stagiaires du laboratoire pour l'ambiance de travail agréable qui y règne. Mais je tiens à mentionner plus particulièrement Marie-Noëlle Matraire et Michel Crépin-Jourdan, qui sont en quelque sorte la mémoire vivante et les racines humaines du laboratoire, sans lesquelles le chercheur aurait trop facilement tendance à se déconnecter du réel. Leur simple présence enracine le laboratoire dans l'humanité. Je les remercie en outre pour tous les menus services rendus au cours de ces huit années.

Mes remerciements vont aussi aux membres du jury qui ont accepté d'évaluer mon travail. La constitution de ce jury tente de couvrir les différentes facettes de mon métier d'enseignant-chercheur :

Michel Baribaud, directeur de l'ENSERG, est coupable de m'avoir introduit dans la spirale infernale de l'enseignement supérieur. Il est, à ce titre, bien placé pour juger de mon activité pédagogique.

Patrick Bouthémy était bien sûr incontournable, puisqu'il est selon moi "Monsieur Champ de Markov" dans la communauté scientifique française. Il était un rapporteur tout désigné sur ce thème.

Alain Chéhikian, ex-chef de l'ex-équipe Vision de l'ex-LTIRF (Laboratoire de Traitement d'Images et Reconnaisance de Formes) et ex-directeur de ce laboratoire, m'a laissé toute liberté pour développer mon projet de recherche lors de mon arrivée au sein du laboratoire. Il m'a simplement proposé une orientation initiale de recherche, qui portait sur l'extraction de primitives spatio-temporelles dans les séquences d'images, sans m'imposer ni méthodes de traitement, ni objectifs, ni applications. Il m'a dès le départ fait confiance (ou a peut-être compris très tôt mon caractère indépendant) et je l'en remercie.

Jean-Louis Lacoume, qui fut mon directeur de thèse, m'inspire respect pour sa compétence en traitement du signal et son "humanisme scientifique". Mon seul regret est de ne pas avoir davantage fait appel à lui au cours de ma thèse, pour apprendre bien plus à ses côtés. Nous retrouvant dans le même laboratoire LIS (Laboratoire des Images et des Signaux), j'entends bien combler cette lacune. Je prends également plaisir, depuis 1995, à participer aux réunions du comité de rédaction de la revue Traitement du Signal qu'il dirige.

Henri Maître a également accepté d'être rapporteur et je l'en remercie sincèrement. Il est l'œil extérieur qui pourra scruter mon travail en toute indépendance et impartialité.

Un grand merci à Jean-Luc Schwartz de l'ICP (Institut de la Communication Parlée) pour avoir accepté bien volontiers d'évaluer ma contribution au projet Labiophone de la fédération ELESA. Sa présence dans ce jury atteste des collaborations inter-laboratoires en cours.

Patrice Senn du CNET/CNS sera mon juge en ce qui concerne les mises en œuvre matérielles des algorithmes que j'ai développés. Il pourra notamment critiquer mes trop modestes compétences en microélectronique, que je souhaite développer dans l'avenir. Sa présence dans ce jury reflète les collaborations en cours ou à venir entre le laboratoire et le monde industriel.

Enfin, je remercie mes thésards et mes stagiaires, puisqu'ils ont été la cheville ouvrière de mon travail de recherche, et que sans eux, mon travail n'aurait pas atteint son stade actuel.

Ce mémoire se divise en deux parties. La permière partie retrace dans sa globalité mon parcours de recherche en recensant mes encadrements et publications. J'essaie d'en souligner les points principaux et les enseignements que j'en ai tirés. Cette partie se veut accessible à un large public.

La deuxième partie est plus technique : elle détaille les principaux thèmes de mes activités de recherche au LTIRF depuis ma nomination en tant que maître de conférences, en les situant par rapport à l'état de l'art. Elle s'achève par une vision en perspective de l'évolution de mes thèmes de recherche au sein du nouveau laboratoire LIS.

Enfin, les annexes résument mon cursus, mes travaux d'encadrement de recherche et de publications. Six papiers représentatifs de mon activité sont joints en fin du document.

Partie I

Un Parcours SIP

Chapitre 1

Un Peu d'Histoire

1.1 Des Origines

C'est en 1983-1984, alors que j'étais élève-ingénieur en deuxième année à l'ENSERG (Ecole Nationale Supérieure d'Electronique et Radioélectricité de Grenoble), que j'ai entendu parler pour la première fois de traitement du signal et de théorie de l'information : Shannon et Fourier faisaient irruption dans ma vie, de façon irréversible.

Avec une note de 9/20 à l'examen de traitement de signal, je pouvais dire que j'étais passé à côté de l'essentiel et n'avais pas compris grand-chose à cette matière. Grâce à mon bagage en mathématiques, j'avais assuré une petite moyenne, mais la signification des grandeurs que je manipulais m'était restée confuse.

Néanmoins, le traitement du signal, et plus généralement le traitement de l'information, m'avait séduit dès l'abord, malgré ma médiocre performance. Les nombreux secteurs de l'ingénierie où intervient un besoin en traitement du signal, typiquement pour éliminer du bruit toujours présent (*filtrage*), extraire les fréquences constitutives des signaux (*analyse spectrale*) ou toute autre information pertinente (*reconnaissance*), ou encore compresser ou transmettre l'information (*codage*), d'une part, et l'outil mathématique sous-jacent, d'autre part, étaient pour moi deux raisons d'explorer plus avant ce domaine vaste, qui m'était encore inconnu. La cybernétique était alors une science jeune et séduisante, prometteuse et futuriste, où la manière de traiter l'information faisait appel aussi bien à l'électronique qu'aux modèles biologiques de fonctionnement du système nerveux humain et du cerveau.

Je décidais donc de faire une thèse au LETI (Laboratoire d'Electronique et de Technologie de l'Information), sur le thème du filtrage et de l'analyse spectrale paramétrique [Lut88a]. C'est là que je devais apprendre les rudiments du métier.

En parallèle, les vacations que j'eus l'occasion d'assurer au CUEFA (centre CNAM de Grenoble), à la Formation Continue de l'INPG sous l'égide de Michel Baribaud, ou encore à l'INSTN, furent pour moi la confirmation de ma vocation à l'enseignement.

Si je devais résumer en quelques mots-clés l'essentiel de mon apprentissage au cours de ma thèse, je mentionnerais :

- La Modélisation (et ses dangers),
- L'Estimation aux Moindres Carrés (EQM)
- La Qualité Logicielle

Je reviendrai sur l'importance de ces trois points.

Après ma thèse, caractérisée évidemment, comme tout travail de doctorat, par une spécialisation assez pointue sur un domaine de recherche très restreint [1], j'ai eu la chance d'effectuer 18 mois de service national de coopération auprès du service scientifique de l'ambassade de France à Stockholm. Détaché par le LETI, j'étais chargé de veille technologique dans le vaste domaine de l'électronique, depuis les technologies de l'information [Lut89a], en passant par l'aérospatial [Lut89b], et jusqu'au nucléaire [Lut90b]. Mon rôle consistait aussi à initier des collaborations et organiser des visites (délégation du LETI à Stockholm). Cette activité de journalisme scientifique, intéressante en soi, l'était d'autant plus après un travail de thèse, pour m'apprendre à élargir mon champ de vision du monde.

1.2 Signal, Image ou Parole ?

De retour de coopération, j'étais nommé maître de conférences à l'ENSERG. Affecté au laboratoire LTIRF dans l'équipe Vision, j'initiai une activité en analyse du mouvement dans les séquences d'images.

Par honnêteté, je me dois de mentionner ma réticence initiale à cette affectation. En effet, mes activités de recherche en thèse sur les algorithmes du type prédiction linéaire me destinaient davantage soit vers l'ICP, autre laboratoire de l'ENSERG, car le traitement de la parole fait un large usage de la prédiction linéaire, soit vers le CEPHAG (Centre d'Etude des Phénomènes Aléatoires et Géophysiques), laboratoire plus spécifiquement Signal, avec lequel j'avais déjà collaboré en analyse spectrale paramétrique.

De plus, par rapport à la science-mère que constitue le traitement de signal, le traitement d'images est une science beaucoup plus jeune. La grande quantité de données à manipuler (images ou séquences d'images, c'est-à-dire signaux 2D ou 3D) interdit des traitements trop complexes, notamment si l'on vise des applications réalistes en terme de temps de calcul, d'encombrement matériel ou de coût. Pour le traitement des signaux 1D au contraire, il existe de "beaux outils", comme par exemple la transformée temps-fréquence de Wigner-Ville, ou la fonction d'ambiguïté [2], outils *a priori* proscrits en traitement d'image à cause de leur coût de calcul.

Un collègue, ancien thésard, qui avait travaillé dans une entreprise de traitement d'images de la région, m'avait d'ailleurs confirmé dans l'idée que le traitement d'image était essentiellement de la "cuisine". Je découvrirai peu à peu qu'il avait raison en partie, mais en partie seulement, et que si c'est de la "cuisine", du moins est-ce une cuisine subtile, qui nécessite de jongler entre algorithmes et architectures matérielles, entre rapidité, robustesse et précision ...

Mais le hasard des recrutements et des affectations en avait décidé autrement. Là encore, deux ou trois mots pourraient résumer mon apprentissage au LTIRF :

- L'Information *a priori*,
- La Régularisation statistique,
- Les Mises en œuvre matérielles (temps réel)

1. Il est loin le temps héroïque de la civilisation grecque où un seul esprit pouvait embrasser tout le savoir scientifique et philosophique de l'époque, tel un Thalès, ou un Aristote... Le dernier grand penseur universel est peut-être Poincaré.

2. Je me souviens notamment d'un exposé brillant du professeur Bernard Escudié de l'ICPI-Lyon (Institut de Chimie et Physique Industrielle) sur la fonction d'ambiguïté à l'occasion d'un stage de traitement de signal organisé par l'INSTN, moment fort qui a marqué ma carrière de traiteur de signal. Je tiens à lui rendre ici hommage.

Par rapport à mon acquis en thèse, intervient ici le principe très général de la minimisation d'une fonction de coût comportant toujours deux types termes : un terme d'attache aux données (énergie externe U_a), et un terme de contraintes de modélisation *a priori* (énergie interne U_m).

$$\min U \qquad \text{avec} \qquad U = U_a + U_m. \tag{1.1}$$

Pour simplifier, je pourrai dire que ma thèse n'avait traité que d'un seul des deux termes d'énergie : celui de l'attache aux données, par l'emploi des méthodes de type EQM ou moindres carrés, sans aborder explicitement les aspects pourtant très complémentaires de la régularisation statistique, à l'aide de contraintes ajoutées sur le modèle (gestion d'un *a priori* pur supplémentaire, qui était seulement mentionné comme possible dans la thèse).

De plus, je n'avais fait que du développement logiciel au cours de la thèse, avec malgré tout un souci de qualité et de pérennité des logiciels produits (cf. [Lut88b]), alors que mon activité au LTIRF m'a fait me pencher sur les problèmes importants de mise en œuvre matérielle en vue d'atteindre des traitements temps réel, point essentiel pour l'analyse du mouvement (*e.g.* pouvoir détecter un obstacle présent dans le champ d'une caméra *avant* de l'avoir percuté !).

1.3 Où la Boucle est Bouclée

Parti du signal (thèse dirigée par Jean-Louis Lacoume du CEPHAG), en passant par l'image (affectation au LTIRF dans l'équipe d'Alain Chéhikian), mon parcours de recherche me fait finalement aborder également le domaine de la parole. En effet, je participe au projet Labiophone [3] en collaboration avec l'ICP. Ce projet pluridisciplinaire (signal, parole, image, télécommunication), soutenu par la fédération ELESA et le département SPI du CNRS, est selon moi un bel exemple de la dynamique de la recherche à Grenoble, où les clivages entre laboratoires ou équipes sont oubliés au profit d'un projet scientifique ouvert, faisant appel aux compétences complémentaires des différents partenaires.

Par ailleurs, la fusion des deux laboratoires CEPHAG et LTIRF dans la création du LIS me fournit l'opportunité de revenir à la science-mère : le traitement du signal. Plusieurs perspectives se dessinent, que j'évoque à la fin du document.

3. Je souhaite rendre hommage à la mémoire de C. Benoît, le père de ce projet Labiophone. J'avais appris a l'apprécier au cours de notre collaboration, à la fois sur le plan scientifique (pour la passion qui animait sa recherche) mais aussi sur le plan humain (pour son côté “post-soixante-huitard” un peu provocateur qui me plaisait bien).

Chapitre 2

Thèse en Signal

2.1 Thème de Recherche

Ma thèse, effectuée au LETI sous la direction scientifique de Jean-Louis Lacoume et sous la responsabilité technique de Roland Blanpain, m'a initié à deux thèmes fondamentaux du traitement du signal : le filtrage et l'analyse spectrale. Le titre annonce les trois aspects originaux de mes travaux de thèse, que je détaille dans les paragraphes qui suivent (§2.2, 2.3 et 2.4) :

Filtrage et analyse spectrale paramétriques.
Applications en spectroscopie et magnétométrie RMN.
Eléments de synthèse pour un analyseur expert.

2.2 Atouts et Risques de la Modélisation

L'analyse spectrale paramétrique tire son origine du domaine de l'automatique numérique. Elle utilise l'outil transformée en Z, mais appliqué au traitement du signal. Elle consiste à modéliser un signal numérique x_n, échantillonné aux instants $t = n\Delta t$, où $\Delta t = 1/F_e$ est la période d'échantillonnage, comme étant la sortie d'un filtre numérique de type ARMA (*Auto-Regressive Moving-Average*) excité en entrée par un bruit blanc centré e_n de variance σ^2 :

$$\hat{x}_n = -\sum_{k=1}^{p} a_k x_{n-k} + \sum_{k=0}^{q} b_k e_{n-k} \tag{2.1}$$

On passe alors au domaine fréquentiel du signal en utilisant la transformée en Z, au lieu de la classique transformée de Fourier. La fonction de transfert en Z du filtre s'exprime par :

$$H(z) = \frac{\sum\limits_{k=0}^{q} b_k z^{-k}}{\sum\limits_{k=0}^{p} a_k z^{-k}} \tag{2.2}$$

En prenant z sur le cercle unité, ($z = \exp j2\pi f/F_e$), on obtient $H(f)$ d'où le spectre du signal modélisé:

$$S_{xx}(f) = |H(f)|^2 S_{ee}(f) \tag{2.3}$$

où $S_{ee}(f) = \sigma^2 \Delta t$ représente le spectre du bruit blanc en entrée. Les paramètres a_k, b_k, σ du modèle sont estimés par une méthode du type moindres carrés. Il s'agit de minimiser une fonction erreur entre le modèle et le signal observé. Différents modèles peuvent être choisis, selon la nature du signal à analyser (AR, MA, ARMA, Prony, Pisarenko). Différents critères de minimisation et algorithmes d'estimation sont également disponibles (EQM, LMS, RLS, filtres en treillis).

Dans la mesure où l'on dispose d'un modèle permettant le calcul d'un échantillon temporel du signal à n'importe quel instant, on s'affranchit de l'inconvénient classique de la FFT (troncature temporelle qui se traduit par une résolution fréquentielle limitée et des ondulations sur le spectre). La Fig. 2.1 illustre l'amélioration potentielle de l'analyse spectrale paramétrique dans le cas d'un signal synthétique de courte durée formé de deux exponentielles complexes dans du bruit.

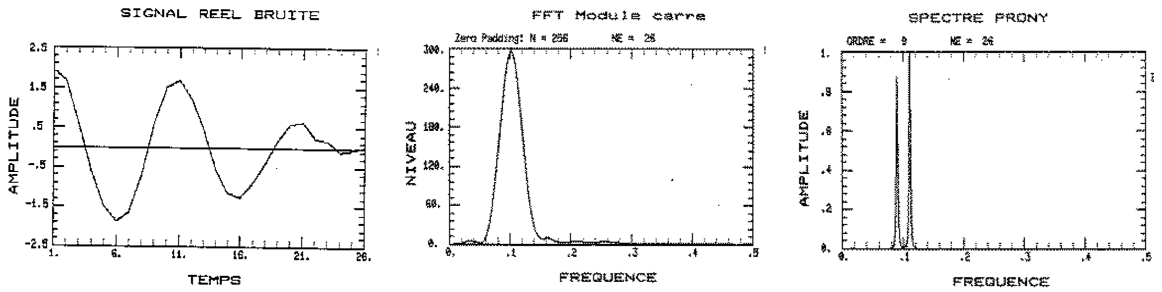


FIG. 2.1 - *Simulation d'analyse spectrale: comparaison FFT et spectre de Prony. Les deux raies spectrales sont noyées dans un seul lobe avec la FFT, alors qu'elles sont bien résolues par l'analyse de Prony.*

Ce principe de modélisation s'applique aussi bien pour la réduction de bruit : il s'agit alors de mettre en œuvre des filtres adaptatifs (du type treillis ou LMS) pour estimer un modèle du bruit additif qui entâche le signal. On obtient là encore des résultats remarquables, sur des signaux de simulation constitués de sinusoïdes noyées dans un bruit MA, AR ou HF (cf Fig. 2.2).

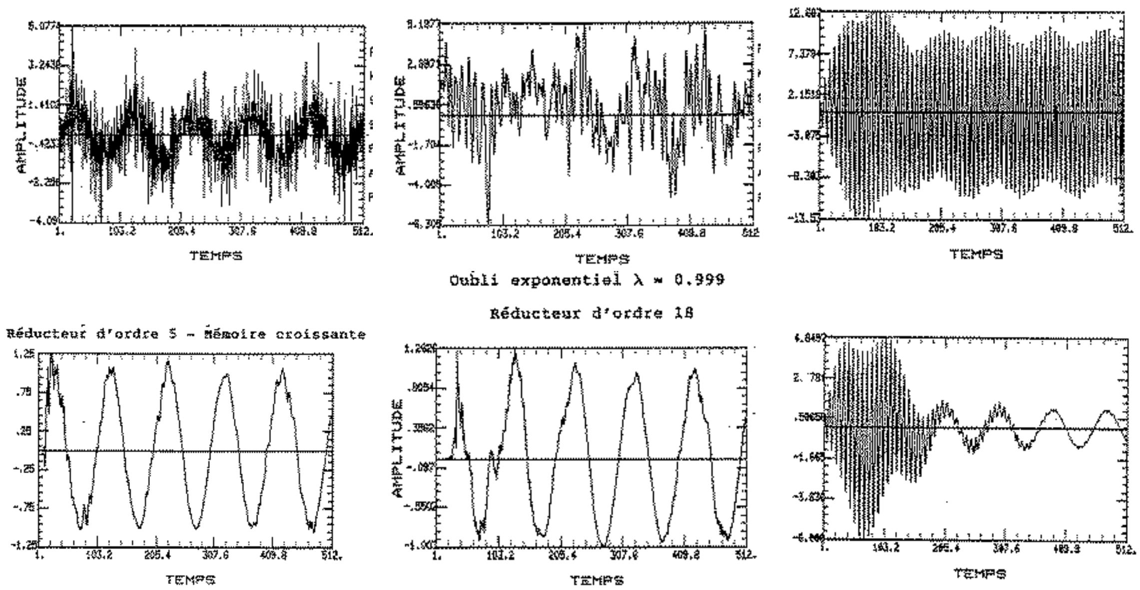


FIG. 2.2 - *Simulation de filtrage : signaux bruités en haut ; signaux filtrés en bas. On remarque l'adaptativité temporelle du filtrage.*

Cependant, ces techniques de modélisation, qui semblent très performantes au premier abord, doivent être employées avec précaution. J'ai en effet appris au cours de ma thèse que l'on peut faire dire à un modèle ce que l'on veut : si l'on sait ce que l'on cherche dans un signal, on peut toujours trouver le bon modèle qui fera ressortir le résultat à trouver. D'où à la fois l'importance de l'information *a priori* pour définir un bon modèle, et le danger des simulations sur des signaux synthétiques correspondant au cas idéal du modèle, avec lesquels on peut exhiber d'excellents résultats, mais qui ne prouvent rien quant à la robustesse et la validation de la modélisation, si ce n'est que l'algorithme a été programmé sans “bugs”...

2.3 Validation par les Applications

On comprend alors toute l'importance de la validation d'une méthode de traitement sur des données réelles (signaux ou images). Au cours de ma thèse, différentes applications en RMN (spectroscopie ou magnétométrie par Résonance Magnétique Nucléaire) ont permis de valider les outils développés.

La Fig. 2.3 montre l'amélioration effective obtenue par analyse spectrale paramétrique dans une application de spectroscopie de signaux RMN in vivo. Il s'agit d'extraire les différentes raies spectrales correspondant à l'ATP (adénosine tri-phosphate) contenue dans les cellules d'un cerveau de rat (pauvre bête, Dieu ait son âme . . .). La détection des multiplets de l'ATP (deux doublets, un triplet) est bien meilleure avec l'analyse paramétrique basée sur le modèle de Prony. Cette application a fait l'objet d'une publication dans une revue [LBDA89]

Une autre application en magnétométrie RMN pour l'étude géologique profonde du sol a permis de valider l'intérêt de filtres paramétriques réducteurs de bruit. La Fig. 2.4 illustre les signaux obtenus : le champ magnétique est enregistré simultanément au sol (référence) et à différentes profondeurs dans le puit de forage (ici 1800m). Il existe *a priori* une forte cohérence entre ces signaux.

Pour extraire l'information géologique, il est nécessaire dans un premier temps de décorréler complétement les deux signaux. Une mesure de cohérence permet de quantifier la qualité du résultat. La technique classique consiste en une simple différence des signaux, ce qui ne diminue que partiellement la cohérence. L'emploi d'un filtre réducteur de bruit adaptatif améliore sensiblement le résultat (cohérence proche de zéro). Il permet d'identifier la fonction de transfert du filtre passe-bas que constitue le sol terrestre et d'en déterminer la fréquence de coupure (Fig. 2.5).

2.4 Evaluation, Expertise et Qualité Logicielle

Dans tous les domaines de la recherche, de nombreux auteurs proposent souvent différents algorithmes pour résoudre un problème donné. De multiples variantes et options fleurissent dans la littérature, et le jeune chercheur ressent alors le besoin d'articles de synthèse qui l'orientent à travers la jungle des idées originales. Le transfert du savoir est essentiel au progrés de la recherche et à la formation des jeunes doctorants. Un bon article de synthèse vaut souvent mieux que l'exposé de la (n+1)ième variante d'un algorithme ...

C'est pourquoi le dernier aspect de mon travail de thèse concerne un essai de synthèse des différents outils et techniques disponibles en analyse spectrale et filtrage paramétriques, en vue de la définition d'un analyseur de spectre expert. Différents modèles paramétriques sont évalués en détail. On souligne l'importance du choix du critère de sélection de l'ordre du modèle, du critère de minimisation de l'EQM (erreur progressive, rétrograde ou conjointe), l'intérêt de la SVD (décomposition en valeurs singulières qui sont les valeurs propres de la matrice de covariance

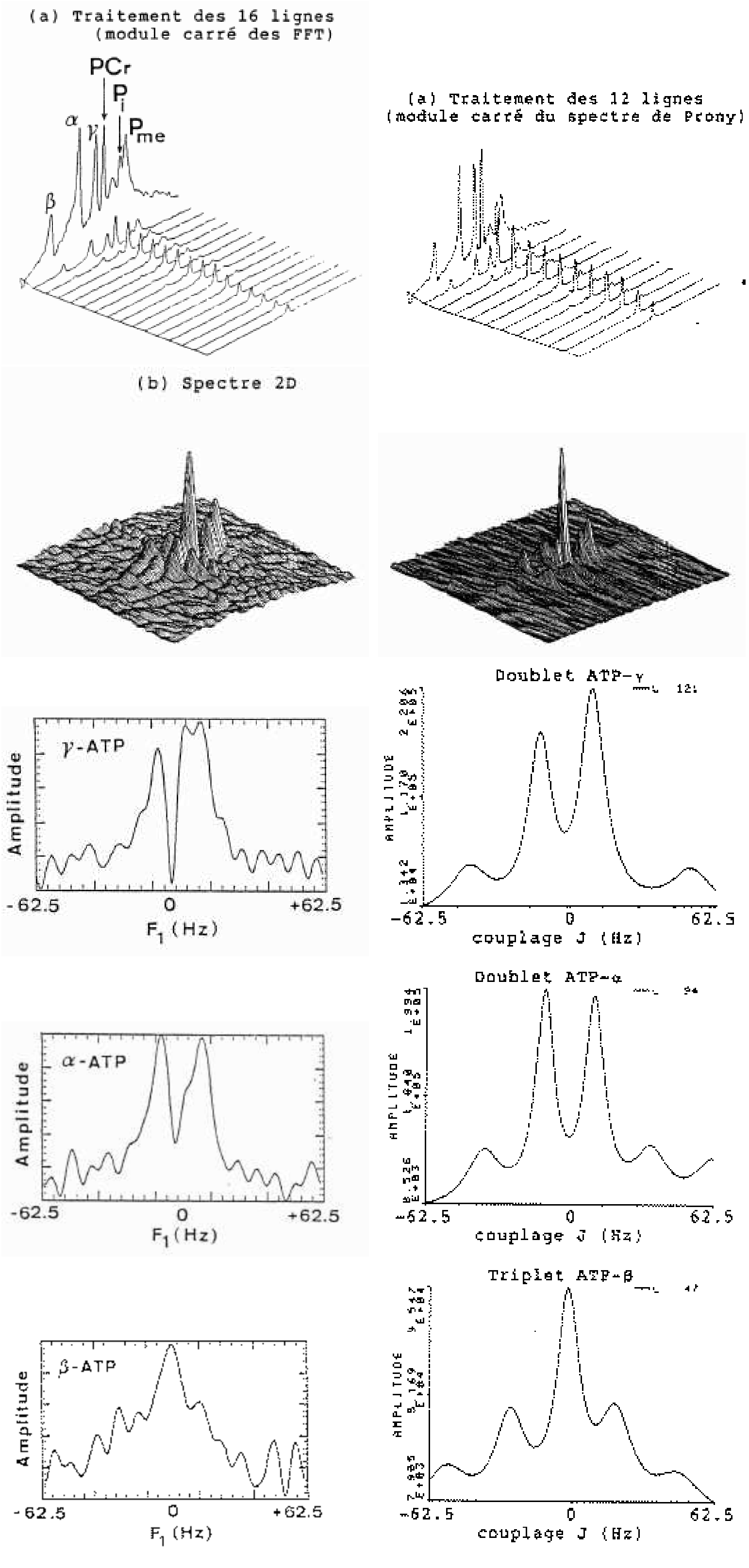


FIG. 2.3 - *Résultats sur signaux in vivo : FFT à gauche ; méthode paramétrique à droite.*

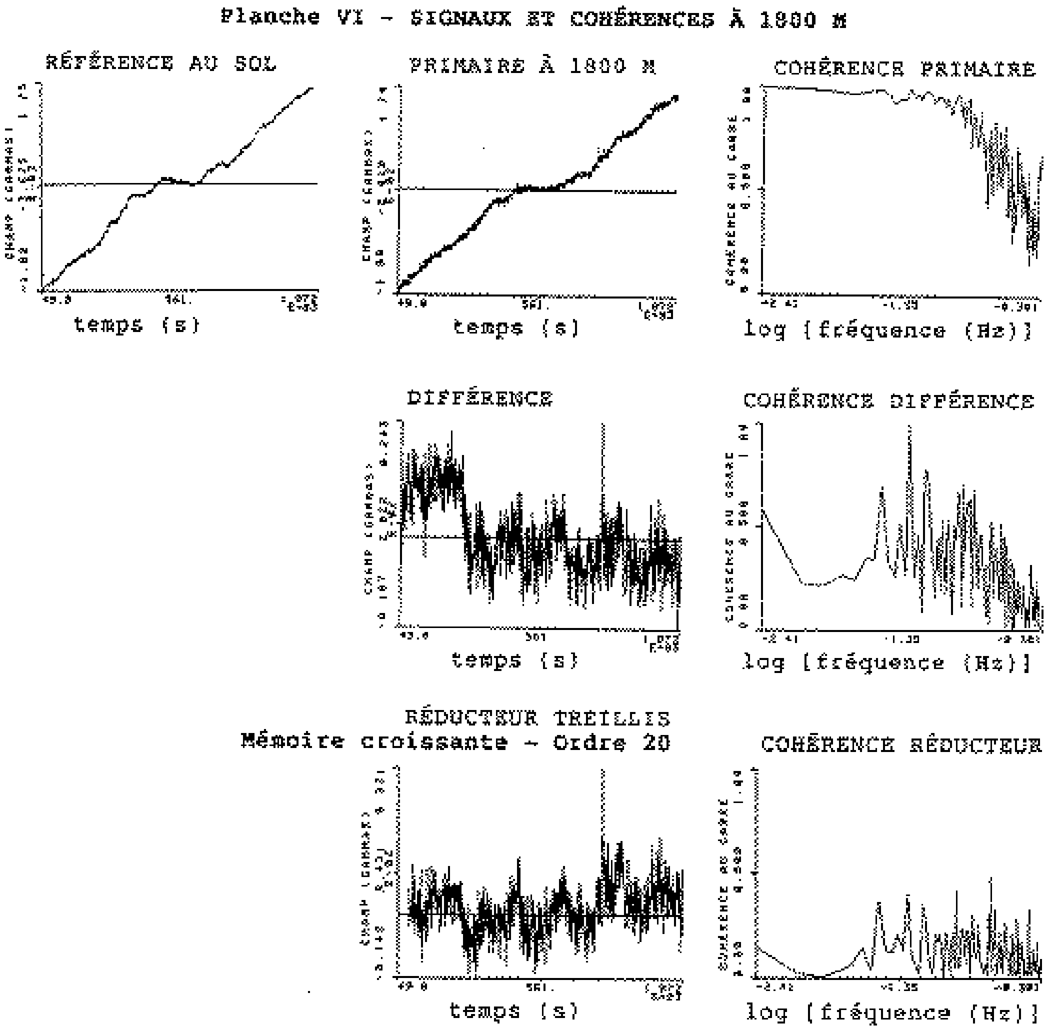


FIG. 2.4 - *Efficacité d'un Réducteur de bruit : comparaison des cohérences.*

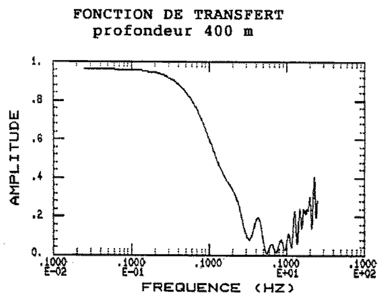


FIG. 2.5 - *Fonction de transfert identifiée du sous-sol à 400m de profondeur. Fréquence de coupure : 0.1 Hz.*

des données). Ce travail se poursuit actuellement au LIS (ex-CEPHAG), dans le projet ASPECT d'aide à l'interprétation spectrale.

D'autre part, l'évaluation (notamment statistique) d'une méthode ou d'un algorithme est selon moi un point très important en recherche : l'expertise est le complément indispensable de la modélisation.

Enfin, la pérennité d'un travail requiert un effort non négligeable de lisibilité et de qualité logicielle, effort que j'ai fourni a l'issue de ma thèse [Lut88b].

2.5 Bilan

Avec le recul, je tire quelques enseignements principaux de ce travail de thèse, hormis le contenu purement scientifique et technique.

D'abord, l'aiguillon de la recherche réside dans les applications : même si le chercheur ne doit pas se contenter du travail d'ingénieur, le moteur de son activité est bien le champ des applications. Selon le mot d'un philosophe, la science (la physique) a la modeste ambition d'expliquer le comment, si elle ne peut expliquer le pourquoi (ce qui reste du ressort de la métaphysique). C'est-à-dire qu'elle a un rôle à jouer dans l'action, dans la pratique.

Deuxièmement, un modèle est bon s'il est efficace. La vérité n'a rien à voir avec l'efficacité, l'action. Le chercheur ne saurait prétendre à la recherche du vrai, mais plus modestement à la recherche de l'effectif. Les modèles physiques ou mathématiques ne sont pas sensés expliquer le monde, mais seulement décrire son fonctionnement. Je reviendrai d'ailleurs sur la notion de modèle, en distinguant modèles mathématiques et modèles biologiques.

Enfin, la validation d'un travail de recherche requiert trois choses, certes ingrates, mais nécessaires :

- une évaluation de type statistique sur des données réelles, pour confirmer la robustesse d'une méthode et toutes ses qualités annoncées : s'arrêter aux simulations est facile mais trompeur, voire malhonnête.

- le positionnement du travail dans le contexte des autres travaux existants, tâche à la fois plus difficile que par le passé (à cause de l'extension rapide du volume du savoir, des connaissances et des méthodes, et de la prolifération cancéreuse des publications), mais aussi plus facile grâce aux nouveaux moyens de communication (serveurs web auxquels on peut accéder pour lancer des algorithmes sans avoir à les reprogrammer, ou banques de données communes sur lesquelles on peut évaluer objectivement sa propre méthode).

- une mise en œuvre robuste et réutilisable (logiciel ou matériel) : production de notes techniques, d'interfaces conviviales et portables, développement d'architectures de traitement rapide, ce qui représente un travail parfois ingrat, mais cependant utile, à la fois pour la communauté scientifique et technique, et pour le chercheur lui-même.

En conclusion, je rappelerai une remarque faite par une enseignante de philosophie s'adressant à ses étudiants préparant leur mémoire de maîtrise, et qui s'inquiétaient de l'originalité de leur rapport. Elle leur disait qu'il fallait avant tout lire, faire de la bibliographie, engranger le savoir et les connaissances découvertes par autrui, bref travailler et apprendre, et elle concluait par ses mots : "L'originalité vient de surcroît ..."

Chapitre 3

Du LTIRF au LIS

3.1 Le Laboratoire d'Accueil

Décrire le laboratoire d'accueil est une tâche à la fois nécessaire et agréable. Nécessaire, car cela permettra de m'y situer, et de mettre en évidence les collaborations nées de mon activité. Agréable, car malgré de réguliers "coups de gueule" avec les collègues ou la direction, l'ambiance du laboratoire reste très conviviale, et c'est un plaisir que de décrire un lieu de travail où l'on se sent libre d'exprimer sa différence, face à des collègues assez intelligents pour l'accepter, sinon y agréer.

Jusqu'en décembre 1997, le LTIRF est un laboratoire d'une quarantaine de personnes, dont une quinzaine de permanents (enseignants-chercheurs, techniciens et secrétaire) et une trentaine d'étudiants (thésards, DEA et stagiaires). Il est constitué en trois équipes :

- l'équipe Vision qui développe des algorithmes et des architectures matérielles de traitement rapide pour la vision par ordinateur et le traitement d'image bas niveau (détection de contours, segmentation d'images, pyramides d'images, poursuite d'indices, machine de vision),

- l'équipe Réseaux Neuronaux qui étudie et applique les réseaux de neurones au traitement de données et à la reconnaissance de formes, avec une approche essentiellement biologique de la vision (séparation de sources, analyse de texture et de couleur, classification, réseaux évolutifs).

- l'équipe Circuits Spécifiques, qui conçoit des circuits intégrés analogiques et numériques pour le traitement du signal et la vision. Malgré sa création plus récente, cette équipe correspond à un savoir faire de longue date en microélectronique, et reflète une des spécificités "hardware" du laboratoire (convolueurs rapides, rétine en silicium).

De par sa taille moyenne, de par son statut relativement indépendant vis-à-vis du CNRS (simple équipe d'accueil jusqu'en 1998), et de part l'homogénéité de statut des personnels ayant une activité de recherche (tous professeurs d'université ou maîtres de conférences), le LTIRF avait la particularité d'être un laboratoire ouvert, accueillant et libéral, voire libertaire, où chacun, du stagiaire au professeur, s'exprime librement.

Depuis le 1er janvier 1998, la fusion avec le CEPHAG a abouti à la création du Laboratoire des Images et des Signaux (LIS-INPG). Cette fusion présente un intérêt évident : d'une part, elle concrétise les collaborations déjà existantes entre les deux laboratoires (notamment en séparation de sources, analyse temps-fréquence, moments d'ordre supérieur ...). D'autre part, elle confère au LTIRF un statut CNRS (UPRESA 5083), reconnaissance de sa valeur scientifique dans le domaine des Sciences pour l'Ingénieur (SPI). La contrepartie étant que le laboratoire, en s'agrandissant, risque de perdre en autonomie (lien plus fort au CNRS) et en souplesse de fonctionnement. On peut compter sur l'indépendance d'esprit des personnels de l'ex-LTIRF pour veiller au grain ...

3.2 Animation Scientifique

3.2.1 Le Groupe Mouvement-Markov (M&Ms)

Affecté à l'équipe Vision du LTIRF en octobre 1990, j'ai initié une activité de recherche nouvelle au sein du laboratoire : l'analyse du mouvement dans les séquences d'images. J'ai constitué un petit groupe de recherche autour de ce thème initial, en l'élargissant peu à peu à la segmentation spatio-temporelle de scènes dynamiques, orientée vers des applications en télésurveillance, en télécommunication audiovisuelle, et en compression de séquences d'images. En moyenne, 5 à 6 personnes travaillent sous ma responsabilité chaque année : 1 ou 2 enseignants-chercheurs, 1 ou 2 doctorants et 2 ou 3 stagiaires (DEA et ingénieurs).

J'ai développé plusieurs axes détaillés ci-après. La Fig. 3.1 en fait la synthèse (thèmes prépondérants en gras), en répondant à la question exhaustive portant sur une chaîne de traitement : “quel type d'outils mathématiques et modèles, appliqués à quel type d'information, pour résoudre quel type de tâche, requise dans quel type d'application, nécessitant quel type de mise en œuvre ?”

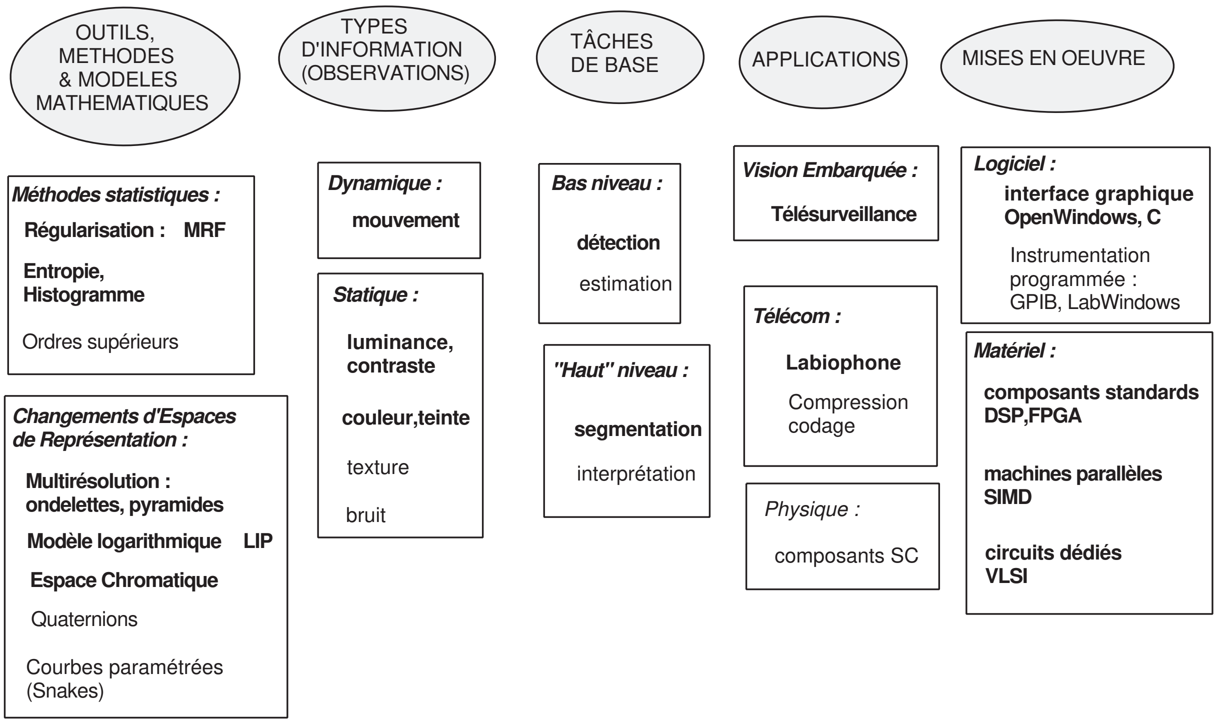


FIG. 3.1 - *Synoptique de mon domaine d'activité.*

3.2.2 Outils et Modèles Mathématiques

Régularisation Statistique par MRF

Pour l'analyse du mouvement et la segmentation spatio-temporelle de séquences vidéo, je me suis orienté vers une approche statistique de modélisation par champ aléatoire de Markov (*Markov Random Field* ou MRF) inspirée des travaux de l'équipe de P. Bouthémy [BL93, LB90, Lal90]. Cette approche est particulièrement intéressante en raison de la régularisation qu'elle assure (face à un problème mal posé comme celui de la détection de mouvement) et de sa robustesse vis-à-vis du type de séquence traitée (stabilité du paramétrage du modèle). Elle offre un cadre statistique bayésien rigoureux permettant de fusionner différentes informations *a priori* et différentes observations. Cet outil statistique permet de calculer un champ de primitives cachées (étiquettes) étant donné un champ de données (observations).

Les deux propriétés essentielles d'un MRF sont :

- une notion de voisinage délimitant l'extension des interactions entre sites-pixels[1],
- une probabilité *a priori* du champ d'étiquettes modélisée par une distibution de Gibbs : $P \propto exp(-U)$ où P représente la probabilité et U la fonction d'énergie associée.

Pour la détection de mouvement par exemple, le champ d'étiquettes est binaire (fixe/mobile), et les observations sont les différences temporelles d'intensité lumineuse. La configuration la plus probable du champ d'étiquettes en accord avec les observations est obtenue par la minimisation d'une fonction d'énergie, ne nécessitant que des calculs locaux et fortement parallélisables, d'où des possibilités intéressantes de mises en œuvre rapides. Des algorithmes de relaxation (stochastiques ou déterministes) permettent de trouver ce minimum.

Cette approche résulte d'une analogie avec la thermodynamique, où l'état d'équilibre d'un gaz correspond à son état d'énergie minimale : ici, le gaz correspond à la séquence d'images, les molécules du gaz sont les pixels, et l'interaction entre molécules voisines correspond à l'interaction locale entre pixels. L'état stable du gaz correspond à la configuration optimale des étiquettes attribuées aux pixels de la séquence vidéo.

Nous avons proposé un modèle spatio-temporel robuste pour la détection de mouvement dans le cas d'une caméra fixe, utilisant un voisinage 3-D des pixels de la séquence avec des cliques binaires (Fig. 3.2) [LCL99].

Les potentiels énergétiques d'interaction $\beta(s,n)$ entre deux sites voisins s et n sont fonction inverse de la distance euclidienne qui les sépare dans le cube, distance éventuellement pondérée par des facteurs d'échelle différents selon l'orientation spatio-temporelle de la liaison (horizontale, verticale, diagonale, passée, future) :

$$\beta(s,n) = \frac{1}{\sqrt{\left(\frac{\delta_x}{\beta_x}\right)^2 + \left(\frac{\delta_y}{\beta_y}\right)^2 + \left(\frac{\delta_t}{\beta_t}\right)^2}}. \tag{3.1}$$

où δ_x, δ_y et δ_t sont les coordonnées dans l'espace 3-D (x,y,t) du vecteur $\overrightarrow{(s,n)}$ correspondant au couple formé par le pixel courant s et l'un de ses voisins n, et où β_x, β_y et β_t sont les facteurs d'échelle. Ces facteurs d'échelle peuvent eux-mêmes être fonction des informations *a priori* dont on dispose [LL98b], ou des contraintes que l'on souhaite imposer (contraintes géométriques sur les formes par exemple).

1. "Un mouvement est une pensée de relations et de comparaisons" ([Ala41], p. 39).

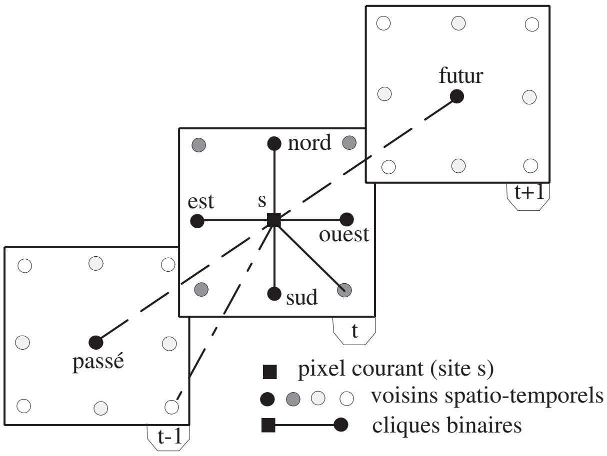


FIG. 3.2 - *Voisinage spatio-temporel d'un pixel.*

Pour le calcul de la configuration optimale du champ d'étiquettes (correspondant au minimum d'énergie), on a développé une technique de relaxation à la fois spatiale et temporelle qui s'applique sur des sections temporelles de longueur N_t de la séquence d'images. On a alors affaire à un MRV (*Markov Random Volume*) au lieu d'un MRF (*Markov Random Field*) [CL95a, CL95c].

Le modèle a par ailleurs été étendu en vue d'un "multi-étiquetage" qui permet de différencier plusieurs objets mobiles. On obtient autant d'étiquettes différentes que de zones mobiles distinctes dans l'image [LC93]. Le multi-étiquetage est géré par le processus markovien lui-même (nombre d'étiquettes automatiquement remis à jour en fonction du contenu de la scène).

Ces travaux ont fait l'objet d'un DEA [PC92] et de la première thèse que j'ai dirigée [Cap95].

L'approche MRF spatio-temporelle est aussi utilisée pour la segmentation des lèvres d'un locuteur. On utilise alors 2 observations (teinte et changement temporel) et 4 étiquettes pour coder 4 configurations pertinentes : présence/absence de rouge et présence/absence de mouvement.

Entropie et Raffinement d'Histogramme

La détermination des paramètres toujours nécessaires dans les algorithmes de traitement d'image est un problème récurrent : nous nous attachons à proposer des algorithmes robustes, c'est-à-dire peu sensibles au choix des paramètres (notamment des seuils).

Pour ce faire, il faut d'une part limiter au maximum le nombre de paramètres requis dans une méthode, pour des raisons évidentes de robustesse et de stabilité du modèle. Un bon contre-exemple se trouve dans [KN95] où le modèle proposé ne possède pas moins de 20 paramètres, ce qui laisse perplexe quant à sa stabilité.

D'autre part, il faut éviter au maximum l'utilisation d'heuristiques (méthodes *ad hoc* souvent criticables) pour déterminer ces paramètres, mais plutôt proposer des méthodes d'estimation automatique et en ligne, en fonction des données disponibles.

Des méthodes simples basées soit sur des histogrammes raffinés, soit sur la notion d'entropie donnent des résultats remarquables pour l'évaluation des seuils nécessaires en segmentation de teinte et de mouvement [LL99]. Evaluer l'entropie d'une différence d'images permet en effet d'évaluer le bruit qui affecte le mouvement, et de positionner ensuite le seuil juste au dessus de ce niveau de bruit, pour ne conserver que les changements temporels significatifs.

Pour la segmentation des lèvres dans un visage de locuteur, une méthode par raffinement d'histogramme est proposée dans [LDC+99] pour estimer automatiquement la teinte correspondant aux lèvres : elle consiste à estimer d'abord le mode dominant correspondant à la teinte de la peau, puis a éliminer les pixels appartenant à ce mode : on calcule alors un deuxième histogramme raffiné où apparaît clairement le mode mineur des lèvres, qui était au départ noyé dans le mode majeur de la peau (Fig. 3.3).

FIG. 3.3 - *Raffinement d'histogramme : à gauche, l'histogramme complet qui fait apparaître deux modes dominants : la peau (pic noir) et le fond; au milieu, les pixels du mode dominant (segmentés en noir); à droite, l'histogramme raffiné après élimination des pixels du mode "peau", qui laisse apparaître le mode mineur des lèvres (pic noir).*

Statistiques d'Ordre Supérieur

La caractérisation des composants électroniques utilisés en technologie VLSI (*Very Large Scale Integration*) implique l'étude de bruits non-gaussiens. En effet, dans les transistors MOS de petite aire (inférieure au μm^2), on observe des bruits de type RTS (*Random Telegraph Signal*) dus au piégeage-dépiégeage d'un électron individuel dans un piège situé dans l'oxyde de grille. Ces fluctuations induisent des fluctuations de conductance du canal, induisant à leur tour des fluctuations de variance de bruit thermique du canal, indétectables sur la densité spectrale de puissance, et nécessitant le recours aux moments d'ordre supérieur.

Un bon indicateur de ce type de fluctuation est la courbe donnant la variance de l'histogramme de la variance glissante en fonction du nombre de points utilisés pour calculer la variance. La recherche d'autres indicateurs, une étude systématique de l'influence de divers facteurs ($50Hz$, dérives d'alimentation) sur les indicateurs en question, et l'expérimentation sur des signaux de bruit provenant d'autres dispositifs (diodes Schottky) nécessitent la mise en œuvre d'un système d'acquisition et de mesures commode, en ligne et si possible temps-réel.

J'ai co-encadré, avec Jean Brini du LPCS, un DEA de Microélectronique sur ce projet [Bon98] : la chaîne d'acquisition, de mesures et de traitement statistique (basé sur le calcul de l'indicateur ci-dessus) a été mise en œuvre sur l'analyseur de signaux HP35665A. Le cadre de travail est désormais posé et va nous permettre d'explorer d'autres indicateurs statistiques d'ordre supérieur. A terme, le projet vise au développement de nouvelles méthodes de caractérisation des composants électroniques, particulièrement adaptées aux phénomènes non-linéaires apparaissant dans les technologies très miniaturisées.

Techniques Multi-résolution

Un de nos axes de travail théorique porte sur l'étude de différentes techniques multirésolution associées soit à la modèlisation MRF pour la détection de mouvement (représentation pyramidale passe-bas des champs d'étiquettes et/ou d'observations, qui a fait l'objet d'un DEA [Hen93]), soit aux algorithmes de calcul du flux optique (champ dense de vecteurs-vitesse) [Bou96]. L'apport des techniques multirésolution permet d'améliorer le comportement des algorithmes dans des cas délicats de mouvement : possibilité de détecter et/ou estimer une plus large gamme de vitesses d'objets (très lents, très rapides), des objets mobiles ayant une luminance uniforme (très peu texturés) ou des mobiles bruités. Nous avons développé un algorithme original de multirésolution à la fois spatiale et temporelle [CL95b, CL95d, CL97], qui donne des résultats intéressants pour la détection de mouvement dans ces cas délicats. Un exemple de pyramide spatio-temporelle est donné Fig. 3.4. Le nombre de niveaux dans la pyramide est fonction de l'amplitude des mouvements présents dans la scène et de la taille des objets mobiles.

FIG. 3.4 - *Exemple de pyramide spatio-temporelle passe-bas à trois niveaux de résolution sur la séquence Trevor (de haut en bas, $k = 0, 1, 2$).*

Pour pallier l'inconvénient du lissage inhérent aux pyramides passe-bas (perte de résolution spatiale et temporelle), on s'intéresse aux ondelettes 2-D, voire 3-D (bancs de filtres miroirs en quadrature ou QMF) [Sim95, Chr96, LS96]. Cela permet de cumuler les avantages des filtrages passe-bas et passe-haut, qui fournissent des informations complémentaires et en principe non redondantes.

La figure 3.5 donne un exemple typique de décomposition 2-D avec des ondelettes de Daubechies à 4 coefficients sur une séquence synthétique d'images comportant un carré sombre se déplaçant horizontalement vers la droite de 1 pixel/image. On constate bien sûr que les coefficients d'ondelettes dans les sous-bandes HB, BH et HH sont non nuls au voisinage des contours de l'objet. Mais il est surtout intéressant de noter l'évolution du signe des coefficients sur cet exemple simple : alors que les coefficients des contours horizontaux gardent le même signe (contours toujours noirs ou toujours blancs), les coefficients des contours verticaux, qui sont perpendiculaires au sens du mouvement, alternent de signe à chaque image : noir, blanc, noir, blanc ...

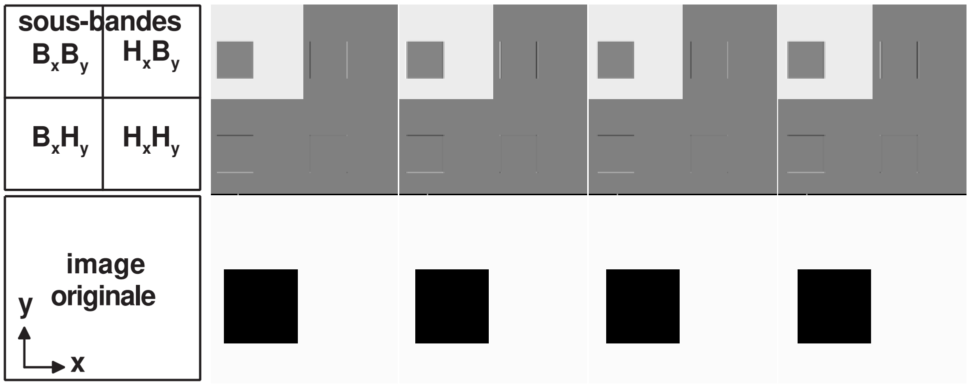


FIG. 3.5 - *En haut) Transformée en ondelettes des images : la tendance (BB) et les 3 fluctuations HB, BH et HH (le fond gris foncé correspond à des coefficients nuls, le noir à des coefficients négatifs, le blanc à des coefficients positifs). En bas) Séquence synthétique : carré noir mobile sur fond clair.*

Ce phénomène nous a conduit à remettre en cause l'hypothèse de cohérence du mouvement dans les sous-bandes, faite par certains auteurs [Baa91]. On a mis en évidence que cette hypothèse, valable dans le cas de filtres continus, n'est plus valable dans le cas de l'utilisation de filtres d'ondelettes discrets. L'invariance par translation n'existe que si le déplacement de l'objet est un multiple d'une puissance du facteur d'échelle de résolution (typ. 2^m pour des ondelettes diadiques). En effet, il faut avoir : $2^{-m}\Delta x = \Delta n \in Z \Leftrightarrow \Delta x = \Delta n.2^m$. Sinon, un déplacement dans le signal ne se traduit pas par un simple déplacement dans les coefficients.

Une contrepartie intéressante de cette constatation est que la recherche de périodicités des coefficients dans les sous-bandes est un moyen d'estimer l'amplitude du déplacement. Une simulation en 1-D fera comprendre notre propos : la Fig. 3.6 présente un créneau lissé qui se déplace d'un échantillon à chaque instant. Au premier niveau de fluctuation, on trouve une périodicité de 2, au deuxième niveau, une périodicité de 4, indiquant que le déplacement est effectivement unitaire à pleine résolution.

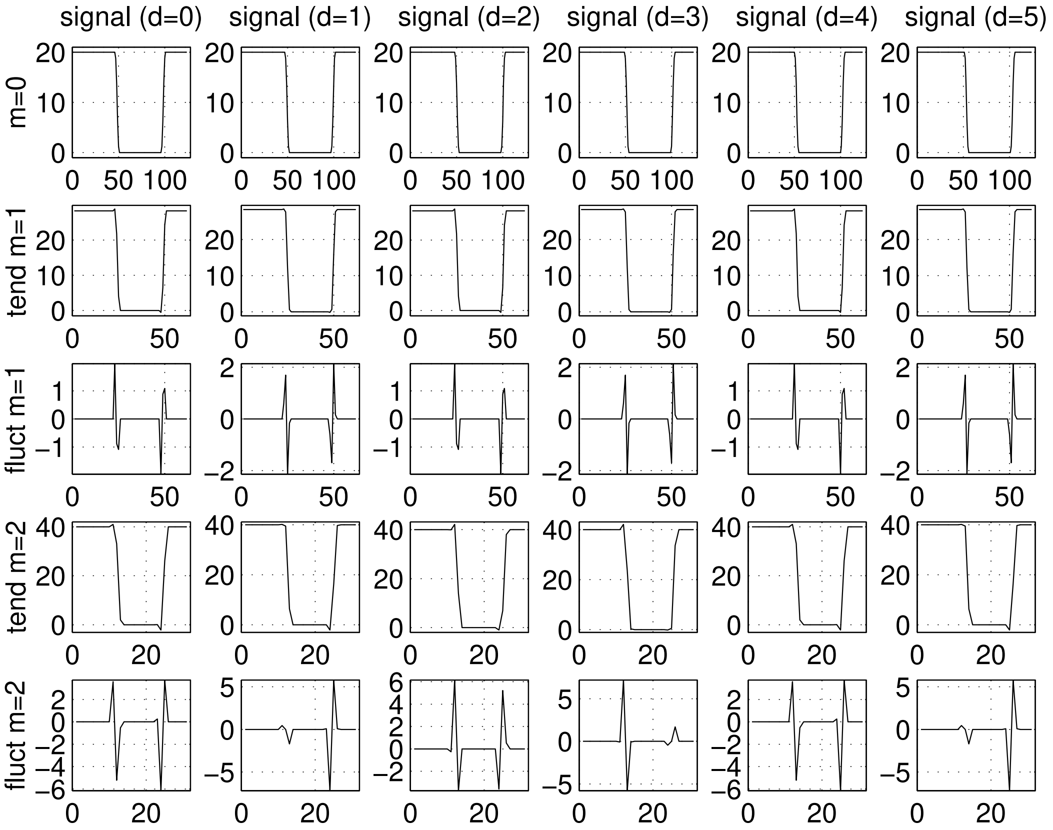


FIG. 3.6 - *Simulation de mouvement en 1-D : En haut) Signal de type créneau à 6 instants successifs (déplacement d de 0 à 5 échantillons) ; En bas) Transformée en ondelettes : tendances et fluctuations à deux niveaux de résolution ($m = 1$ et $m = 2$).*

Nous nous orientons à présent vers l'utilisation d'ondelettes 3-D appliquées à des sections temporelles de séquences d'images. Cette approche rejoint la méthode fréquentielle proposée par Heeger, mais avec des avantages spécifiques à l'emploi des ondelettes.

Modèle LIP

Le modèle LIP (*Logarithmic Image Processing*) a été développé par l'équipe de Michel Jourlin de l'INSA Lyon [JP95, MPJ96], initialement pour le traitement d'images obtenues en transmission (radiographies), mais s'étend aux images obtenues par réflexion. Sous sa forme simplifiée, le modèle LIP consiste à associer à la fonction intensité $I(x,y)$ d'une image en niveaux de gris une fonction de tons définie par : $f(x,y) = M - I(x,y)$, où $M = 256$ typiquement. Il existe un isomorphisme fondamental ϕ entre l'espace des tons de gris E et l'espace des réels R, défini par :

$$\phi(f) = -M \log\left(1 - \frac{f}{M}\right). \tag{3.2}$$

On dispose ainsi d'une structure d'espace vectoriel réel où l'on redéfinit de façon cohérente toutes les opérations algébriques de base entre tons de gris (addition, soustraction, multiplication par un scalaire, norme).

Les principaux points forts de ce modèle sont un cadre mathématique algébrique très rigoureux, une transformation logarithmique de l'information d'intensité lumineuse (respectant le phénomène biologique de compression logarithmique réalisée par l'œil), une redéfinition du contraste conforme aux lois psychovisuelles de Weber et Fechner, et des opérations qui assurent de travailler sur des grandeurs toujours comprises dans l'intervalle borné [0,255], permettant notamment des seuillages robustes [Pin97].

C'est au cours d'une école d'été que j'ai organisée [CCL97] que j'ai pris connaissance de ce modèle. Il m'a paru répondre à un problème que je me posais : alors que l'œil discerne (*i.e.* segmente) sans difficulté les lèvres dans un visage, le traitement classique d'images (*e.g.* détection de contours) a beaucoup de difficulté pour atteindre les mêmes performances. Le contraste est en effet très faible dans certaines zones de contours des lèvres. Une transformation de type logarithmique de l'information d'une part, et une redéfintion du contraste d'autre part, me semblaient donc dignes d'intérêt pour tenter de résoudre le problème.

Nous avons appliqué avec succès ce modèle à la segmentation de la teinte des lèvres dans un visage de locuteur.

Transformées de l'Espace Couleur

Pour la segmentation spatio-temporelle des lèvres d'un locuteur, l'information de teinte joue un rôle important, le rouge étant la couleur prédominante sur les lèvres [2]. On a donc été amené à traiter des séquences vidéo couleur de locuteurs, et à s'intéresser à différentes transformées de l'espace des couleurs (Fig. 3.7), du type $RVB \rightarrow TLSP$ (teinte, luminance, saturation, pureté) [PB95].

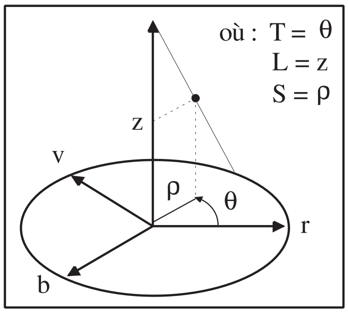


FIG. 3.7 - *Espace couleur.*

Les transformations angulaires classiques, utilisant la fonction arctangente appliquée à des

2. "Et le mouvement, apparent ou réel, ne serait point perçu sans quelque changement dans les couleurs et les lumières." ([Ala41], p. 45).

rapports de différences de composantes couleur (Eq. (3.3)), sont très sensibles au bruit.

$$T = \frac{\pi}{2} - \arctan\left(\frac{2R - V - B}{\sqrt{3}(V - B)}\right) + k \quad \text{où: } k = 0 \text{ si } V > B \text{ et } k = \pi \text{ sinon.} \tag{3.3}$$

Afin de s'affranchir des problèmes d'ombre et de variations d'éclairement, et de gagner en robustesse par rapport au type de caméra utilisée (caméra professionnelle tri-CCD ou micro-caméra mono-CCD), on a développé une transformée couleur très simple de type logarithmique, inspirée du modèle LIP et qui donne des résultats de segmentation robustes. Dans la mesure où le bleu est très peu présent dans le visage, très bruité et fortement corrélé au vert, on n'utilise que les informations portées par le rouge et le vert. On définit donc la teinte par :

$$T = 256 \times \frac{V}{R} \tag{3.4}$$

La Fig. 3.8 illustre l'information apportée par la teinte, qui fait bien ressortir quatre zones principales : les lèvres, l'intérieur de la bouche, la peau et le fond.

FIG. 3.8 - *Séquence typique de locuteur : luminance en haut, teinte en bas (représentée en niveaux de gris).*

On souhaite désormais formaliser notre approche et étendre le modèle LIP au traitement algébrique de la couleur [DP98].

3.2.3 Les Quatre Tâches de la Vision

La vision par ordinateur repose sur quatre tâches principales : la détection, la segmentation, l'estimation et l'interprétation.

Détection

La détection correspond à un processus de décision bas niveau, c'est-à-dire très proche de l'information élémentaire qu'est le pixel (*e.g.* présence/absence d'une caractéristique).

La détection du mouvement vidéo est logiquement la première étape dans l'analyse du mouvement. Elle est basée sur les deux hypothèses de caméra fixe et d'éclairement quasi-constant de la scène observée. Pour assurer la robustesse de la décision, une régularisation des changements temporels détectés est toujours nécessaire. Pour cela, nous avons opté pour une modélisation par champ aléatoire de Markov.

Le modèle a été étendu à la détection avec multi-étiquetage pour différencier plusieurs objets mobiles. L'adjonction d'un filtre de Kalman au modèle markovien permet en outre un suivi temporel efficace (*tracking*) des zones détectées en mouvement dans la séquence, et une estimation de leur vitesse et trajectoire [Cap95]. Un exemple de détection multi-étiquette est donné Fig. 3.9.

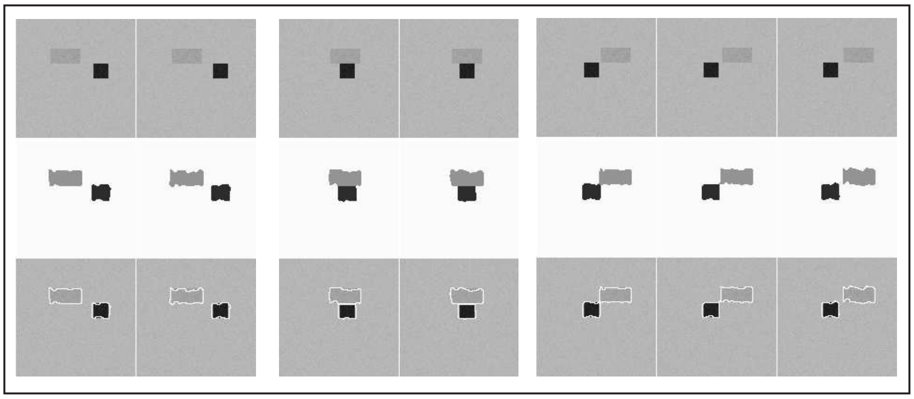


FIG. 3.9 - *Détection de mouvement multi-étiquette, permettant de distinguer différents mobiles en les étiquetant avec un niveau de gris distinct (ici un carré et un rectangle qui se croisent). En haut : la séquence synthétique. Au milieu : les champs d'étiquettes obtenus. En bas : superposition des contours des masques détectés sur les images originales.*

Segmentation

Contrairement à la détection, le terme de segmentation sous-entend généralement l'usage (et le calcul) de critères discrimants pour distinguer différentes régions de l'image : estimation d'un modèle de mouvement (constant, linéaire ou quadratique), d'une teinte, d'une texture, etc ... par zone d'image.

La segmentation spatio-temporelle de séquences d'images revêt un grand intérêt dans le cadre des nouveaux besoins apparus en télécommunications audio-visuelles (norme de compression MPEG), que ce soit pour la transmission et l'archivage de séquences vidéo, l'indexation par le contenu de scènes dynamiques, la visiophonie haut de gamme, les téléconférences, la réalité virtuelle, les interfaces homme/homme et homme machine.

L'approche MRF spatio-temporelle est utilisée pour la segmentation spatio-temporelle et le suivi automatique des régions correspondant aux lèvres d'un locuteur, dans des conditions naturelles d'éclairage et sans maquillage particulier, ce qui fait l'objet de la troisième thèse que je dirige actuellement [Lie00] (cf. §3.2.4). Cette segmentation est basée sur un double critère de teinte et de mouvement. Le choix d'un voisinage 3-D pour modéliser les interactions entre pixels s'avère très pertinent pour cette application [LCL99]. Un exemple de résultat de segmentation est donné Fig. 3.10.

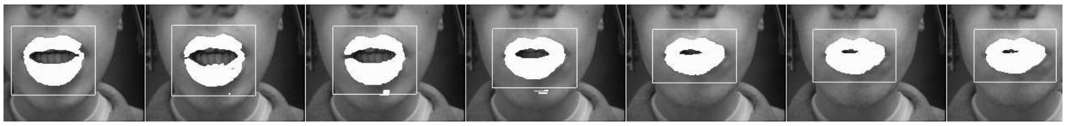


FIG. 3.10 - *Segmentation des lèvres d'un locuteur avec détermination automatique de la zone d'intérêt (bounding box).*

Concernant plus particulièrement la segmentation au sens du mouvement, on étudie à présent les méthodes applicables au cas d'une caméra mobile (compensation de mouvement basée sur les modèles de mouvement et les estimateurs robustes) [Riq96] et les possibilités de mise en œuvre rapide sur DSP [Chb98].

Estimation

L'estimation est une tâche conjointe à la segmentation : elle porte sur le calcul local de primitives bas-niveau (par exemple l'estimation d'un vecteur-vitesse par pixel, réflétant son déplacement d'un image à la suivante). Mais on se trouve confronté à un problème bien connu en traitement d'image : celui "de la poule et de l'œuf". Laquelle des deux tâches, segmentation ou estimation, doit précéder l'autre ? Une estimation réalisée sur deux zones de nature différente sera biaisée (moyenne des informations des deux zones). Une pré-segmentation s'avère donc nécessaire pour guider l'estimation sur des zones homogènes selon un certain critère. Mais a contrario, pour réaliser une bonne segmentation, il faudrait disposer d'une estimation correcte des primitives pertinentes. Il s'agit là d'un problème *a priori* insoluble. Je préciserai dans le bilan ma position sur ce point.

En ce qui concerne l'estimation de mouvement, nous nous intéressons au calcul du flux optique (champ dense de vecteurs-vitesses) selon trois types d'approches :

– les approches différentielles basées sur l'équation de contrainte du mouvement (ECM) :

$$\frac{\partial I}{\partial x}v_x + \frac{\partial I}{\partial y}v_y + \frac{\partial I}{\partial t} = 0 \quad (3.5)$$

Cette équation traduit l'hypotèse de constance spatio-temporelle de l'illumination dans une scène au cours du déplacement. Mais elle porte sur des informations très locales et pose le problème classique d'ouverture, qui est résolu en ajoutant des contraintes de lissage [HS81, NE86, Nag87]. Nous étudions l'association de ce type d'approche avec la transformée en ondelettes pour les applications en codage vidéo (méthodes hybrides basées sur la DFD (*Displaced Frame Difference*)) [Fra94, Sim95, Chr96].

– les approches par mise en correspondance [ADM81], qui consistent en la recherche (exhaustive ou non) de blocs similaires entre deux images consécutives. L'implantation de ce type de méthode dans un cadre hiérarchique [Ana89] (pyramides passe-bas) permet de réduire les temps de calculs [Bou96].

– les approches fréquentielles [Hee87], basées sur le fait que le spectre d'énergie d'un bruit blanc en mouvement de translation dans l'image est concentré dans un plan de l'espace des fréquences spatio-temporelles :

$$f_x v_x + f_y v_y + f_t = 0 \quad (3.6)$$

Ce troisième type d'approche a été l'occasion d'une coopération entre l'équipe Vision et l'équipe Neurones du LTIRF, car la décomposition en fréquences spatio-temporelles correspond, dans l'approche biologique du processus de vision, à l'analyse faite par l'œil pour traiter l'information de mouvement. Dans le cadre du programme TEMPRA de collaboration avec les pays de l'est, j'ai recruté un étudiant roumain de l'Université Polytechnique de Bucarest (UPB) qui a soutenu une thèse [Spi98] dirigée par un collègue du laboratoire, avec qui j'ai initialement collaboré sur l'approche fréquentielle avec des filtres de Gabor [Yao94].

Interprétation

L'interprétation consiste à produire une information de haut niveau, par exemple de type sémantique (reconnaissance d'objets, structuration d'une scène 3-D).

La vocation du laboratoire concernant plutôt les aspects bas niveau, l'interprétation haut niveau, qui relève du génie informatique et de l'intelligence artificielle, n'entre pas directement dans notre champ d'activité.

Cependant, certaines informations sémantiques pertinentes peuvent être extraites facilement de nos traitements bas-niveau, comme par exemple l'information d'ouverture/fermeture des lèvres à partir des masques segmentés. Ce type de renseignement peut ensuite être utilisé dans un processus de fusion pour la reconnaissance audio-visuelle.

3.2.4 Applications

Vision Embarquée

Parmi les multiples applications de la vision par ordinateur (robotique, contrôle non destructif, diagnostic médical, météorologie...), nous avons surtout investigué les applications de la détection de mouvement à la télésurveillance (entrées de bâtiments, terrains ...) et au contrôle du trafic routier. La Fig. 3.11 est une illustration typique de surveillance vidéo d'une scène de rue.

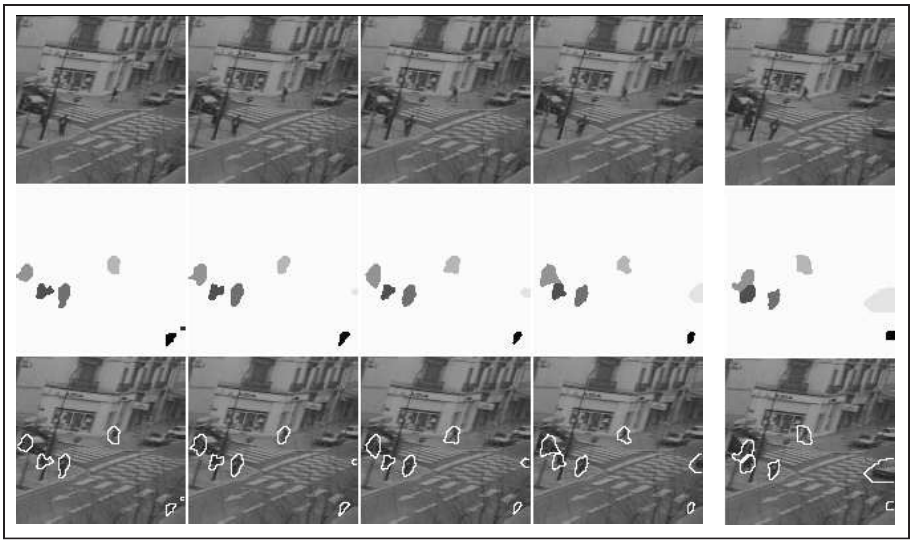


FIG. 3.11 - *Télésurveillance vidéo. En haut : la séquence d'images (scène de rue avec piétons et voiture). Au milieu : les masques obtenus par détection de mouvement multi-étiquette, permettant de distinguer et suivre les différents mobiles. En bas : superposition des contours des masques détectés sur les images originales.*

Ce type d'application requiert un matériel embarqué (autonomie, faible encombrement, faible coût, fiabilité) et fonctionnant en temps-réel, d'où l'importance des mises en œuvre matérielles des algorithmes, qui font l'objet du §3.2.5.

Télécommunications

Un autre champ d'applications en plein développement à l'heure actuelle est celui des télécommunications audiovisuelles : compression de séquences d'images pour la transmission de TV numérique (standard MPEG [Kle95]) ou l'archivage vidéo, interfaces multi-modales homme/machine.

Dans le cadre de la fédération de laboratoires ELESA, le projet Labiophone auquel je participe vise au développement d'une plate-forme de communication homme-homme et homme-machine par la parole audio-visuelle. Il est en effet connu que la lecture labiale (et faciale) permet à un individu, malentendant ou non, de rehausser substantiellement l'intelligibilité d'un message acoustique dégradé (bruit, langue étrangère ...) La parole est en fait un vecteur de communication multi-sensoriel.

D'où l'idée de prendre en compte la dimension visuelle du langage pour un traitement automatique, que ce soit dans le sens machine → homme (synthèse de visages parlants), dans le sens homme → machine (reconnaissance automatique de la parole) ou dans le sens machine → machine (compression conjointe du son et de l'image pour codage bas débit).

Le contexte du projet est décrit sur la figure Fig.3.12.

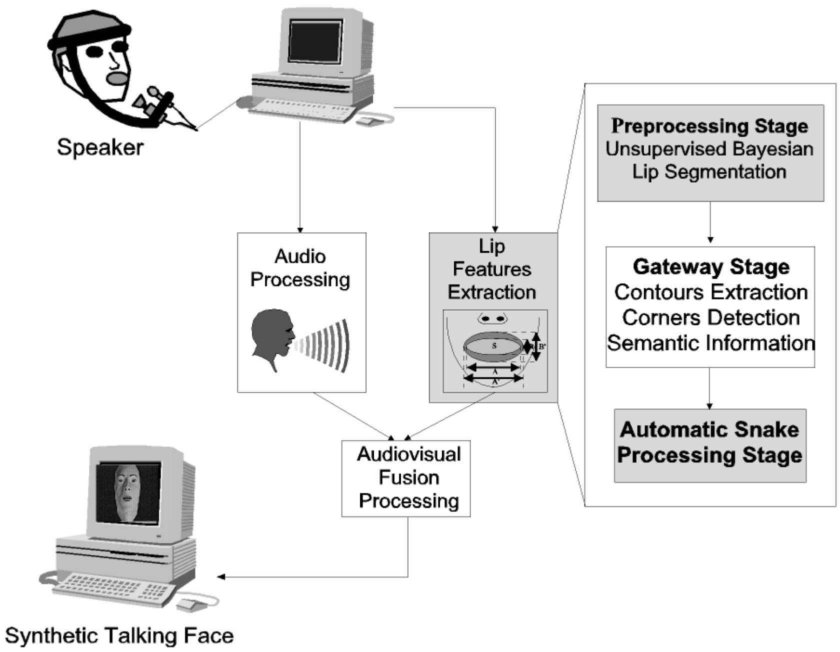


FIG. 3.12 - *Contexte du projet Labiophone.*

Le LIS participe activement à ce projet à deux niveaux : d'une part au niveau de l'analyse de visages parlants (segmentation de lèvres et estimation de leurs contours, analyse géométrique pour l'extraction de points pertinents en vue du rendu réaliste et de l'animation d'un modèle texturé de visage de locuteur), et d'autre part au niveau de la fusion audio-visuelle.

Notre contribution à ce projet se situe au niveau de la phase de prétraitement qui consiste à extraire automatiquement les masques des lèvres et la région d'intérêt de la bouche [LL97a, LL97b, LL98a, LL98b]. Ce prétraitement fournit des points initiaux fiables qui permettent ensuite de lancer un processus de contours actifs (*snakes*) qui viennent converger précisément sur les contours des lèvres (Fig. 3.13). Ce travail est le fruit d'une collaboration étroite avec Pierre-Yves Coulon et son thésard Patrice Delmas [LDC+99].

3.2.5 Mises en œuvre Matérielles et Logicielles

En parallèle aux études théoriques, le laboratoire, dont la compétence reconnue porte sur la réalisation de circuits, de cartes électroniques et d'architectures adaptées au traitement rapide d'images, s'intéresse aux possibilités de mise en œuvre matérielle des algorithmes. Nous étudions plus particulièrement l'implantation des algorithmes de détection de mouvement par champ de Markov et de segmentation des lèvres sur trois types de matériels, en vue d'atteindre un

FIG. 3.13 - *Extraction automatique de contours de lèvres. En haut et au milieu : points initiaux (externes et internes) fournis par prétraitement automatique. En bas : contours actifs positionnés sur les lèvres.*

traitement “temps réel”, c’est-à-dire à la cadence vidéo (25 images/s). Pour ces mises en œuvre, nous collaborons fructueusement avec d’autres chercheurs des trois équipes de l’ex LTIRF, collaboration valorisée par des publications dans des revues [CDLC96, DLC98b].

Circuits Cellulaires Analogiques (ASIC VLSI)

Ce travail, initié en 1993 suite à une discussion avec Jeanny Hérault de l’équipe Neurones autour du concept de rétine en silicium, porte sur la mise en œuvre sur un circuit intégré mixte analogique-numérique de l’algorithme de détection de mouvement par MRF. L’idée est de réaliser la minimisation d’énergie, non pas de façon logicielle par un algorithme de relaxation déterministe ou stochastique (du type recuit simulé ou ICM), mais de façon matérielle en laissant relaxer un réseau résistif vers son état de dissipation minimale d’énergie.

La Fig. 3.14 présente l’architecture de l’ASIC détecteur de mouvement ainsi qu’un layout de la cellule élémentaire du réseau, qui intègre à la fois le capteur (photo-récepteur) et le traitement bas niveau (détection de mouvement).

Les simulations électriques effectuées sont concluantes et permettent d’envisager la réalisation d’une caméra “intelligente” qui détecte en temps-réel le mouvement dans la scène observée. Mais la contre-partie à ce capteur intelligent est un facteur de remplissage pas encore satisfaisant : seulement 10 % de la cellule sont occupés par le photo-récepteur, le reste étant occupé par les circuits de traitement et les connexions. La conception du circuit, qui utilise la technique des courants commutés, se heurte aux problèmes de la haute intégration en technologie CMOS, et est actuellement à l’étude. Nous utilisons, pour la conception du circuit et les simulations, les facilités dont dispose le laboratoire, puisqu’il est situé à proximité du CIME (Centre Interuniversitaire de MicroElectronique) et collabore avec le CNET de Meylan sur un projet de rétine en silicium [Sic99].

J’ai encadré trois stages d’étudiants de l’UPB sur ce projet [Pop93, Pop94, Dra96]. Ce travail a fait l’objet de plusieurs publications [Lut93, LPC94b, LPC94a, LP95, PL95], dont un article dans la revue IEEE Trans. on Circuits and Systems [LD99]. Le projet se poursuit en collaboration avec Gérard Bouvier de l’équipe Circuits Spécifiques (co-encadremenet d’un DEA de microélectronique [ElA98]).

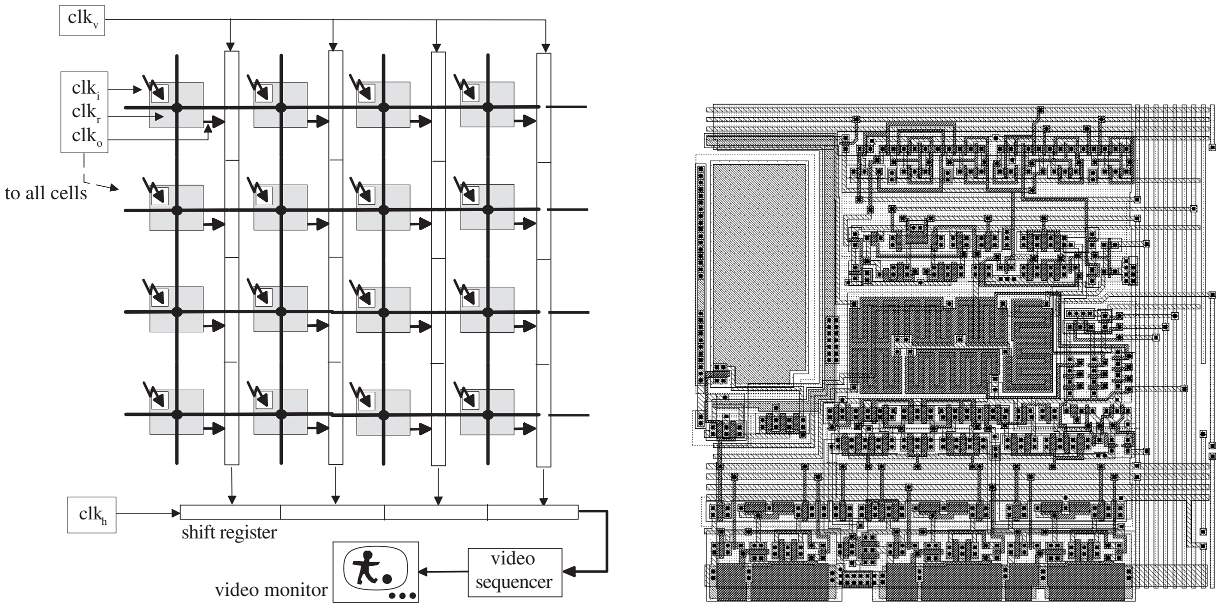


FIG. 3.14 - *A gauche) Architecture du réseau VLSI analogique ; A droite) Layout d'une cellule élémentaire (technologie CMOS* $1.2\mu m$*).*

Composants Programmables Standards (DSP, FPGA)

Ce second type d'implantation a fait l'objet de la deuxième thèse que j'ai dirigée [Dum96]. Il s'agit de la mise en œuvre de la détection de mouvement en temps-réel sur une carte d'acquisition et de traitement au format PC (bus ISA), à base de circuits logiques programmables FPGA (Field Programmable Gate Array) et d'un DSP Motorola 96002 de fréquence d'horloge 33 MHz (Fig. 3.15).

Cette carte, développée au laboratoire par J.P. Charras, avait été initialement conçue pour la détection de contours. Une étude architecturale des besoins pour la détection de mouvement a conduit à la réalisation d'un second prototype opérationnel. Le point clé de cette architecture est le traitement asynchrone des données, rendu possible par la présence d'une mémoire tampon.

Notre implantation aboutit à une carte électronique autonome, qui réalise la détection de mouvement à une cadence de traitement de 15 images/s pour une taille d'image 128×128. Cette mise en œuvre a fait l'objet de plusieurs publications [DLC96b, DLC96a, DLC97, CLD95, CLD98], dont un article dans la revue IEEE Trans. Imaging Processing [DLC98b]. Une réflexion architecturale est proposée dans [DLC98a]. Ce travail s'est déroulé en étroite collaboration avec J.P. Charras et P.Y. Coulon de l'équipe Vision pour la partie électronique (DSP, FPGA).

Machines Multi-Processeurs Parallèles

Un troisième type d'implantation matérielle concerne l'utilisation de machines parallèles de type SIMD.

L'algorithme de détection de mouvement a d'abord été implanté sur une machine parallèle SIMD à 256 processeurs (la machine CNAPS de chez Adaptive Solution). Une cadence de traitement d'environ 12 images/s est obtenue. Mais, cette machine n'étant pas dédiée au traitement d'images vidéo, mais plutôt aux réseaux de neurones, elle ne nous a servi que de critère de comparaison par rapport aux autres implantations matérielles.

Une collaboration en cours avec le CNET Meylan concerne l'implantation de nos algorithmes

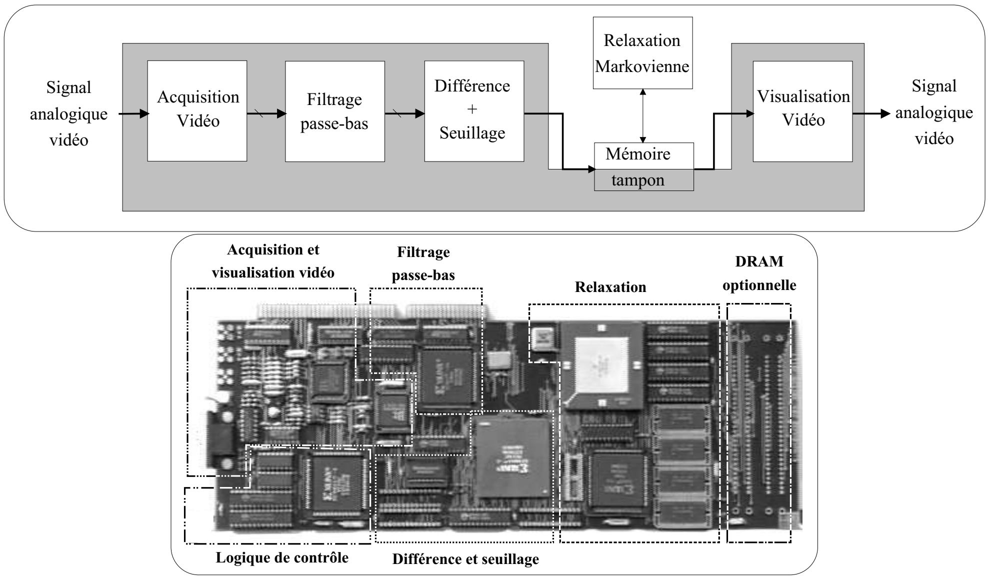


FIG. 3.15 - *Carte PC prototype développée au laboratoire, sur laquelle est implantée la détection de mouvement en temps réel. En haut) Architecture de la carte : la partie grisée correspond aux étages fonctionnant en mode synchrone, c'est-à-dire à la cadence du flux pixel. En bas) Photo du prototype.*

sur un processeur vidéo parallèle programmable (le PVP) développé par le CNET. Il s'agit d'une structure SIMD à 8 (ou 16) processeurs avec 7 flux vidéo d'entrée/sortie. Elle offre une puissance de traitement de 2 Gops et autorise des entrées/sorties à haut débit (4 Gbits/s). Pour la détection de mouvement, une cadence de 150 images/s a été obtenue sur le simulateur logiciel du PVP. Quant à la segmentation des lèvres, une cadence de 15 images/s a été évaluée. J'ai encadré deux stages-ingénieurs sur ce projet [Lab98, Gje98]. La conception et la réalisation d'une carte au format PC (bus PCI) équipée d'un processeur PVP doit faire l'objet d'un contrat avec le CNET.

En parallèle, une mise en œuvre sur le processeur vidéo TMS320C80 (structure MIMD avec 4 DSPs et un processeur RISC) est prévue, ce qui permettra de comparer les performances de ces deux types de processeurs vidéo d'architecture assez différente.

Environnement Logiciel MAIS

J'ai été rapidement confronté, en encadrant de nombreux stagiaires chaque année, aux problèmes d'insertion informatique du stagiaire, et de pérennité et réutilisation des logiciels développés. C'est pourquoi j'ai développé sous OpenWindows un environnement de programmation avec interface graphique conviviale (Fig. 3.16), qui permet de manipuler des séquences d'images (allocation, affichage statique et dynamique, stockage, visualisation, conversion de format).

Cette interface est largement utilisée par de nombreux stagiaires et thésards du laboratoire, encadrés par moi-même ou par mes collègues. Chaque nouvel utilisateur peut ainsi se concentrer sur la phase de programmation des traitements proprement dits, et insérer sa contribution dans cette interface, tout en bénéficiant des algorithmes développés par ses prédécesseurs.

Cette interface évolue en permanence avec le temps, avec les suggestions des utilisateurs, avec les nouvelles versions de XView, s'acheminant peu à peu vers la portablilité (Unix, Linux), et essayant de respecter une démarche de "qualité logicielle".

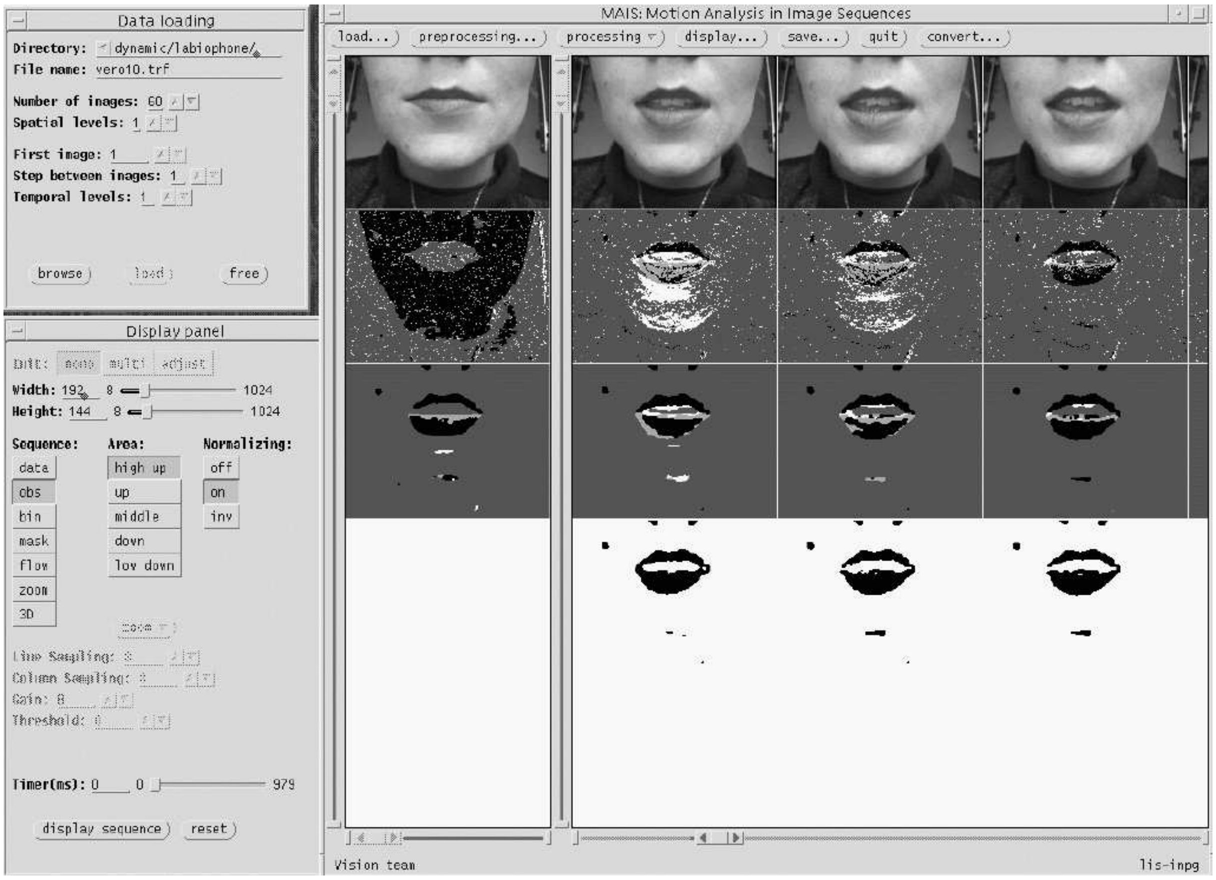


FIG. 3.16 - *Interface MAIS: la fenêtre de visualisation gauche sert à l'affichage dynamique des séquences, celle de droite à l'affichage statique des images. Les menus de chargement et d'affichage sont représentés à gauche.*

3.2.6 Relations Industrielles, Collaborations, Contrats

En 1994, une collaboration avec SGS-Thomson portant sur la compression de séquences d'images (MPEG) a fait l'objet d'un contrat dans le cadre d'une thèse CIFRE. Malheureusement, l'étudiante en thèse a démissionné à mi-parcours pour des raisons personnelles. J'en étais le responsable au LTIRF (responsable industriel : R. Bramley).

En 1995, un contrat de collaboration avec le CNET a été signé pour la réalisation d'une rétine en silicium pour la détection de contours et de mouvement. Le responsable industriel est P. Senn et le responsable au LTIRF est le G. Bouvier. Nos travaux sur le circuit VLSI implantant l'algorithme markovien de détection de mouvement bénéficient de cette collaboration et la fructifient.

Le LIS est fédéré avec sept autres laboratoires grenoblois rattachés au CNRS au sein de la fédération ELESA en Automatique-Traitement du Signal. Un important projet soutenu par la fédération est le projet Labiophone (visiophone haut de gamme bas débit avec fusion d'informations auditives et visuelles venant du locuteur). Ce projet fait intervenir, outre le LIS et l'ICP (Institut de la Communication Parlée), plusieurs industriels de la région en partenariat (SGS-Thomson, CNET-CNS, Ganymedia ...). Un des thèmes d'étude de ce vaste projet concerne l'analyse du mouvement d'un visage parlant (lèvres, menton), thème dont je suis co-responsable avec J.L. Schwartz de l'ICP et P.Y. Coulon du LIS. J'ai encadré sur ce sujet un DEA [Lie96a], deux stages ingénieurs [Lie96b, Asf97], et je dirige actuellement une thèse démarrée en octobre 1997 [Lie00].

Toujours dans le cadre de la fédération ELESA, je collabore avec J. Brini du LPCS sur l'analyse du bruit dans les composants électroniques, basée sur des méthodes utilisant les moments d'ordre supérieur (co-encadrement d'un DEA de microélectronique [Bon98]).

En 1998, nous avons contracté avec Thomson-Plasma par l'intermédiaire d'INPG-Entreprise pour le développement d'algorithmes visant à l'amélioration du rendu visuel de la couleur (*dithering*) sur écrans à cristaux liquides, qui ne disposent que de 4 bits pour coder chaque composante couleur. Les responsables au LIS étaient A. Caplier et moi-même (responsable industriel : P. Zorzan). Cette prestation a été assurée par mon thésard M. Liévin.

En 1997, une collaboration a été initiée avec le CNET-Meylan autour du processeur PVP, à la fois sur des aspects logiciels et matériels : mise en œuvre de nos algorithmes sur le processeur, et conception d'une carte au format PC (bus PCI) intégrant le processeur PVP, pour le traitement temps-réel de séquences vidéo. Les responsables de ce projet sont D. Barthel du CNET Meylan, et moi-même au LIS.

L'interface logicielle MAIS semble également intéresser le CNS/CNET, qui accueille chaque année de nombreux stagiaires, pour résoudre le même type de problème que j'ai moi-même rencontré avec mes stagiaires : offrir un environnement de travail permettant à tout nouveau développeur de se concentrer sans perte de temps sur la phase de programmation de ses traitements, sans avoir à re-développer sa propre interface pour la gestion de données vidéo et l'affichage graphique.

Enfin, nous participons au GDR PRC ISIS, le groupe de recherche du CNRS sur le traitement du signal et des images, par des présentations dans des réunions informelles [CL93, Lut93] ou des conférences avec actes [DLC96a].

3.2.7 Relations Internationales

Dans le cadre du programme TEMPRA de collaboration de la région Rhône-Alpes avec les pays de l'est, j'ai initié une collaboration entre le LTIRF et le département d'Electronique Appliquée de l'Université Polytechnique de Bucarest (Pr. V. Buzuloiu et D. Steriu). Depuis 1993, le LTIRF a accueilli 6 étudiants boursiers roumains (stages de fin d'études, DEA, thèses en co-tutelle) et je me suis rendu deux fois à Bucarest pour le recrutement des stagiaires. Une collaboration bilatérale plus équilibrée entre Grenoble et Bucarest est cependant souhaitable et je m'efforce d'y contribuer.

J'ai reçu par ailleurs une invitation pour assurer un séminaire en analyse du mouvement à l'occasion d'une école internationale de printemps organisée à Bucarest par l'Université Polytechnique tous les ans.

3.3 Bilan

En résumé, mon activité au sein du laboratoire a conduit à la constitution d'une petite équipe de recherche autour d'un thème nouveau pour le laboratoire : l'analyse du mouvement vidéo. Mon effort a porté sur l'encadrement des étudiants sous ma responsabilité, à la fois sur le plan théorique et sur le plan informatique (développement d'un environnement logiciel commun avec interface graphique sous OpenWindows pour la manipulation, le traitement et la visualisation de séquences d'images), ainsi que sur la valorisation des travaux de l'équipe (publications et contrats). Depuis la rentrée 1996, je bénéficie d'une Prime d'Encadrement Doctoral et de Recherche.

Le laboratoire a désormais acquis une compétence reconnue en analyse du mouvement dans les séquences d'images. Ce nouveau savoir-faire, auquel j'ai contribué, porte à la fois sur des aspects théoriques de traitement (modélisation MRF, multirésolution ...), sur des aspects d'implantation matérielle (architectures, circuits, cartes ou machines de traitement rapide) et enfin sur des aspects applicatifs (télésurveillance, compression, Labiophone ...). Ces travaux ont été l'occasion de renforcer sur plusieurs thèmes la collaboration entre les trois équipes Vision, Réseaux Neuronaux et Circuits Spécifiques, ce qui est un point positif pour le laboratoire.

3.3.1 Points Clé d'un Traitement

Le problème de segmentation de lèvres d'un locuteur dans des conditions naturelles d'éclairement et de maquillage nous a fait sentir l'importance de quatre points clés dans une chaîne de traitement de séquences d'images : l'étape d'acquisition du signal vidéo, le choix des observations, la connaissance d'informations *a priori*, et enfin l'adéquation algorithme/architecture (cf. Fig. 3.17).

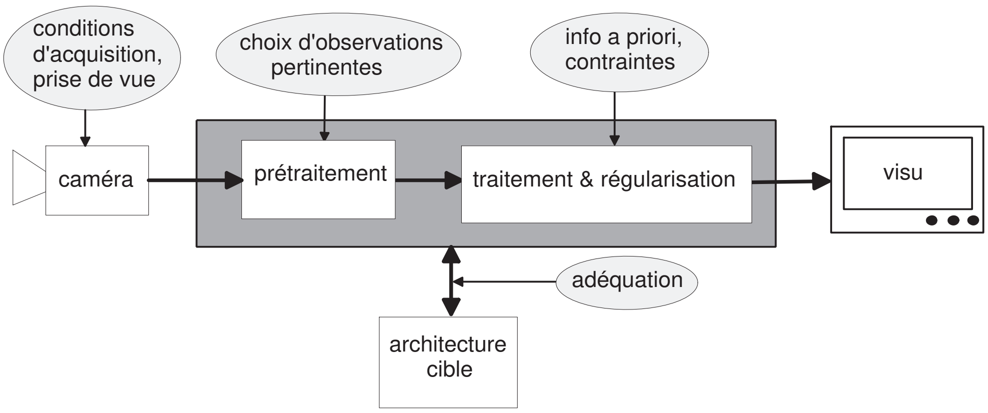


FIG. 3.17 - *Points clés dans une chaîne de traitement.*

- Conditions d'acquisition du signal vidéo : il n'est pas inutile de rappeler ici une petite évidence : meilleure est l'acquisition, moindres sont les performances requises du traitement. Utiliser une caméra professionnelle tri-CCD, ou une micro-caméra mono-CCD pour acquérir les séquences couleur implique des contraintes de robustesse différentes au niveau du traitement. Dans tous les cas, il est donc essentiel de bien évaluer la phase d'acquisition (réglage des blancs, éclairage, corrélation des plans couleurs, ombres et reflets) avant de se lancer dans un traitement.

- Importance des observations choisies pour réaliser un traitement donné : quelles que soient la performance et la robustesse de l'algorithme utilisé, le choix de différentes observations, complémentaires et non redondantes, est essentiel en traitement d'images : en l'occurrence, la teinte, le contraste sur la luminance, l'information de mouvement, voire même la texture, doivent être pris en compte simultanément dans un unique processus de segmentation pour optimiser les résultats.

- Importance de l'information *a priori* : l'intégration conjointe d'information spatiale et temporelle sous contraintes constitue l'originalité de notre approche et s'avère essentielle dans un projet comme le Labiophone. La plupart des travaux sur le suivi des lèvres consistent en effet en une segmentation uniquement spatiale, image par image.

- L'Adéquation entre l'Algorithme est l'Architecture cible est un domaine de recherche à part entière (cf §3.3.2 ci-desous).

3.3.2 Prises de Position

Mon Point de Vue sur l'Evolution du Hard

Concernant les études d'architectures matérielles, plusieurs points sont à souligner. Plus que le matériel final ou une réalisation concrète et opérationnelle, c'est surtout la démarche et les concepts architecturaux qui comptent (cf. réflexion architecturale proposée dans [DLC98a] où l'on présente deux types de modules : module "soft" type FPGA + module "hard" type DSP). La clé d'une mise enœuvre temps-réel réside souvent plus dans une bonne gestion des communications, que dans la cadence de traitement du ou des processeurs élémentaires. Peu de systèmes opérationnels intégrant *toute* la chaîne de traitement vidéo, depuis l'acquisition jusqu'à la visualisation, sont présentés dans la littérature. C'est sur ce créneau que nous nous positionnons au laboratoire.

Mais la valorisation de ce type de travail *hardware* représente un double challenge. D'une part, l'évolution rapide de la technologie des processeurs, rend vite caduque une réalisation matérielle initiée deux ou trois ans auparavant à l'occasion d'un travail de thèse par exemple. Car la prochaine génération de processeurs génériques (type Pentium) vient souvent outre-passer les performances d'une machine dédiée. Et la programmation séquentielle supplante la programmation ardue en langage assembleur ou en langage parallèle. L'avenir me semble donc davantage résider dans l'utilisation de processeurs génériques (du type processeurs vidéo intégrant un certain parallélisme), plutôt que dans les machines dédiées.

D'autre part, la publication de ce genre de travaux représente un challenge et demande un effort certain, que je crois avoir su fournir pour relever ce défi (cf. publications dans des revues). On ne peut donc nier que la recherche *hardware* présente un certain handicap, qu'il faut savoir surmonter et valoriser.

Contours ou Régions ?

Dans mes choix de recherche, j'ai un parti pris pour les approches régions plutôt que contours. La détection de contours est en effet une tâche délicate (problèmes non triviaux de seuillage par hystérésis, de chaînage de contours, d'approximation polygonale [Run92]) qui, même une fois résolue, doit encore subir un post-traitement, pour ensuite seulement pouvoir être exploitée en vue d'en déduire une information de forme extérieure des objets. L'approche région me semble plus proche de l'information sémantique, et donc plus proche des besoins d'interprétation à haut niveau.

Cette remarque est aussi valable pour le flux optique dense, qui est toujours difficile à estimer et fournit rarement une information exploitable en soi. Un flot épars, ou un vecteur vitesse attaché à chacun des objets segmenté, est une information plus directement exploitable à plus haut niveau. C'est pourquoi j'oriente mes recherches davantage vers la segmentation que vers l'estimation du flux optique.

Mais en fait, c'est dans la **coopération** entre les deux types d'approches (contours et régions, ou estimation et segmentation) que réside la clé d'un traitement performant, comme on l'illustrera au §1.4.3, p. 60.

Modèles Mathématiques *versus* Modèles Biologiques ?

Contrairement à l'équipe Réseaux Neuronaux du laboratoire qui axe ses recherches sur les modèles biologiques, psychovisuels et cognitifs pour les appliquer au domaine des sciences de l'ingénieur, j'ai quant à moi une préférence pour les modèles mathématiques plutôt que biologiques, due en partie à ma formation initiale d'ingénieur[3].

J'ai conscience que ma recherche est de ce fait plus avale, et plus modestement utilitariste, visant à des applications concrètes. L'effort de pluridisciplinarité est une œuvre de recherche plus amont et plus fondamentale, plus ambitieuse aussi, et plus noble, mais également à plus long terme. Mais, à ma connaissance, les algorithmes issus de modèles biologiques ont rarement approché l'état de l'art de la vision par ordinateur basée sur les approches "computationnelles", ou du moins pas dans les mêmes délais. Cependant, sur le long terme, nul doute que c'est dans les approches inspirées de la biologie et de la psycho-vision que réside le plus gros potentiel à venir (chronique d'une rétine annoncée). Mais dans l'intervalle, les modèles mathématiques ont leur rôle à jouer, en vue d'applications effectives rapidement, et ceci pour un certain temps encore ...

De mon point de vue, et suivant mon expérience, l'inspiration à partir du système de perception humain ou animal (perception du mouvement, des couleurs) peut être utile pour guider la modélisation mathématique.

3.3.3 Expérience Personnelle

J'aimerais finalement souligner quelques points issus de mon expérience personnelle :

- D'abord la motivation par les applications et les mises en œuvre. Ma formation d'ingénieur me fait me tourner plutôt vers la recherche appliquée. En outre, les mises en œuvre temps-réel sont un point incontournable quand il s'agit d'analyse du mouvement, car détecter le

3. Une citation de Poincaré, qui décrit parfaitement l'approche MRF, éclaire ma position : "C'est donc grâce à l'homogénéité approchée de la matière étudiée par les physiciens que la physique mathématique a pu naître. Dans les sciences naturelles, on ne retrouve plus ces conditions : homogénéité, indépendance relative des parties éloignées, simplicité du fait élémentaire, et c'est pour cela que les naturalistes sont obligés de recourir à d'autres modes de généralisation." ([Poi68], p. 172).

mouvement avec un retard important rend inutile le résultat (évitement d'obstacle trop tardif ...)

- La publication dans certaines revues comme les IEEE Trans. représente un exercice de style en soi, nécessitant de respecter à trois points principaux : la maîtrise de la langue évidemment, le positionnement du travail pour rapport à la communauté scientifique (chapitre "related works"), et enfin l'évaluation du travail (évaluation de type statistique par exemple pour étayer la robustesse d'un algorithme). Cette démarche est en règle générale plus ancrée dans la culture anglo-saxonne que dans la nôtre.

- La participation à des conférences est bénéfique à plus d'un titre : pour l'occasion qui est donnée de nouer des contacts scientifiques, pour la publicité et l'image de marque du laboratoire, et pour suivre l'évolution de l'état de l'art, les tendances et les modes scientifiques.

- Ma mobilité géographique (LETI, Suède, LTIRF) et thématique (Electronique, Signal, Image, Parole) a finalement été pour moi un atout plus qu'un handicap.

- Enfin, l'importance du travail d'équipe me paraît essentielle dans le domaine de la recherche, où l'individu isolé connaît parfois des traversées du désert désespérantes. Ayant travaillé essentiellement seul pendant ma thèse, j'ai pu apprécier a contrario tout le bénéfice de la collaboration en équipe au sein du LTIRF.

Chapitre 4

Enseignement

4.1 A l'ENSERG

L'Ecole Nationale Supérieure d'Electronique et Radioélectricité de Grenoble (ENSERG) est une des dix écoles rattachées à l'Institut National Polytechnique de Grenoble (INPG). Elle forme des ingénieurs électroniciens en trois ans. Les promotions sont d'environ 120 étudiants. Les deux permières années consistent en un enseignement de tronc commun (mathématique, physique, électronique de base ...) alors que la troisième année apporte une spécialisation dans une des trois options proposées par l'école : systèmes intégrés, traitement de l'information, systèmes de transmission.

J'interviens principalement dans les enseignements de 2ème année (cours, TD et TP), et dans une moindre mesure dans ceux de première année (TD), ainsi que dans les projets de troisième année (tutorats).

Mon activité d'enseignement s'articule autour de trois pôles :

- les mathématiques de l'ingénieur (Cours et TD) : 60h
- le traitement du signal et la théorie de l'information (TD) : 40h
- l'électronique et l'automatique (TP) : 150h

Ces trois pôles distincts s'organisent de façon assez cohérente autour du traitement du signal, en partant des outils théoriques jusqu'aux aspects pratiques et matériels.

En effet, le cours et les TD de mathématique que j'assure en 1ère année abordent les principaux outils nécessaires au traitement du signal : fonctions de la variable complexe et théorème des résidus, transformées de Laplace, de Fourier, en Z, séries de Fourier, systèmes d'équations différentielles et équations aux dérivées partielles. J'ai rédigé un document de cours destiné aux étudiants de première année [Lut96].

Les travaux dirigés que j'encadre en 2e année sont plus directement couplés à mon activité de recherche. Il s'agit, en effet, d'une part de traitement du signal : rudiments sur les distributions, étude spectrale des signaux certains, filtres, signaux à bande étroite, échantillonnage, traitement et filtrage numériques, fonctions aléatoires, processus faiblement stationnaires, bruit dans les systèmes de transmission. D'autre part de théorie de l'information : notions de source et d'entropie, extension d'une source, capacité d'un canal, codage sans bruit et avec bruit (codage source et codage canal), efficacité, théorèmes de Shannon, codes linéaires, correcteurs, cycliques, cas des signaux continus. Ces TD sur la théorie de l'information m'ont servi pour l'approche statistique par champ aléatoire de Markov que j'utilise, pour la compression d'images et, d'un point de vue plus philosophique, pour la façon d'envisager un problème en terme d'information ou entropie. A l'inverse, mon activité de recherche sur les MRF m'a permis d'enrichir les TD au sujet des sources de Markov.

Enfin, en tant que responsable de la plate-forme de TP2A (Travaux Pratiques) depuis 1992, j'ai mis en place ou rénové un certain nombre de manipulations, dont quatre en rapport direct avec le traitement de signal numérique :

- analyseur de spectre numérique à FFT
- transformée de Fourier numérique : initiation à Matlab
- transformée en Z et systèmes discrets : simulation avec Matlab
- asservissement numérique d'une pompe à chaleur : identification avec Matlab
- instrumentation programmée par bus GPIB : environnement LabWindows
- amplificateur contre-réactionné à circuits intégrés avec didacticiel Toolbook

J'encadre une vingtaine de manipulations d'électronique, automatique et traitement du signal sur les thèmes suivants (manipulations de 4h en binômes) :

- composants électroniques : diodes et transitors en commutation, transistor MOS,
- circuits électroniques : oscillateurs, relaxateurs, ampli BF, contre-réaction, TV noir et blanc,
- traitement de signal analogique et numérique : transformée de Fourier discrète, analyseur de spectres analogiques et à FFT, transformée en Z et systèmes discrets, initiation à MATLAB
- modulations : d'amplitude, de fréquence, récepteur AM, TV couleur
- asservissements analogiques et numériques : PLL, asservissement de position et de vitesse, pompe à chaleur,
- instrumentation programmée : bus GPIB, environnement LabWindows

4.2 En Formation Continue

J'ai par ailleurs une expérience en formation continue :

- Conférences de Traitement de Signal : de 1988 à 1994, je suis intervenu dans des stages biannuels de Traitement de Signal organisés par l'INSTN au CENG/CEA, destinés aux techniciens et ingénieurs qui veulent se perfectionner en traitement de signal. J'y assurais une conférence sur les nouvelles méthodes de filtrage et d'analyse spectrale : prédiction linéaire d'un signal, modélisation paramétrique AR,ARMA, modèle deprony, algorithmes de SVD, des moindres carrés en treillis... Un document de 50 pages est rédigé à l'attention des stagiaires [Lut90a].
- Cours et TD de mathématiques : au cours de ma thèse, j'ai assuré plusieurs vacations à la Formation Continue de l'INPG pour des techniciens de différentes entreprises de la région grenobloise (Merlin-Gerin, Thomson, HP ...) et au CUEFA (centre CNAM de Grenoble) pour des étudiants de DEST (maths B1)

4.3 Quelques Remarques

Cette expérience en formation continue, de même que le contact avec les admis sur titres de l'ENSERG issus de l'industrie (à l'occasion des remises à niveau en maths que j'assure en début d'année), est très enrichissante, car elle permet d'avoir un retour du monde industriel et ainsi de porter un éclairage différent sur les qualités et les défauts de la formation initiale en école d'ingénieur. Notamment, mon expérience personnelle m'a appris qu'il serait souhaitable d'éviter l'effet "boîte noire" qu'on rencontre trop souvent dans les salles de travaux pratiques : les élèves-ingénieurs travaillent au niveau "système", sans entrer dans la circuiterie, comme peuvent le faire les techniciens et dépanneurs, ce qui est un handicap à un bon apprentissage pratique.

Mes perspectives pour faire évoluer l'enseignement des travaux pratiques à l'ENSERG concernent le renforcement de l'instrumentation programmée (LabView, LabWindows), et le développement de didacticiels (Toolbook) mis en place à la fois en salle de TP et sur le site web de l'école pour permettre aux étudiants de préparer ou réviser leurs TP hors de la salle de classe. J'ai encadré une stagiaire roumaine pour la mise en place de ce genre d'outils sur quelques unes des manipulations déjà existantes [Dum97].

J'ai également le projet de valoriser mon acquis et mon expérience en enseignement par la publication de deux ouvrages, qui sont en cours de rédaction :

- l'un portant sur les mathématiques de l'ingénieur, manuel pratique à l'usage des étudiants du second cycle, avec rappels de cours et exercices corrigés,
- l'autre consistant en un recueil de sujets de TP d'électronique à l'usage des enseignants (avec description du matériel requis, buts pédagogiques des manipulations, textes et corrigés). J'ai en effet constaté un manque évident dans ce domaine : chaque année, je suis contacté par des universitaires français ou étrangers, chargés de la mise en place ou de la rénovation de plates-formes de TP, et qui seraient très demandeurs de ce genre de document.

Enfin, il ne me déplairait pas que mon enseignement s'oriente un peu plus vers les aspects matériels (atelier et CAO électronique, processeurs numériques, VHDL) car c'est un thème d'intérêt dans mon activité de recherche sur les architectures de traitement d'images.

Chapitre 5

Autres Responsabilités

5.1 Responsabilité d'une Plate-Forme de TP

Depuis 1992, je suis responsable de la plate-forme TP2A à l'ENSERG :

- gestion du personnel : 1 technicien, 10 enseignants,
- gestion d'un budget annuel de 100 à 200kF,
- gestion du parc de matériel évalué à 2 MF (analyseurs de spectres numériques et analogiques, oscilloscopes combinés analogiques/numériques, générateurs wobulables synthétisés, multimètres et alimentations programmables, parc de PCs et imprimantes connectés en réseau)
- choix des investissements et négociations avec les fournisseurs,
- gestion administrative du service (calendriers, binômes, examens, rattrapages)
- choix pédagogiques, rédaction des notices d'appareils et textes de manipulations
- rénovation du parc de manipulations

5.2 Comité de Rédaction d'une Revue

Depuis 1995, je suis membre du comité de rédaction de la revue `Traitement du Signal`. Cette revue, éditée par le GRETSI, avec le concours du CNRS et de la DGA, paraît tous les trimestres (4 numéros, plus 2 numéros spéciaux). Le Directeur de publication est J.L. Lacoume. Le comité de rédaction, composé d'une douzaine de membres, essaie de se réunir régulièrement pour gérer les quelques 120 articles adressés à la revue chaque année. L'augmentation du nombre de papiers soumis portant sur le traitement d'image est la raison de ma participation au comité.

Je suis par ailleurs expert pour la revue `Signal Processing`.

5.3 Organisation d'Ecole d'Eté et Stages Pédagogiques

En 1997, j'ai été chargé, de l'organisation scientifique et matérielle de la seconde session de l'Ecole des Techniques Avancées en Signal-Image-Parole (ETASIP'97). Cette école est organisée alternativement par le LIS et l'ICP. Son objectif est de diffuser les résultats et les techniques les plus récents dans ces trois domaines, auprès de jeunes thésards, de chercheurs et d'enseignants-chercheurs ainsi que d'ingénieurs en R&D du secteur industriel. La méthode pédagogique fait appel à des exposés théoriques présentés par des spécialistes, illustrés par des ateliers sur des signaux et images réels, et par des présentations de prototypes et des démonstrations d'applications industrielles.

Mon rôle a consisté à concevoir et éditer la plaquette (Fig 5.1), assurer la diffusion de l'information (mailing listes), coordonner et produire les actes, gérer et équilibrer le budget de l'école (120kF dont la moitié en subventions), et participer, avec A. Chéhikian et P.Y. Coulon, à la définition du programme scientifique et à l'organisation matérielle (35 participants, 10 conférenciers, 10 ateliers).

ELESA
FR CNRS 08

CEPHAG - ICP - TIRF

Ecole des Techniques Avancées Signal Image Parole

Direction scientifique : Christian Jutten

Session 2
8-12 Septembre 1997

Traitement d'images et de séquences d'images ...

Méthodes, applications et mises en œuvre.

Responsables scientifiques
A. Chehikian, P.-Y. Coulon, F. Luthon

Lieu
Institut National Polytechnique de Grenoble
I.N.P.G., 46 Avenue Félix-Viallet
38031 Grenoble Cedex, France

Ecole des Techniques Avancées Signal Image Parole

Session 2 : 8-12 Septembre 1997

Traitement d'images et de séquences d'images ... Méthodes, applications et mises en œuvre.

L'objectif de cette seconde session de l'Ecole des Techniques Avancées en Signal, Image et Parole est d'une part de présenter les méthodes de base, mais aussi des approches plus récentes, utilisées pour traiter des images ou des séquences d'images, et d'autre part de passer en revue différents domaines d'application d'actualité, qui requièrent des algorithmes rapides et des mises en œuvre performantes pour manipuler les grandes quantités d'information que constituent les images et les séquences vidéo.

Une approche à la fois théorique et pratique sera adoptée pour présenter ces nouvelles techniques et leurs mises en œuvre. Dans ce but, l'école s'articulera :

- le matin, autour d'exposés théoriques sur les bases méthodologiques (filtrages, morphologie mathématique, multirésolution), et d'exposés sur des domaines d'application d'actualité (segmentation, compression, analyse de mouvement, télésurveillance, contrôle et diagnostic, interprétation).
- l'après-midi, autour d'ateliers de démonstrations et de simulations illustrant les exposés (qualité de compression, estimation de flux optique), et de présentations de réalisations matérielles (cartes de traitement rapide pour le filtrage, la détection de mouvement, circuits spécifiques).

Des pauses seront aussi prévues pour des discussions libres avec les intervenants, au cours desquelles les participants pourront exposer leurs centres d'intérêt et leurs préoccupations pratiques. En effet, le concept d'image est plus générique que la simple représentation d'une information selon deux dimensions spatiales, et peut s'étendre à tout type de signal à deux dimensions, quelle que soit la nature de ces dimensions. De nombreux outils mathématiques utilisés actuellement en image sont donc potentiellement applicables à d'autres problématiques où intervient la notion de signal 2D ou 3D (temps-fréquence, traitement d'antennes...).

Cette école d'été se veut tout à la fois un lieu d'initiation aux techniques actuelles du traitement des images statiques et dynamiques pour les non-spécialistes désireux d'acquérir une culture de base dans ce domaine en pleine évolution, un forum de discussion pour spécialistes confrontés à des problèmes spécifiques, et l'occasion de sensibiliser la communauté scientifique sur la portée et la généralité de certains outils pertinents en traitement des images et des séquences et généralisables à d'autres domaines.

Programme prévisionnel :

- **Thème 1** (Lundi, Mardi) : ***Fondements méthodologiques***
  Filtrage linéaire, filtrage d'ordre, filtrage non linéaire
  Représentations multirésolution, multiéchelle, pyramidale
  Morphologie mathématique
- **Thème 2** (Mercredi) : ***Segmentation d'images***
  Contours, Interpolation, Régions
  Coopération contours-régions
  Texture
- **Thème 3** (Jeudi) : ***Segmentation et estimation de mouvement***
  Approche markovienne, Flot optique
  Mises en œuvre matérielles « temps réel »
- **Thème 4** (Vendredi) : ***Compression et qualité d'images***
  Norme JPEG
  Décomposition sur des bases d'ondelettes, Fractales
  Homogénéité spatiale et temporelle, régularisation
  Couleur
- **Ateliers** (Après-midi) : ***Démonstrations de prototypes, simulations***

Intervenants

- **D. Barba,** SEI, IRESTE, Nantes
- **P. Bolon,** LAMII, Univ. Savoie, Annecy
- **P. Bouthemy,** IRISA, INRIA, Rennes
- **J.-M. Chassery,** TIMC, IMAG, Grenoble
- **A. Chehikian,** TIRF, INP Grenoble
- **P.-Y. Coulon,** TIRF, INP Grenoble
- **J.-M. Jolion,** INSA, Lyon
- **M. Jourlin,** CPE, Lyon
- **F. Luthon,** TIRF, INP Grenoble
- **S. Philipp,** ENSEA, Cergy-Pontoise

Ateliers et démonstrations

- **D. Alleyson,** TIRF, INP Grenoble
- **P. Bertolino,** TIMC, IMAG, Grenoble
- **A. Caplier,** TIRF, INP Grenoble
- **C. Dumontier,** TIRF, INP Grenoble
- **V. Fristot, TIRF,** INP Grenoble
- **A. Guérin-Dugué,** TIRF, INP Grenoble
- **S. Olympieff,** TIRF, INP Grenoble
- **D. Pellerin,** TIRF, INP Grenoble
- **G. Sicard,** TIRF, INP Grenoble
- **A. Spinei,** TIRF, INP Grenoble
- **A. Torralba,** TIRF, INP Grenoble

FIG. 5.1 - *Programme de l'ecole d'été ETASIP'97.*

L'organisation de cette école a non seulement été une expérience personnelle enrichissante, mais elle a aussi eu des répercussions positives pour le laboratoire : en plus de la publicité, de la valorisation et de la diffusion des recherches faites au laboratoire, elle a été l'occasion de prises de contacts avec plusieurs participants du monde industriel ou universitaire, contacts qui ont débouché sur des collaborations effectives ou des avancées dans la recherche du laboratoire (collaboration avec le CNS-CNET Meylan sur le proceseur PVP, utilisation du modéle LIP développé par l'équipe de M. Jourlin, développement de nos algorithmes de détection de mouvement sur processeur TMS320C80).

D'autre part, j'organise et j'anime tous les deux ans des Stages de Formation Pédagogique en Travaux Pratiques. Il s'agit de stages de 2 jours dans le cadre de la formation "Pratique Pédagogique en Electronique" dispensée par le C.I.E.S. (Centre d'Initiation à l'Enseignement Supérieur) aux doctorants-moniteurs de l'INPG et de l'Université Joseph Fourier de Grenoble : choix d'appareils de mesure et choix d'investissement pour plate-forme de TP, rédaction de notices techniques pédagogiques, prise en main d'instruments et de logiciels récents et performants.

5.4 Commissions Diverses

- Membre élu à la CCE : Commission Consultative d'Enseignement, chargée d'appréhender les problèmes de pédagogie à l'ENSERG.
- Membre élu au CDL : Conseil de Laboratoire du LIS.
- Membre CSE suppléant : Commission de Spécialistes 61éme section, chargée notamment du recrutement et de la promotion des personnels enseignants-chercheurs.

5.5 Participation à des Jurys

En plus de ma participation au jury des deux thèses que j'ai dirigées, j'ai également été sollicité pour être membre du jury de la thèse de Yann Ricquebourg, intitulée :

"Analyse de mouvements articulés dans des séquences d'images : Mesure et suivi 2D ; Application à la télé-surveillance".

Doctorat de l'Université de Rennes I, mention Informatique, soutenue le 10 janvier 1997.

Composition du jury :

B. Arnaldi	Professeur, IRISA-Rennes I, Président
M.J. Aldon	Chargée de Recherche, LIRMM Université de Montpellier, Rapporteur
D. Barba	Professeur, IRESTE Nantes, Rapporteur
P. Bouthémy	Directeur de Recherche, IRISA Rennes I
F. Heitz	Professeur, ENSPS Illkirch
F. Luthon	MCF, INPG Grenoble
M. Parent	Directeur de Recherche, INRIA Rocquencourt

En outre, je prends part chaque année aux jurys de DEA de l'Ecole Doctorale de l'INPG (Option Signal-Image-Parole).

Partie II

Le Mouvement dans les Images

Présentation

L'analyse du mouvement [1] dans les séquences d'images est un domaine de recherche active à l'heure actuelle, en raison de son importance dans de nombreuses applications : télésurveillance, compression pour les télécommunications ou l'archivage, diagnostic médical, météorologie, contôle non destructif, robotique mobile, etc.

A la base de quasiment toutes les méthodes d'analyse de mouvement, qu'elles soient différentielles, fréquentielles, ou par mise en correspondance, on trouve toujours une information élémentaire liée à la variation temporelle d'intensité lumineuse entre deux instants, que ce soit une simple différence de trames (*frame difference* FD), une différence de trame déplacée (DFD) [NR79, WR84], un gradient temporel, ou un écart entre blocs dans deux images successives [ADM81]. Mais cette information élémentaire et bas niveau requiert des traitements souvent complexes pour aboutir à une décision ou une interprétation pertinente, d'où des coûts de calcul élevés qui nécessitent de s'intéresser aux mises en oeuvre temps réel, si l'on vise une application opérationnelle de ces traitements.

On distingue principalement quatre tâches [MB96]: la *détection* (des changements temporels ou des zones mobiles dans le plan image), l'*estimation* (des vecteurs-vitesses, en chaque pixel ou pour chaque objet), la *segmentation* (en régions cohérentes au sens du mouvement selon un certain critère) et l'*interprétation* de haut niveau (reconnaissance de formes faisant appel à l'intelligence artificielle). Ces quatre étapes ne sont en aucun cas indépendantes ni forcément séquentielles, mais au contraire fortement interdépendantes (notamment en ce qui concerne les deux phases estimation-segmentation [CTS97]). Nos travaux ont jusqu'à présent porté principalement sur la phase de détection du mouvement des objets mobiles dans le cas d'une caméra fixe par rapport à la scène observée, et sur la segmentation de visages de locuteurs, basée sur la couleur et le mouvement. Nous menons également des travaux en estimation de mouvement associée à la transformée en ondelettes.

Cette partie expose plus en détail 6 travaux de recherche, sélectionnés pour leur représentativité à la fois de mon activité et de la palette des problématiques propres au mouvement. Elle tente de situer nos propres travaux par rapport à l'état de l'art.

Elle est organisée en deux chapitres, chacun divisé en trois thèmes : l'un porte sur les traitements et les algorithmes (détection de mouvement, segmentation de lèvres, ondelettes et estimation de mouvement). L'autre porte sur les implantations matérielles (machine SIMD, ASIC

1. "La représentation du changement par le mouvement est bien un préjugé. Oui. Même dans le cas le plus simple de cette bille qui roule, rien dans les apparences n'impose l'hypothèse du mouvement ; et ce n'est toujours pas la bille qui le présente, puisqu'elle n'est jamais en même temps que dans un lieu, comme le subtil Zénon l'avait remarqué. Et rien n'empêche de supposer que la bille est détruite aussitôt, et qu'une autre bille naît à côté, comme il est vrai dans le cinématographe, où ce n'est point le même cheval qui court, mais des images différentes qui se remplacent sur l'écran. Seulement le mouvement est préféré et choisi ; et le mécanisme est de même préféré et choisi ; non comme facile, certes, car rêver est le plus facile, mais comme libérateur, exorciseur, arme de l'esprit contre tous sortilèges. Soutenu par la nature qui le vérifie, oui, mais qui le vérifie à la condition que l'esprit le pose, le maintienne, le construise, le complique assez." ([Ala41], p. 153).

VLSI analogique, DSP, processeur vidéo DVP). Ces 6 thèmes correspondent à mes encadrements principaux (ayant débouché sur des résultats concrets ou des publications) et donc aux travaux des étudiants qui ont le plus marqué ma carrière à ce jour : A. Caplier, C. Dumontier, D. Dragomirescu, G. Laborde, M. Liévin, S. Popescu et V.G. Popescu.

Après quelques rappels sur les principes et outils de base utilisés, nous présentons nos principaux axes d'innovation. Nos algorithmes se veulent simples mais robustes et implantables en temps-réel, influencés à l'origine par les travaux d'Alain Chéhikian et sa manière d'envisager les problèmes de traitement d'images [CC91, Che92].

Chapitre 1

Traitements et Algorithmes

1.1 Détection de Mouvement

1.1.1 Principes

La détection de mouvement consiste à attribuer à chaque pixel ou *site* $s = (x, y, t)$ des images d'une séquence un attribut, qu'on appelle *étiquette* ou label l_s, indiquant si le pixel appartient à un objet mobile ou au fond fixe de la scène observée :

$$l_s = \begin{cases} m = \text{"1"} & \text{si } s \in \text{zone mobile}, \\ b = \text{"0"} & \text{si } s \in \text{fond fixe}. \end{cases} \tag{1.1}$$

En faisant l'hypothèse d'une caméra fixe et d'un éclairement quasi-constant de la scène observée, on peut extraire en chaque pixel une information de bas niveau, qu'on appelle *observation* o_s, portant sur la variation temporelle de l'intensité lumineuse I du pixel :

$$o_s = |I(x, y, t) - I(x, y, t-1)| \tag{1.2}$$

Dans le cas idéal, cette variation temporelle entre deux instants d'acquisition $t-1$ et t est nulle pour un pixel du fond fixe, non nulle s'il y a eu mouvement d'un objet d'intensité non uniforme. Cependant, cette observation de bas niveau est peu robuste (bruit d'acquisition, légères variations d'éclairement ...) et ne donne qu'une information partielle sur les objets en mouvement. Un simple seuillage ne permet pas d'extraire les zones mobiles. En effet, on distingue typiquement 4 zones dans le champ d'observations (Fig. 1.1) :

- le fond fixe,
- la zone d'écho (zone de fond découverte par l'objet à l'instant t),
- la zone de glissement de l'objet sur lui-même,
- la zone de recouvrement (zone de fond recouverte par l'objet à l'instant t).

Il faut d'une part éliminer l'écho, d'autre part reconstruire l'intérieur des objets mobiles (zone de glissement). Différentes techniques, très simples dans leur principe, mais relativement peu robustes, permettent d'obtenir les zones mobiles :

- la technique de simple différence par rapport à une image de référence censée ne contenir que le fond fixe de la scène : le problème est bien sûr l'obtention de cette image de référence, qu'il faut remettre à jour régulièrement en fonction de l'évolution de la scène,

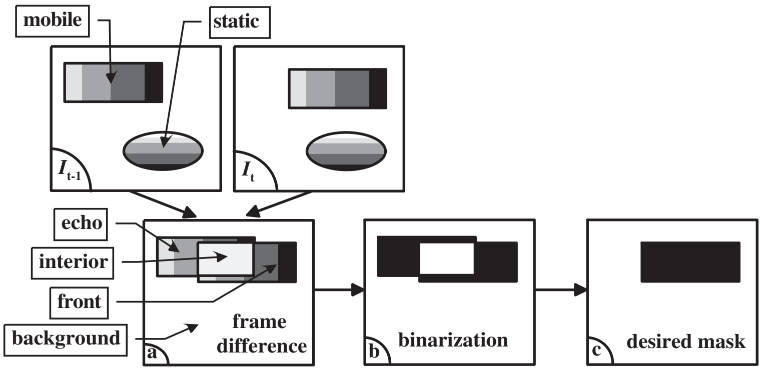


FIG. 1.1 - *Processus de détection de mouvement. Cas d'un rectangle mobile et d'une ellipse fixe. a) observation : différence temporelle entre les instants t et $t-1$ b) carte binaire des changements temporels ; c) masque idéal en sortie.*

- la technique du ET logique entre cartes binaires de changements temporels, qui ne donne de bons résultats que si le mouvement est assez rapide pour qu'il n'y ait pas de recouvrement entre deux positions successives de l'objet, ce qui est assez restrictif.

1.1.2 Régularisation par Approche Markovienne

C'est pourquoi on fait appel à une approche statistique probabiliste pour régulariser le problème qui est initialement mal posé (sous-déterminé) [PTK85]. Le principe consiste à introduire une modélisation *a priori* du champ des étiquettes, à l'aide de la théorie des champs aléatoires de Markov (*Markov Random Fiels* ou MRF) [CJ93]. On contraint alors la solution vers la configuration la plus probable du champ des étiquettes étant donné les observations et le modèle *a priori* [BL93].

Fonctions d'Energie

Adoptons les notations suivantes :

S	l'image à l'instant courant t
s	un pixel (*site*) quelconque de S
η_s	un voisinage spatio-temporel de s, par exemple celui de la Fig. 1.2
n	n'importe quel voisin de s (spatial ou temporel)
C	l'ensemble des cliques[1] binaires $c = (s, n)$ constituant les voisinages η_s
$O = \{O_s, s \in S\}$	le champ aléatoire des observations
$L = \{L_s, s \in S\}$	le champ aléatoire des étiquettes
$o = \{o_s, s \in S\}$	une réalisation particulière du champ d'observations O à l'instant t
$l = \{l_s, s \in S\}$	une réalisation particulière du champ d'étiquettes L à l'instant t
Λ	l'ensemble des configurations possibles l du champ aléatoire L.

Si nécessaire, on ajoutera un indice temporel pour préciser l'instant considéré : t, $t-1$, etc.

Les interactions spatiales et temporelles entre étiquettes sont modélisées par un MRF, qui constitue le modèle *a priori* du champ des étiquettes. La propriété essentielle d'un MRF relativement à une structure de voisinage est le caractère local des interactions : la probabilité d'avoir en un pixel s l'étiquette l_s ne dépend que des étiquettes de ses voisins, et non pas de toute l'image.

$$\forall l \in \Lambda, \qquad P(L = l) > 0 \qquad (1.3)$$

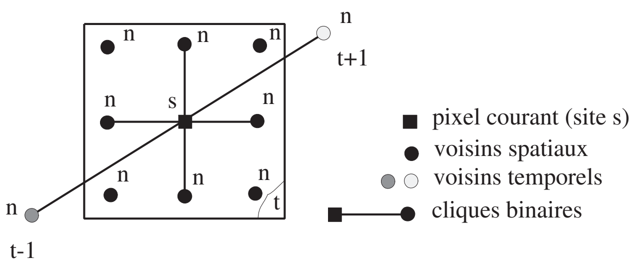


FIG. 1.2 - *Voisinage spatio-temporel de base et cliques binaires associées.*

$$P\left(L_s = l_s \mid L_n = l_n, n \neq s, n \in S\right) \;\; = P\left(L_s = l_s \mid L_n = l_n, n \in \eta_s\right) \tag{1.4}$$

Cette propriété est cruciale car elle implique des calculs purement locaux et fortement parallélisables, d'où les possibilités de mise en œuvre temps réel [CLD98, DLC98b].

Un MRF étant équivalent à une distribution de Gibbs, on peut exprimer la probabilité *a priori* du champ d'étiquettes à l'aide d'une fonction d'énergie [MB87]:

$$P(L = l) = \frac{1}{Z}\, exp(-U_m(l)) \tag{1.5}$$

où Z est une constante de normalisation (appelée *fonction de partition*) et U_m une fonction d'énergie associée au modèle *a priori*. Le champ d'étiquettes le plus probable relativement aux observations est obtenu par le critère du Maximum A Posteriori (MAP). A l'aide du théorème de Bayes et de l'équation (1.5), on montre que ce critère équivaut à la minimisation d'une fonction d'énergie totale U [GG84] :

$$\max_l P(L = l \,|\, O = o) \Longleftrightarrow \min_l U \quad \text{ où } \quad U = U_m(l) + U_a(o, l) \tag{1.6}$$

$U_m(l)$ est un terme d'énergie qui assure la régularisation de la solution (terme analogue à une contrainte de lissage par exemple). Il s'exprime comme une somme de potentiels énergétiques élémentaires sur les cliques :

$$U_m(l) = \sum_{c \in C} V_c(l_s, l_n) \tag{1.7}$$

Ces potentiels élémentaires sont définis en fonction du problème à traiter. On peut par exemple prendre des potentiels à niveaux du type :

$$V_c(l_s, l_n) = \begin{cases} -\beta & \text{si } l_s = l_n \\ +\beta & \text{si } l_s \neq l_n \end{cases} \tag{1.8}$$

ou des potentiels quadratiques du type :

$$V_c(l_s, l_n) = \beta(l_s - l_n)^2 \tag{1.9}$$

où $\beta > 0$ prend une des trois valeurs β_s, β_p ou β_f selon la clique considérée (spatiale, passée ou future). Ces potentiels énergétiques favorisent un étiquetage homogène puisque des étiquettes semblables sur des pixels voisins induisent une moindre contribution énergétique, donc une configuration plus favorable. En pratique, on prend $\beta_f > \beta_p$, pour bien éliminer les zones d'écho du mouvement. On peut aussi utiliser des estimateurs robustes [OB94, Odo94, OB95], dont certains sont implantables en analogique (*antibump function* [Del93]).

$U_a(o,l)$ est l'énergie d'adéquation, qui assure une bonne attache aux données. Elle traduit le lien entre observations et étiquettes et s'exprime à l'aide d'une fonction Ψ censée modéliser les observations :

$$U_a(o,l) = \frac{1}{2\sigma^2} \sum_{s \in S} [o_s - \Psi(l_s)]^2 \tag{1.10}$$

où σ^2 est la variance des observations et Ψ est la fonction d'attache aux observations. Pour la détection de mouvement, on peut par exemple prendre une fonction simple du type :

$$\Psi(l_s) = \begin{cases} 0 & \text{si } l_s = b \\ \alpha > 0 & \text{sinon} \end{cases} \tag{1.11}$$

où α correspond à la valeur moyenne des observations non nulles. Une autre définition pour Ψ est proposée dans [BL93] :

$$\Psi(l_s) = \begin{cases} 0 & \text{if } \; l_s = l_p = b & \text{(fond fixe)} \\ \alpha_1 > 0 & \text{if } \; l_s = l_p = m & \text{(mouvement)} \\ \alpha_2 >> \alpha_1 & \text{if } \; l_s \neq l_p & \text{(transition)} \end{cases} \tag{1.12}$$

où α_1 et α_2 sont des paramètres qui peuvent être prédéfinis ou estimés en ligne, et où l_p représente l'étiquette attribuée au voisin passé.

Tous les paramètres qui interviennent dans la modélisation (α_i, β_i, σ) peuvent soit être déterminés de façon empirique après des tests sur des séquences typiques correspondant à l'application envisagée, soit être estimés automatiquement grâce à des algorithmes du type EM (*Expectation-Maximisation*) ou SEM (*Stochastic Expectation Maximisation*) [DLR77, BP95, Pie94].

Algorithmes de Relaxation Spatiale

Pour calculer la configuration d'étiquettes donnant l'énergie minimale, différents algorithmes, dits de relaxation, peuvent être utilisés.

- Les algorithmes de relaxation stochastiques, du type recuit simulé [KGV83], consistent à explorer l'espace Λ des configurations possibles avec une loi de descente en température adéquate (*i.e.* suffisamment lente) qui assure en principe de converger vers le minimum global de la fonction d'énergie, mais au prix d'un coût de calcul prohibitif.

- C'est pourquoi on leur préfère souvent les algorithmes de relaxation déterministes, du type ICM (*Iterated Conditional Modes*) [Bes86], GNC (*Graduated Non Convexity*) ou MFA (*Mean Field Annealing*) [BSGG92, ZC93], moins lourds en calculs mais qui ne garantissent pas de converger vers le mimimum global d'énergie. Ils risquent de rester piégés dans un minimum local, car ils n'autorisent aucune remontée en température pour sortir d'un minimum local. C'est pourquoi ce type d'algorithme requiert une bonne configuration initiale du champ d'étiquettes (Fig. 1.3).

Ces algorithmes déterministes sont récursifs et itératifs : on visite chaque pixel de l'image et on calcule, pour chaque étiquette qu'on peut lui attribuer, la contribution énergétique sur son voisinage. On retient l'étiquette qui donne l'énergie minimale, puis on passe au pixel suivant et ainsi de suite sur toute l'image. On réitère la visite des sites de l'image jusqu'à la convergence

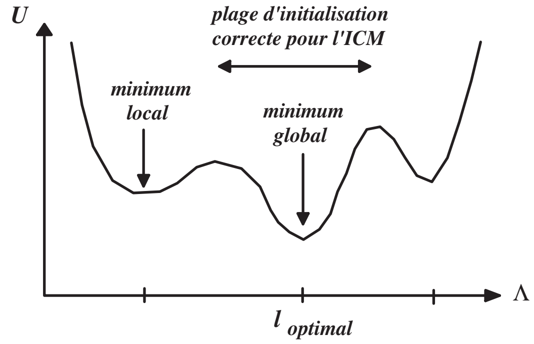


FIG. 1.3 - *Influence de l'initialisation sur le résultat de l'ICM : évolution de la fonction d'énergie U en fonction de la configuration l du champ d'étiquettes.*

de la fonction d'énergie totale U vers une valeur qui n'évolue plus, ou très peu, au cours des itérations. Différentes politiques de visite de sites sont envisageables :

- visite séquentielle classique (*line scanning*) (ligne par ligne de gauche à droite et image par image de haut en bas),
- visite aléatoire (pile d'instabilité [CB90]),
- visite chaînée (ligne par ligne alternativement de gauche à droite puis de droite à gauche et image par image alternativement de haut en bas puis de bas en haut) pour bénéficier au mieux des derniers voisinages mis à jour.

Différentes politiques de mise à jour des sites sont aussi envisageables, le choix dépendant notamment du type de matériel utilisé pour la mise en œuvre (cf. §2) :

- pixel-récursif : on met à jour chaque pixel au fur et à mesure de la visite,
- ligne-récursif : on attend d'avoir visité tous les pixels d'une ligne avant de mettre à jour leurs étiquettes,
- image-récursif : on attend d'avoir visité toute une image avant de faire une mise à jour simultanée de toutes les étiquettes de l'image.

De même, différents critères de convergence (arrêt des itérations) peuvent être utilisés :

- nombre de pixels instables inférieur à un seuil prédéterminé,
- variation relative d'énergie totale inférieure à un seuil prédéterminé (typiquement $\Delta U \ / \ U = 0.01\%$),
- nombre fixe d'itérations (en pratique un faible nombre d'itérations par image, typiquement 5 à 10, est suffisant). Ce critère est évidemment le moins coûteux pour une mise en œuvre matérielle.

Organigramme de Détection Robuste

Le synoptique de l'algorithme est donné Fig. 1.4. Sur ce schéma-bloc, la notation l^0 représente un champ d'étiquettes initial grossièrement estimé par binarisation du champ correspondant d'observations. Pour réaliser la binarisation, on peut soit faire un seuillage simple (technique peu

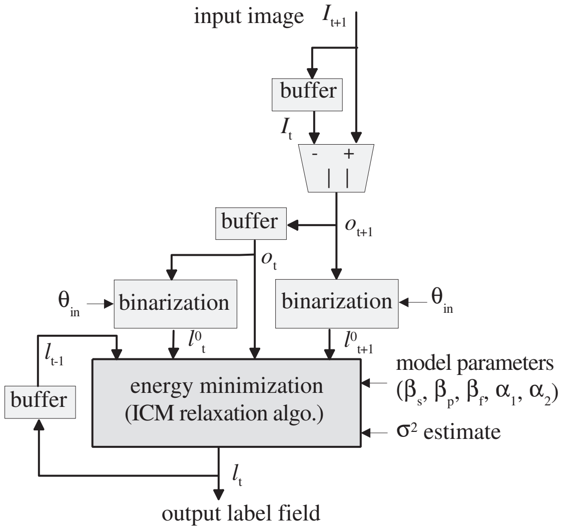


FIG. 1.4 - *Synoptique de l'algorithme de détection de mouvement.*

coûteuse mais peu robuste aux fluctuations), soit développer des techniques de détermination automatique des seuils, basées par exemple sur l'entropie de la différence d'images pour évaluer automatiquement le niveau de bruit [LL99], soit utiliser des techniques robustes basées sur des tests de vraisemblance, avec modèles de luminance (constant, linéaire ou quadratique) sur un voisinage des pixels [HNR84, AKM93].

Etant donné le voisinage spatio-temporel utilisé pour détecter le mouvement, voisinage qui inclut un voisin futur et un voisin passé pour chaque pixel (Fig. 1.2), le résultat de traitement est obtenu avec un retard d'une image par rapport à l'acquisition, puisqu'il faut disposer de l'image en $t+1$ pour estimer le champ d'étiquettes en t.

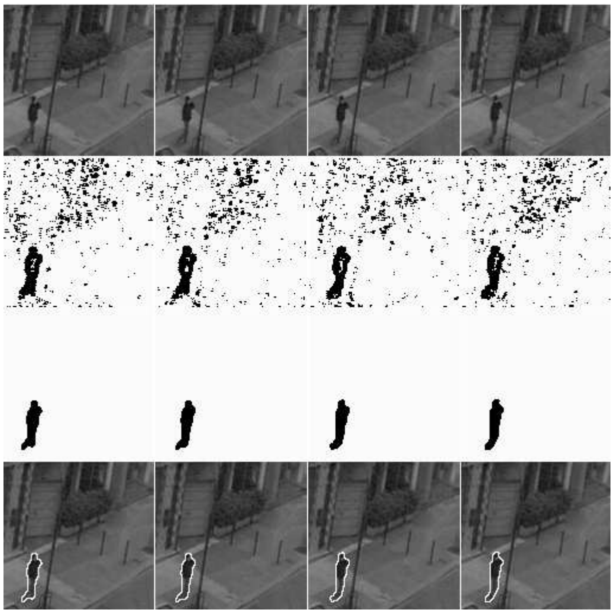
La Fig. 1.5 donne un exemple de résultat typique sur une scène de rue.

Postionnement du Travail

Par rapport à l'algorithme original proposé par Bouthémy [BL93], notre contribution porte sur plusieurs modifications, conduisant à une version de l'algorithme mieux adaptée aux mises en œuvre matérielles (§2), tout en offrant les mêmes performances de robustesse et de qualité que l'algorithme de détection original.

La première différence concerne la façon de traiter l'information temporelle : dans [BL93], une détection initiale est réalisée à partir du couple d'images I_t et I_{t-1}, puis elle est mise à jour quand on prend en compte le couple d'images I_{t+1} et I_t. Deux champs d'étiquettes successifs sont mis à jour simultanément (optimisation en deux passes qui permet de différer la décision concernant les zones découvertes par le mouvement). L'algorithme modifié proposé ici travaille sur un seul champ d'étiquettes optimisé une seule fois (traitement en une seule passe). Ceci est rendu possible grâce à une initialisation différente : on utilise une estimation grossière du champ d'étiquettes futur, au lieu de simplement répéter le passé. Les zones découvertes par le mouvement sont gérées en donnant plus de poids au futur (anisotropie temporelle introduite grâce à $\beta_f > \beta_p$). Ainsi, le nombre de champs requis pour la relaxation est de 5 dans notre algorithme, au lieu de 6 dans [BL93], d'où un besoin moindre pour les stockages en mémoire.

La seconde différence concerne le coût de calcul : on utilise quatre paramètres de modélisation au lieu de cinq car la fonction d'adéquation est simplifiée (Eq. (1.11) au lieu de Eq. (1.12)). La



Nb_itération	Temps cpu /image
4	5s

FIG. 1.5 - *Résultat typique de détection de mouvement. De haut en bas : 1) séquence de rue avec un piéton mobile ; 2) champs d'étiquettes initiaux ; 3) masques finaux après relaxation par ICM ; 4) superposition des contours des masques sur les images de la séquence. Le temps de calcul indiqué a été obtenu sur station Sun Sparc-2.*

décision concernant la clique temporelle (passé ou futur) nécessite un seul test conditionnel pour choisir entre deux configurations, alors que 8 configurations différentes sont testées dans la version originale de Bouthémy (Table 1 dans [BL93]).

Notons enfin que dans les voisins temporels ne sont pas strictement markoviens, puisqu'ils ne sont pas affectés par la relaxation. La technique de relaxation spatio-temporelle sur une section temporelle de la séquence, proposée au §1.2.2, est une réponse à ce constat.

1.2 Nos Variantes

1.2.1 Multi-Etiquetage

On a étendu la détection binaire au cas de la détection multi-étiquette. Le but est d'attribuer une étiquette distincte à chaque zone mobile (étiquette m_i pour le iéme objet mobile). Pour la mise à jour automatique du nombre d'étiquettes (apparition d'objets dans la scène), on introduit une étiquette supplémentaire m_0 de mouvement indéterminé, attribuée temporairement aux pixels incertains. On a défini une initialisation multi-étiquette et une relaxation multi-voisinage pour gérer correctement ce modèle multi-étiquette [Cap95]. L'adjonction d'un filtre de Kalman à ce modèle multi-étiquette permet en outre une prédiction et un suivi temporel des masques en mouvement dans les cas d'occultation [Cap95]. Le filtre de Kalman fournit un voisin virtuel prédit, que l'on introduit dans le modèle markovien.

Cette variante permet de traiter le cas délicat des occultations temporaires d'un mobile par un objet fixe dans la scène. Un exemple de résultat de cette coopération entre MRF multi-étiquette et filtrage de Kalman est présenté sur la Fig. 1.6.

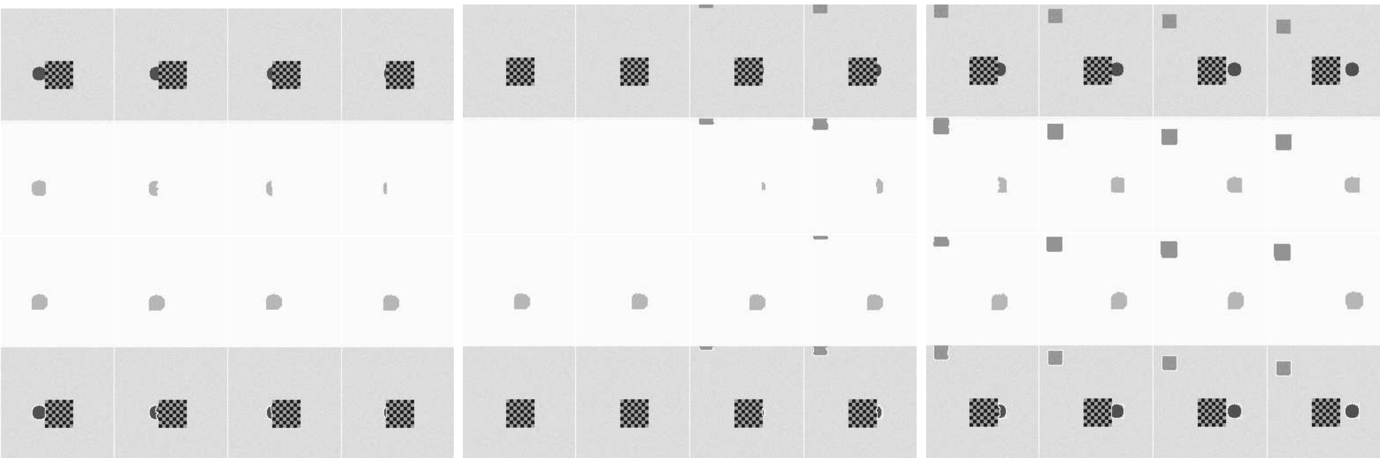


FIG. 1.6 - *Détection de mouvement multi-étiquette et suivi temporel d'un mobile occulté par le fond fixe. De haut en bas : 12 images non consécutives d'une scène synthétique (cercle mobile occulté par un damier fixe et apparition d'un carré) ; masques obtenus après relaxation ; champs des prédictions obtenus par filtrage de Kalman ; superposition des contours des masques (en blanc) sur la séquence initiale.*

1.2.2 Voisinage Cubique et Relaxation Spatio-Temporelle

Un autre type de variante sur le thème de la modélisation markovienne porte sur le choix de la structure de voisinage. On peut envisager une structure de voisinage 3-D comme présenté sur la

Fig. 1.7, associée à une relaxation spatio-temporelle sur une section temporelle de longueur N_t de la séquence vidéo, au lieu d'une relaxation uniquement spatiale image par image [CL97, LCL99].

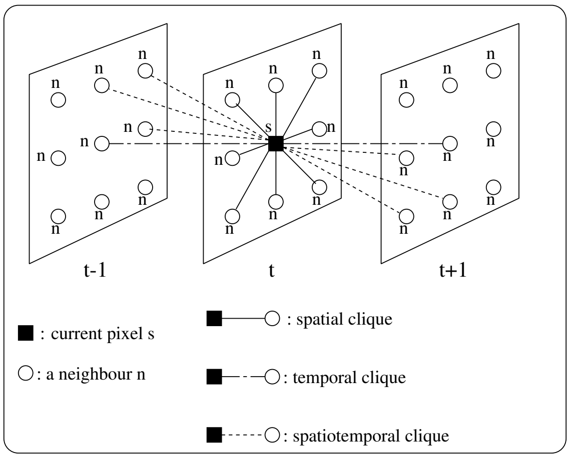


FIG. 1.7 - *Voisinage spatio-temporel cubique, et cliques binaires associées (toutes les cliques ne sont pas représentées).*

Si on note $\delta_x, \delta_y, \delta_t$ les coordonnées du vecteur représentatif de chaque clique dans l'espace (x, y, t) centré en le pixel courant s, on définit :

- 4 cliques spatiales horizontales et verticales (δ_x ou $\delta_y = \pm 1$, $\delta_t = 0$);
- 4 cliques spatiales diagonales (δ_x et $\delta_y = \pm 1$, $\delta_t = 0$);
- 2 cliques temporelles (δ_x et $\delta_y = 0$, $\delta_t = \pm 1$);
- 8 cliques spatio-temporelles horizontales et verticales (δ_x ou $\delta_y = \pm 1$, $\delta_t = \pm 1$);
- 8 cliques spatio-temporelles diagonales (δ_x et $\delta_y = \pm 1$, $\delta_t = \pm 1$).

Il faut alors modéliser les énergies d'interactions spatio-temporelles sur toutes les cliques $c = (s, n)$ de ce voisinage cubique par des potentiels $\beta(s, n)$ qui sont fonctions du type de clique considérée. Dans [LCL99], on propose une expression qui fait intervenir l'inverse du carré de la distance séparant les deux pixels de la clique :

$$\beta(s,n) = \frac{1}{d^2(s,n)\ \left(\frac{\delta_x(s,n)^2}{\beta_s} + \frac{\delta_y(s,n)^2}{\beta_s} + \frac{\delta_t(s,n)^2}{\beta_t}\right)} \tag{1.13}$$

où $d(s, n) = \sqrt{\delta_x^2 + \delta_y^2 + \delta_t^2}$ est la distance euclidienne entre le pixel courant s et le voisin considéré n. Cette formule donne :

- $\beta(s, n) = \beta_s$ pour les cliques spatiales horizontales ou verticales ($d(s, n) = 1$) ;
- $\beta(s, n) = \frac{\beta_s}{4}$ pour les cliques spatiales diagonales ($d(s, n) = \sqrt{2}$) ;
- $\beta(s, n) = \beta_t$ pour les cliques temporelles ($d(s, n) = 1$) ;
- $\beta(s, n) = \frac{\beta_s \beta_t}{2(\beta_s + \beta_t)}$ pour les cliques spatio-temporelles horizontales ou verticales ($d(s, n) = \sqrt{2}$) ;

– $\beta(s,n) = \frac{\beta_s \beta_t}{3(\beta_s+2\beta_t)}$ pour les cliques spatio-temporelles diagonales ($d(s,n) = \sqrt{3}$).

Notons que cette définition de $\beta(s,n)$ ne fait intervenir que deux paramètres β_s et β_t.

Cette variante de l'algorithme permet d'améliorer les performances de la détection dans le cas de séquences bruitées comme illustré sur la Fig. 1.8.

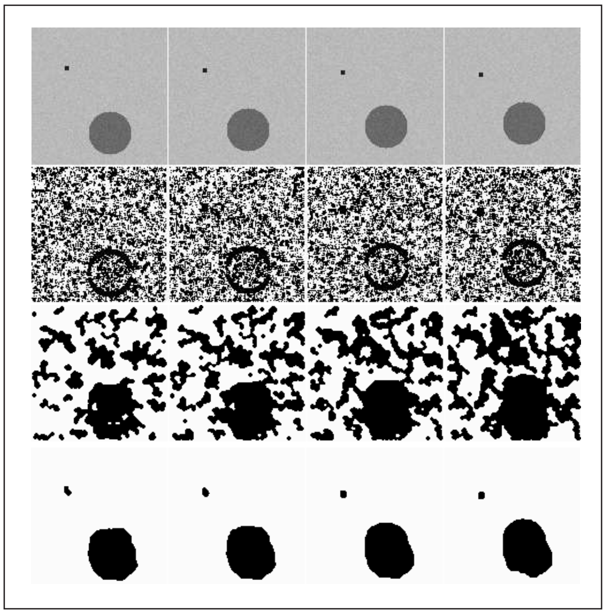


FIG. 1.8 - *Comparaison des deux variantes de l'algorithme de détection. De haut en bas : 1) séquence synthétique bruitée ; 2) initialisation binaire ; 3) masques obtenus avec voisinage de base et relaxation spatiale ; 4) masques obtenus avec voisinage cubique et relaxation spatio-temporelle.*

Une seconde formulation, plus rigoureuse, du potentiel d'interaction $\beta(s,n)$ peut être obtenue par analogie avec l'électrostatique ou la mécanique (notions de potentiel électrique ou d'énergie potentielle). On sait qu'un potentiel entre deux points suit une loi d'évolution en $1/d$, où d est la distance entre les points.

Si l'on note d_x, d_y, d_t les coordonnées du vecteur représentatif de chaque clique dans l'espace des potentiels $(\beta_x, \beta_y, \beta_t)$, on peut adopter une approche tensorielle en définissant la relation entre les coordonnées "géométriques" et les coordonnées "potentielles" :

$$\begin{bmatrix} \delta_x \\ \delta_y \\ \delta_t \end{bmatrix} = T. \begin{bmatrix} d_x \\ d_y \\ d_t \end{bmatrix} \tag{1.14}$$

où T est un tenseur (matrice 3×3), que l'on définit à son gré pour une modélisation adéquate des interactions.

Le potentiel s'exprime alors simplement comme l'inverse de la distance virtuelle ou énergétique :

$$\beta(s,n) = \frac{1}{||d||} = \frac{1}{\sqrt{d_x^2 + d_y^2 + d_t^2}} \tag{1.15}$$

La transformation T la plus simple consiste à prendre l'expression (1.16), pour se retrouver dans un cas analogue au précédent :

$$T = \begin{bmatrix} \beta_s & 0 & 0 \\ 0 & \beta_s & 0 \\ 0 & 0 & \beta_t \end{bmatrix}. \tag{1.16}$$

On obtient alors :

- $\beta(s,n) = \beta_s$ pour les cliques spatiales horizontales ou verticales ;
- $\beta(s,n) = \frac{\beta_s}{\sqrt{2}}$ pour les cliques spatiales diagonales ;
- $\beta(s,n) = \beta_t$ pour les cliques temporelles ;
- $\beta(s,n) = \frac{\beta_s\beta_t}{\sqrt{\beta_s^2+\beta_t^2}}$ pour les cliques spatio-temporelles horizontales ou verticales ;
- $\beta(s,n) = \frac{\beta_s\beta_t}{\sqrt{\beta_s^2+2\beta_t^2}} = \frac{\frac{\beta_s}{\sqrt{2}}\beta_t}{\sqrt{\left(\frac{\beta_s}{\sqrt{2}}\right)^2+\beta_t^2}}$ pour les cliques spatio-temporelles diagonales.

Cette modélisation est utilisée avec succès pour la segmentation des lèvres dans un visage, car elle permet, par le choix judicieux d'une matrice T, d'introduire de façon cohérente des contraintes géométriques (anisotropie verticale/horizontale par exemple).

1.2.3 Multirésolution Spatio-Temporelle

Pour améliorer les performances de l'algorithme dans le cas du mouvement d'objets très lents (mouvement sub-pixel), ou de gros objets peu texturés, on peut utiliser une technique du type multirésolution [Ros84]. On présente ici quelques résultats obtenus avec une technique originale de multirésolution spatio-temporelle, qui consiste à filtrer passe-bas et à sous-échantillonner la séquence d'images à la fois en espace et en temps selon le schéma de la Fig. 1.9, où est aussi représenté le noyau du filtre 3D appliqué à la séquence. Ce noyau de filtrage 3-D est une simple extension 3-D du filtre 2-D passe-bas proposé dans [Che92] pour la génération de pyramides d'images [Bur84].

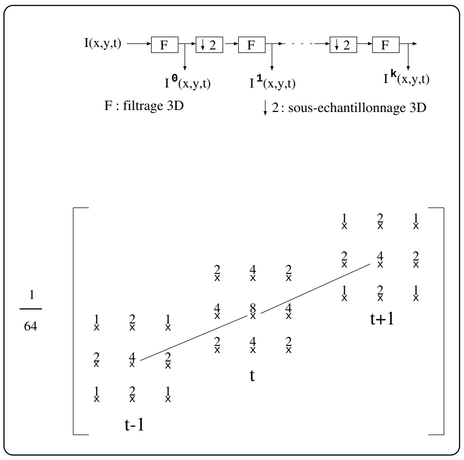


FIG. 1.9 - *Principe de construction de la pyramide spatio-temporelle et noyau du filtre 3-D.*

Un exemple de pyramide spatio-temporelle est donné Fig. 3.4, p. 15.

Le filtrage spatial permet de renforcer les observations dans le cas de grandes zones uniformes d'où une meilleure détection (Fig. 1.10 gauche). Quant au filtrage temporel, il permet d'intégrer l'information de mouvement de plusieurs images successives, ce qui est indispensable pour détecter des objets au mouvement très lent (Fig. 1.10 droite).

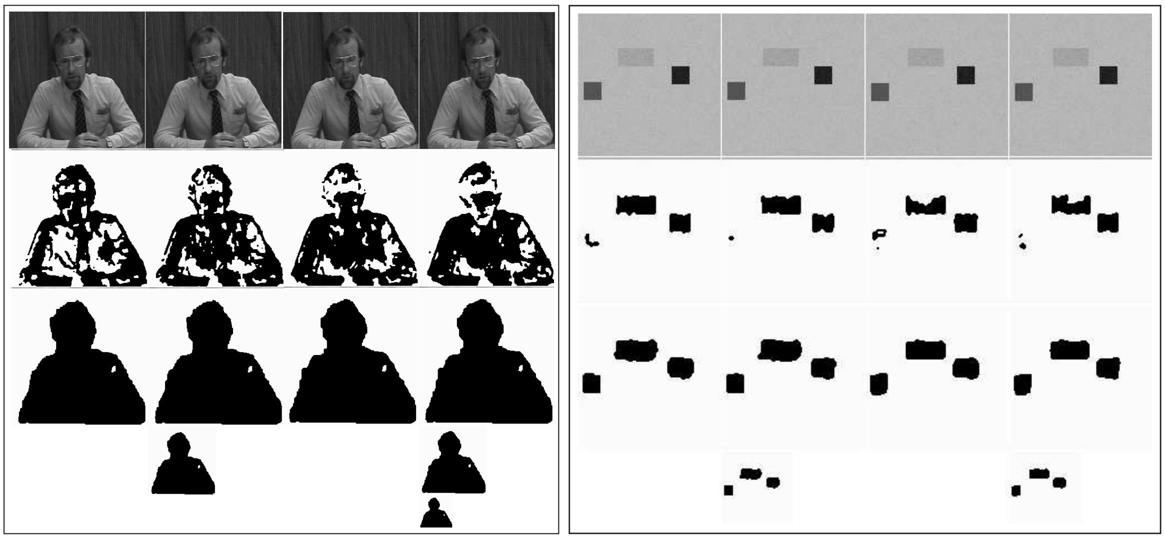


FIG. 1.10 - *Détection de mouvement multirésolution. A Gauche) Cas de zones peu texturées : De haut en bas : 1) séquence Trevor; 2) masques monorésolution ; 3) masques multirésolution (3 niveaux d'analyse* $k = 0, 1, 2$*). A Droite) Cas d'un mouvement très lent : De haut en bas : 1) séquence synthétique avec 3 objets mobiles dont l'un a un mouvement sub-pixel; 2) masques monorésolution ; 3) masques multirésolution (2 niveaux d'analyse* $k = 0, 1$*).*

1.2.4 Discussion

Le schéma de multirésolution spatio-temporelle proposé ici forme un tout cohérent avec le modèle MRF 3-D à voisinage cubique et la relaxation spatio-temporelle associée. Il permet de traiter deux cas difficiles : celui des mouvements sous-pixel et celui des zones mobiles très peu texturées. Cependant, dans le cas de mouvements rapides, la version multirésolution spatio-temporelle donne de moins bons résultats que la version monorésolution, car le filtrage temporel induit un lissage de l'information de mouvement sur plusieurs images, qui empêche de détecter avec précision les frontières du mouvement. En conséquence, les masques sont plus gros que les zones mobiles effectives. Une multirésolution uniquement spatiale est alors mieux adaptée, car elle permet de linéariser spatialement l'intensité, sans introduire de flou temporel. Les versions mono- et multirésolution étant complémentaires, il serait intéressant de développer une stratégie permettant une commutation automatique entre les deux versions, en fonction d'informations *a priori* dont on disposerait (taille, texture et vitesse des objets). De plus, la pyramide spatio-temporelle implique un filtrage passe-bas 3-D. Pour limiter l'effet de lissage (flou), l'utilisation d'ondelettes 3-D (bancs de filtres 3-D orthogonaux passe-haut et passe-bas) pourrait être une alternative intéressante.

On a passé en revue un certain nombre de techniques applicables en détection de mouvement dans les séquences d'images. Mais les outils de traitement présentés ici sont de portée très générale et peuvent être utilisés dans d'autres domaines d'applications ou sur d'autres types de données que des séquences d'images. Ils sont en effet applicables dans tout problème mal posé portant sur des signaux à deux ou trois dimensions, quelle que soit la nature de ces dimensions, et où il est nécessaire de régulariser la solution en introduisant des contraintes ou des connaissances *a priori*. On peut citer par exemple le cas des images temps-fréquence ou des images sonar (détection et suivi de cibles en acoustique sous-marine).

1.3 Estimation de Mouvement et Ondelettes

Les méthodes hybrides de compression de séquences d'images, qui reposent conjointement sur une prédiction par compensation de mouvement et sur l'utilisation d'une transformée, présentent un grand intérêt pour la transmission à bas débit [TL94]. Nous nous sommes intéressé à une approche d'estimation de mouvement basée sur une méthode différentielle [NR79, WR84, BLBP87], associée à une transformée en ondelettes des images [LS96].

L'hypothèse de travail étant l'invariance de la luminance, le champ de déplacement est estimé en minimisant la *DFD* (*Displaced Frame Difference*) en chaque pixel ou sur un voisinage causal constitué de N pixels incluant le pixel courant, selon une technique pixel-récursive et itérative du type descente de gradient. Ceci donne une image prédite à partir de l'image passée et du champ de déplacement estimé.

Pour estimer des déplacements importants et/ou accélerer la convergence de l'algorithme, il est intéressant d'intégrer cette estimation de mouvement dans une approche multi-résolution : à un niveau de résolution spatiale grossière, on estime les grands déplacements, puis on propage l'information vers des niveaux de résolution plus fine.

Reprenant les travaux de Baaziz [Baa91], nous avons utilisé les ondelettes dyadiques de Daubechies à 4 coefficients [Dau88, Dau92], en s'arrêtant au premier niveau de décomposition (4 sous-bandes). Baaziz base son approche avec transformée en ondelettes sur l'hypothèse de cohérence du mouvement dans les sous-bandes Nous avons mis en évidence que, dans le cas d'ondelettes diadiques discrètes, cette hypothèse suppose le respect de conditions restrictives sur l'amplitude du mouvement des objets dans l'image, et s'avère erronée dans le cas général.

La Fig. 1.11 illustre le phénomène. On y représente en 3-D les quatre sous-bandes (1er niveau de résolution), correspondant à deux images successives de la séquence synthétique présentée sur la Fig. 3.5, p. 15, où un carré noir se déplace de 1 pixel/image selon l'axe des x. Alors que les coefficients correspondant aux contours parallèles au mouvement restent de même signe, les coefficients correspondant aux contours orthogonaux au mouvement (ainsi qu'aux coins du carré) changent de signe d'une image à la suivante. On n'a donc pas invariance par translation.

Pour l'interprétation de ce phénomène, considérons le cas d'un signal 1-D et notons $\psi_{m,n}$ une ondelette-fille, dilatée et translatée de l'ondelette-mère $\psi(x)$:

$$\psi_{m,n} = \sqrt{2^{-m}}\psi(2^{-m}x - n)_{n \in Z} \tag{1.17}$$

Il est alors facile de montrer dans quel cas une translation de Δx dans le signal original discret correspond à une translation de $\Delta n \in Z$ dans les coefficients, c'est-à-dire pour quelles valeurs de Δx l'égalité (1.18) est vérifiée.

$$\langle f(x + \Delta x), \psi_{m,n}(x) \rangle = \langle f(x), \psi_{m,n+\Delta n}(x) \rangle \tag{1.18}$$

On a : $$\int_{-\infty}^{+\infty} f(x + \Delta x)\psi_{m,n}(x)dx = \int_{-\infty}^{+\infty} f(x + \Delta x)2^{-\frac{m}{2}}\psi(2^{-m}x - n)dx$$

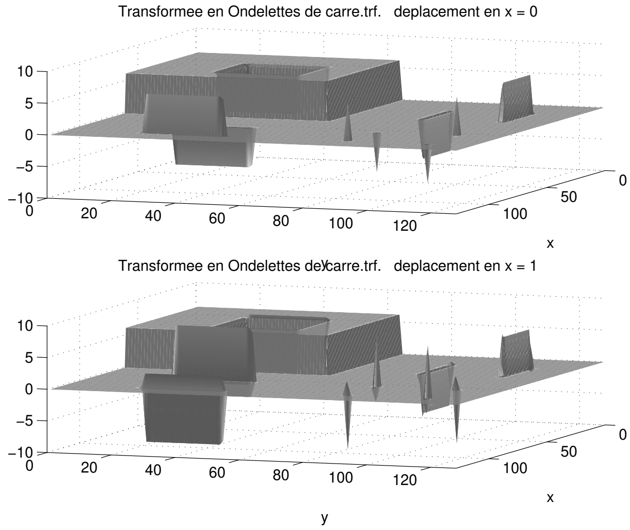


FIG. 1.11 - *Evolution des coefficients d'ondelettes pour un déplacement en x de 1 pixel entre deux images (utilisation des ondelettes discrètes de Daubechies à 4 coefficients).*

$$= \int_{-\infty}^{+\infty} f(y)2^{-\frac{m}{2}}\psi(2^{-m}(y-\Delta x)-n)dy = \int_{-\infty}^{+\infty} f(y)2^{-\frac{m}{2}}\psi\left(2^{-m}y-(2^{-m}\Delta x+n)\right)dy$$

Donc la relation (1.18) n'est vérifiée que si le déplacement de l'objet est un multiple de 2^m. En effet, il faut avoir : $2^{-m}\Delta x = \Delta n \in Z \Leftrightarrow \Delta x = \Delta n.2^m$. Sinon, un déplacement dans le signal ne se traduit pas par un simple déplacement dans les coefficients. Ce résultat correspond en fait à la non invariance par translation, déjà signalée par Mallat [Mal89].

Cependant, ce constat pourrait être utilisé à profit pour estimer le mouvement, l'idée étant de rechercher les périodicités temporelles sur les coefficients de chaque sous-bande, et à différents niveaux de résolution. Cette voie de recherche est en cours d'investigation.

1.4 Segmentation des Lèvres d'un Locuteur

1.4.1 Contexte

Il est bien connu que l'homme s'aide d'informations visuelles pour améliorer sa reconnaissance de la parole, notamment dans un environnement bruité. L'intelligibilité d'un message est meilleure quand l'interlocuteur voit le visage de celui qui parle. En complément du signal auditif, la vision du mouvement des lèvres du locuteur est donc une donnée précieuse, qui peut être exploitée pour réaliser un système automatique de reconnaissance de la parole, ou de synthèse de visages parlants.

Dans ce but, nous avons développé un algorithme de segmentation automatique des lèvres d'un locuteur [LL97b, LL98b], dans le cadre d'un projet de visiophonie haute qualité qui se fixe l'objectif de fusionner les informations issues du geste et du son, pour la transmission bimodale de la parole à distance (compression et télécommunication audio-visuelle).

Le projet consiste à équiper le locuteur d'un casque léger qui comporte, en plus du microphone, une micro-caméra couleur solidaire du crâne et dirigée vers la zone des lèvres (Fig. 1.12.a). On se retrouve donc dans le cas d'une caméra fixe par rapport au locuteur [2] et on peut appliquer

2. "Il faut considérer enfin l'effet de mes propres mouvements, évidemment de première importance quand

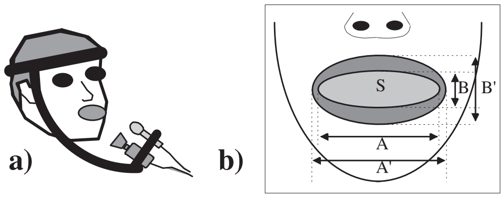


FIG. 1.12 - *a) Micro-caméra et microphone solidaires du crâne ; b) Caractéristiques géométriques des lèvres à extraire de la séquence vidéo.*

les principes de détection de mouvement exposés précédemment, en adaptant les observations et le modèle MRF à l'application envisagée, afin d'extraire de la séquence vidéo les paramètres pertinents du geste de parole : ouverture B et étirement A (Fig. 1.12.b).

1.4.2 Coopération Couleur/Mouvement

Le synoptique de l'algorithme de segmentation est donné sur la Fig. 1.13. Sa structure

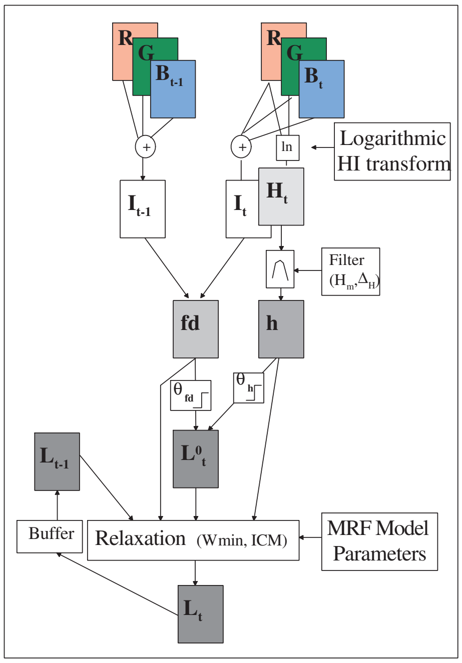


FIG. 1.13 - *Synoptique de l'algorithme Lip-MAD (Motion Automatic Detection).*

générale est similaire à celle de l'algorithme de détection de mouvement, mais on fait coopérer ici deux observations (couleur et mouvement), au lieu de n'en gérer qu'une.

Les grandes lignes du traitement sont les suivantes : on acquiert une séquence couleur RVB (rouge-vert-bleu) du mouvement de la bouche, la région filmée allant des narines au menton. On transforme l'espace RVB en un espace de type HISP (teinte ou *Hue* en anglais, Intensité ou luminance, Saturation, Pureté) en utilisant un traitement d'image de type logarithmique

j'imagine le mouvement des choses. Le moindre mouvement de ma tête fait mouvoir toutes choses." ([Ala41], p. 67).

(modèle LIP) :

$$
\begin{aligned}
H &= 256 \times \frac{V}{R} && (1.19)\\
I &= \frac{R+V+B}{3} && (1.20)\\
S &= 1 - \frac{\min(R,V,B)}{I} && (1.21)\\
P &= I.S = I - \min(R,V,B) && (1.22)
\end{aligned}
$$

Cette transformation s'avère bien adaptée au problème car :

- la teinte rouge, qui est prédominante sur les lèvres, est une observation spatiale pertinente pour la segmentation,
- le bleu est peu informatif, car il est peu présent sur la peau, donc assez bruité et fortement corrélé au vert dans le cas d'une caméra mono-CCD.
- la pureté [PHS95] permet éventuellement de s'affranchir des zones d'ombre (où elle est voisine de zéro),
- enfin, les variations temporelles de luminance donnent l'information de mouvement, qui est prédominant au niveau de la bouche du locuteur.

On utilise deux observations de bas niveau en chaque site s :

- une information spatiale sur la teinte rouge (*hue*) : $h(s)$,
- une information temporelle sur la différence d'images (*frame difference*) : $fd(s)$.

$$
\begin{aligned}
fd(s) &= |I_t(s) - I_{t-1}(s)| && (1.23)\\
h(s) &= \left[256 - \left(\frac{H(s)-H_m}{\sigma_H}\right)^2\right] \times 1_{|H(s)-H_m| \leq 16.\sigma_H} && (1.24)
\end{aligned}
$$

où la notation $1_{condition}$ dénote une fonction binaire qui vaut 1 si la condition est vraie, 0 sinon. H_m et σ_H représentent respectivement la valeur moyenne et l'écart-type de la teinte des lèvres. Ces deux paramètres sont determinés automatiquement par une méthode de raffinement d'histogramme [LL99].

La Fig. 1.14 illustre les champs d'observations obtenus pour deux séquences typiques acquises avec une micro-caméra mono-CCD, l'une dans le cas d'un maquillage naturel, l'autre sans maquillage. On constate qur la transformée couleur utilisée fait bien ressortir les lèvres, et que les mouvements sont essentiellement localisés autour des lèvres.

La combinaison de ces deux observations permet, après un seuillage robuste [LL99], de définir un jeu de quatre étiquettes (Table 1.1).

Chaque étiquette correspond à un codage sur 2 bits de l'information bas niveau relative au pixel s : le bit $m(s)$ code la présence ou l'absence de mouvement, le bit $r(s)$ code la présence ou l'absence de teinte à dominante rouge. L'étiquette a_1 par exemple correspond à la présence concommittante d'une teinte rouge à l'instant t et d'un mouvement significatif entre $t-1$ et t.

On définit alors un modèle énergétique markovien du problème à traiter [LL98b], basé sur une structure de voisinage cubique (Fig. 1.15), et qui intègre des contraintes géométriques sur les lèvres et les informations binaires bas-niveau $m(s)$ et $r(s)$ portées par le pixel s.

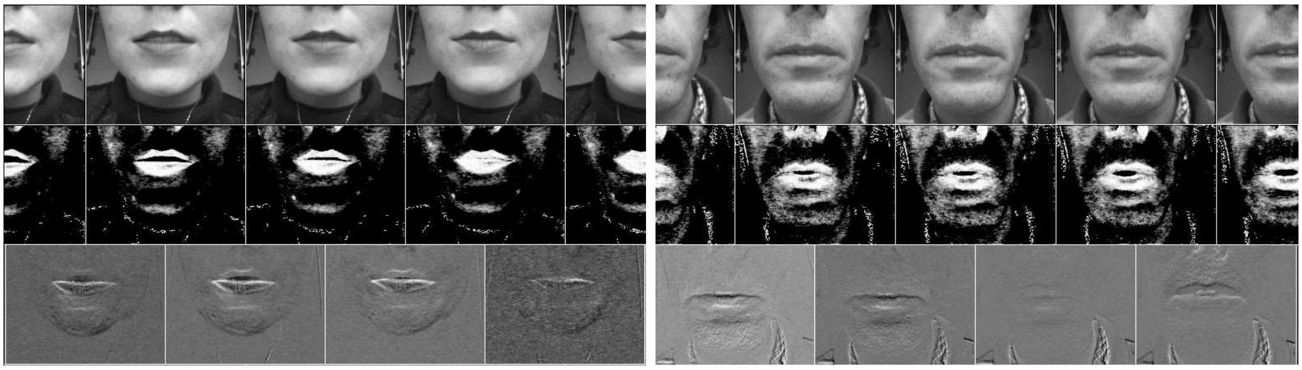


FIG. 1.14 - *Observations sur deux séquences acquises avec micro-caméra solidaire du crâne.* Haut *: séquence des luminances (à gauche, Véronique avec maquillage ; à droite, Benny sans maquillage) ;* Milieu *: séquence des observations spatiales de teinte h (teinte à dominante rouge en blanc) ;* Bas *: séquence des observations temporelles fd (différences positives en blanc, négatives en noir, nulles en gris).*

TAB. 1.1 - *Bits d'information bas-niveau $m(s)$ et $r(s)$ et les quatre étiquettes initiales correspondantes.*

Configuration initiale		mouvement	teinte rouge	
$fd(s)$	$h(s)$	$m(s)$	$r(s)$	l_s
$< \theta_{fd}$	$< \theta_h$	0	0	b_0
$> \theta_{fd}$		1		b_1
$< \theta_{fd}$	$> \theta_h$	0	1	a_0
$> \theta_{fd}$		1		a_1

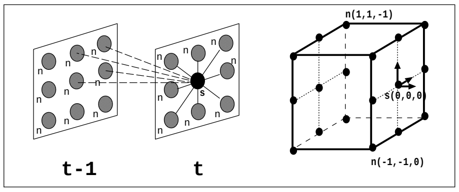


FIG. 1.15 - *Voisinage cubique.*

On est amené à minimiser une fonction d'énergie totale faite de trois termes : deux termes d'attache aux données (car on a deux observations), et un terme de contrainte de modélisation des interactions spatio-temporelles :

$$U = \sum_{s \in S} [U_h(s) + U_{fd}(s) + \lambda U_m(s)] \tag{1.25}$$

où λ est un coefficient de pondération entre l'énergie associée au modèle *a priori* et les énergies d'attache aux données. Pour la définition précise de ces énergies, on se reportera à [LL98b].

La minimisation de cette fonction d'énergie permet de segmenter correctement les lèvres d'un locuteur non maquillé (Fig. 1.16).

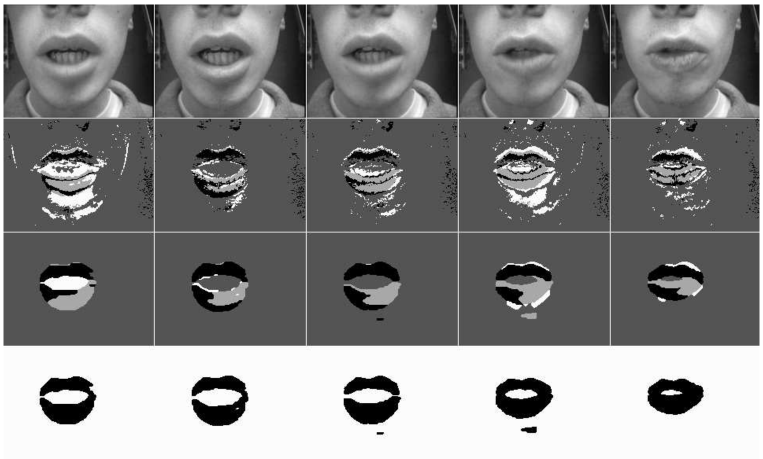


FIG. 1.16 - *Segmentation de lèvres sans maquillage.* De haut en bas *: séquence d'images de luminance ; étiquettes initiales ; champs d'étiquettes après relaxation (les 4 étiquettes sont représentées en niveaux de gris (du noir au blanc : b_1, a_1, b_0, a_0)) ; séquence des étiquettes correspondant aux lèvres (union des étiquettes a_0 et a_1)*

On extrait automatiquement de la séquence la région d'intérêt, comme illustré sur la Fig. 1.17.

Les masques de lèvres obtenus permettent de mesurer avec une précision correcte le paramètre B d'ouverture des lèvres, comme illustré sur la Fig. 1.18, mais aussi et surtout d'initialiser un processus de contours actifs (§1.4.3).

1.4.3 Coopération Segmentation Bayésienne/Contours Actifs

Un contour actif est une courbe paramétrée définie par ses coordonnées cartésiennes x et y en fonction de l'abscisse curviligne s qui évolue suivant la minimisation d'une fonctionnelle F faite de deux termes d'énergie :

$$v(s) = [x(s), y(s)], \quad s \epsilon [0,1] \tag{1.26}$$

$$F : v(s) \longrightarrow \int_0^1 (E_{int}(s) + E_{ext}(s))\, ds \tag{1.27}$$

L'énergie interne est un terme de régularisation du second ordre, qui contrôle le lissage de la courbe par les paramètres de tension α et de courbure β (Eq. (1.28)). L'énergie externe

FIG. 1.17 - *Segmentation de lèvres avec maquillage naturel.* De haut en bas *: séquence des luminances ; séquence des teintes ; Champs initiaux des étiquettes correspondant aux lèvres ; Champs d'étiquettes après relaxation et bounding box détectée ; Superposition des masques des lèvres sur la séquence d'images.*

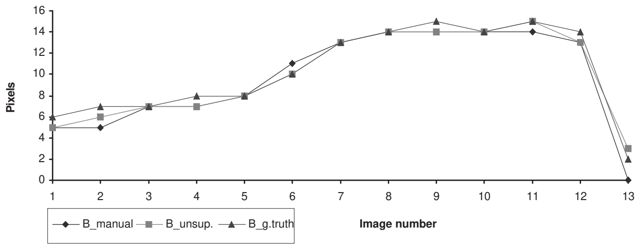


FIG. 1.18 - *Mesures de l'ouverture interne B des lèvres sur la séquence Benny. B_{manual} : mesure obtenue avec paramétrage manuel; $B_{unsup.}$: mesure obtenue avec paramétrage non-supervisé ; $B_{g.truth}$: vérité terrain.*

représente l'attache aux données, qui sont ici les gradients de l'image, puisqu'on recherche les frontières des lèvres (Eq. (1.29)):

$$E_{int}(s) = \alpha|v'(s)|^2 + \beta|v''(s)|^2 \tag{1.28}$$

$$E_{ext}(s) = -\left|\nabla\left(G_\sigma \otimes I\right)(v(s))\right|^2 \tag{1.29}$$

Le vecteur des points de contrôle du *snake* évolue selon un schéma dynamique itératif. Mais le problème bien connu des *snakes* est précisément le choix des points d'initialisation.

D'où notre idée de faire coopérer un prétraitement basé sur la segmentation bayésienne décrite ci-dessus avec un traitement par contours actifs [LDC+99]. La segmentation préalable des lèvres fournit les points initiaux, ce qui lève le problème classique d'initialisation du *snake*. On obtient ainsi des résultats performants pour l'estimation des contours de lèvres, même dans le cas de visages barbus, comme le montre la Fig. 1.19.

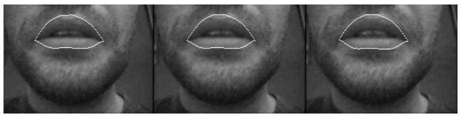


FIG. 1.19 - *Cas d'un visage barbu : contours externes correctement positionnés sur la séquence d'images, grâce à la coopération entre segmentation bayésienne basée couleur/mouvement et contours actifs.*

1.4.4 Positionnement du Travail

Beaucoup de travaux portant sur l'extraction des contours des lèvres imposent des contraintes sévères (maquillage précis des lèvres en bleu, conditions d'éclairement bien adaptées [BLMA92]), ou des hypothèses fortes sur les paramètres de teinte de peau ou sur la localisation de la bouche [SH96]. Certains auteurs utilisent des modèles déformables et insistent sur l'intérêt de gérer la dimension temporelle du signal, au lieu d'un traitement purement statique image par image [Sil96].

Notre travail vise à proposer une méthode automatique et robuste, fonctionnant dans des conditions normales (non contraignantes) d'éclairage et de maquillage, et pour tout type de visage (barbu, peau mate ...).

Les résultats que nous avons obtenus sont encourageants et nous confortent dans l'idée qu'une estimation performante des contours des lèvres passe nécessairement par une coopération entre différentes approches (ici segmentation bayésienne/contours actifs).

Suite à notre travail, certains points clés se sont révélés à nous. D'abord l'utilisation de l'information apportée par la couleur (au lieu des simples niveaux de gris de la luminance) améliore notablement la qualité du traitement. Mais le problème essentiel est de trouver une bonne transformée couleur, qui décorrèle les différentes composantes. La transformée logarithmique que nous utilisons s'avère robuste vis-à-vis de l'éclairement et du type de caméra.

Par ailleurs, l'intégration conjointe d'information temporelle et spatiale dans un unique processus de traitement est une originalité de notre approche, qu'on peut qualifier d'approche dynamique. En effet, beaucoup de méthodes proposées jusqu'à présent adoptaient un traitement purement spatial image par image (approches statiques). Or l'information tirée du passé récent est une aide précieuse à une bonne segmentation de l'instant courant, car la meilleure prédiction du temps qu'il fera demain est encore le temps qu'il fait aujourd'hui (sauf information supplémentaire *a priori* ...).

Chapitre 2

Implantations Matérielles

Beaucoup de travaux portant sur les MRF mentionnent la possibilité d'implantation rapide (calculs locaux sur les voisinages et fortement parallélisables). Mais assez peu de travaux présentent une mise en œuvre opérationnelle et temps-réel. C'est précisément sur ce créneau que nous nous positionnons.

Les algorithmes basés sur une approche markovienne et une relaxation par ICM se décomposent en deux étapes principales (cf. Fig. 1.4, p. 48 et Fig. 1.13, p. 57) :

- un prétraitement qui calcule les observations et les champs initiaux d'étiquettes,
- le traitement proprement dit qui calcule le minimum d'énergie sur les voisinages spatio-temporels.

Prenons l'exemple de notre algorithme de détection de mouvement. Le prétraitement qui consiste simplement en des calculs de différences d'images, de valeurs absolues et des binarisations par seuillage ne pose pas de gros problème d'implantation temps réel. Par contre, la minimisation d'énergie est un processus itératif gourmand en calculs et en mémoire et doit donc faire l'objet d'une mise en œuvre sur une architecture adaptée. Une évaluation grossière montre en effet que la relaxation markovienne compte pour 90% de la charge totale de calcul. On a alors affaire à un problème d'adéquation algorithme-achitecture. Plusieurs types de mises en œuvre sont envisageables :

- les machines parallèles : MIMD (*Multiple Instruction Multiple Data*) ou SIMD (*Single Instruction Multiple Data*)
- les réseaux cellulaires analogiques (ASIC VLSI)
- les composants programmables standards (DSP, FPGA, DVP)

2.1 Les Machines Parallèles : Notre Position

La mise en œuvre des algorithmes markoviens sur une machine de type parallèle est une solution assez naturelle pour exploiter au mieux le parallèlisme intrinsèque à ce type de traitement.

2.1.1 Exemple de Machine MIMD : la TransVision

Un exemple de machine parallèle développée pour la détection de mouvement est la machine Transvision [DBG91]. Le synoptique général de cette machine est donné Fig. 2.1. Son architecture

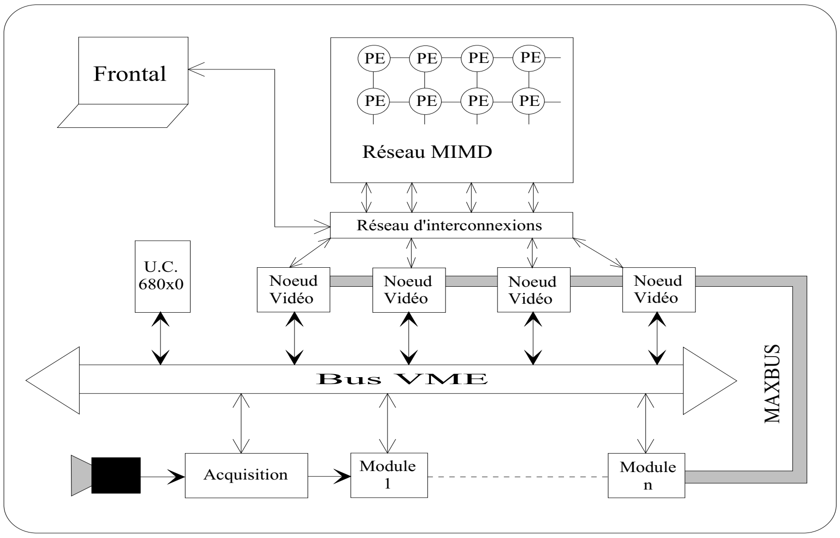


FIG. 2.1 - *Architecture de la machine TransVision.*

est centrée sur l'utilisation de processeurs spécifiques : les transputers. Basée sur l'utilisation conjointe d'une structure pipeline et d'une structure MIMD, la machine Transvision met en œuvre un concept de nœud vidéo faisant office de passerelle entre ces deux structures. Chaque nœud vidéo comporte un transputer et une mémoire vidéo double port accessible par les deux mondes pipeline et MIMD. Un mécanisme de synchronisation (concept de rendez-vous OCCAM) permet d'alimenter et de resynchroniser le réseau MIMD sans goulet d'étranglement. Les modules pipeline et MIMD sont gérés via un bus VME et une unité centrale Motorola 68040 fonctionnant sous VxWork. Le développement d'applications sur les transputers est réalisé sur station de travail ou sur PC en langage OCCAM ou C parallèle.

Le module pipeline considéré dans cette application se restreint à une carte d'acquisition vidéo de la société DataCube permettant l'acquisition et la digitalisation d'images 512 x 512 points en temps réel. Ces images sont alors transmises aux nœuds vidéo par l'intermédiaire d'un réseau MaxBus. Le module MIMD est constitué de douze transputers Inmos T800.

Cette machine fournit donc une chaîne complète de traitement, de l'acquisition vidéo à l'interprétation des résultats.

Indépendamment des performances très intéressantes de cette machine, cette description sommaire met en avant la complexité architecturale d'un tel système, et donc a fortiori ses deux défauts majeurs : son encombrement et son coût.

2.1.2 Exemple de Machine SIMD : la CNAPS

Une machine d'architecture SIMD peut prendre en compte à la fois le caractère parallèle et le caractère identique des calculs effectués en chaque pixel. Les points clés de ce type d'implantation sont la répartition des données entre processeurs et la gestion des communications.

La mise à jour des sites au cours de la relaxation ne peut pas se faire de façon pixel-récursive avec ce genre d'architecture. Selon le nombre de processeurs élémentaires (PE) et la taille d'image à traiter, la mise à jour des sites se fera soit par lignes (ou colonnes), soit par bandes dans l'image, soit par image.

Un exemple d'architecture SIMD est la machine CNAPS (*Connected Network of Adaptive ProcessorS*), représentée sur la Fig. 2.2.

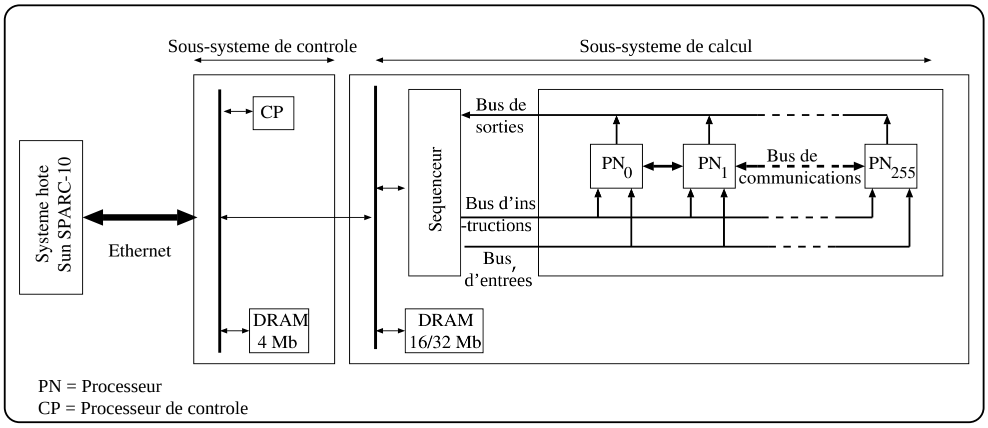


FIG. 2.2 - *Architecture de la machine CNAPS*

Avec cette architecture à 256 processeurs, il est naturel d'attribuer une ligne d'image par processeur (pour des images de taille 256×256). Dans ce cas, la remise à jour des étiquettes doit être colonne-récursive ou image-récursive. En raison des contraintes de communications inter-processeurs, il est préférable d'adopter le schéma image-récursif [Cap95]. De même la visite des sites devra respecter les contraintes imposées par les communications et transferts entre processeurs [CLD98, Cap95].

L'implantation de l'algorithme de détection de mouvement sur cette machine, dont l'architecture est dédiée plutôt aux algorithmes de réseaux de neurones qu'aux algorithmes de traitement de séquences vidéo, conduit à une performance correcte (traitement à la demi-cadence vidéo), sans toutefois atteindre le temps-réel vidéo. Cette performance résulte d'un compromis ente deux limites : la charge de calculs et la taille mémoire locale nécessaire (on ne dispose que de 4 koctets).

Au vu de l'expérience acquise, notre point de vue se résume en trois points. D'une part, les coûts financiers de ces machines parallèles (140 000 US$ pour la CNAPS) sont prohibitifs pour le prototypage rapide d'applications de vision au sein d'unités de recherche à budget limité. D'autre part, leur encombrement matériel (taille) est également prohibitif pour des applications opérationnelles de vision embarquée. Enfin, malgré les caractéristiques impressionnantes de ces machines, les cadences de traitement effectivement obtenues ne sont pas toujours à la hauteur des espérances. Ceci est dû notamment au fait que ces machines intègrent rarement la chaîne complète de traitement du flux vidéo (depuis l'acquisition jusqu'à l'interprétation ou la décision). Ceci nous conduit à nous orienter vers d'autres types d'implantation.

2.2 ASIC Cellulaires en VLSI Analogique

Une alternative, qui prend essentiellement en compte le caractère local des calculs sur un petit voisinage spatio-temporel, consiste à s'inspirer du fonctionnement des réseaux de neurones pour implanter l'algorithme sur un circuit résistif analogique en technologie VLSI. Il faut alors développer une analogie électrique du modèle markovien.

2.2.1 Etat de l'art

La conception de capteurs intelligents (*smart sensors*) pour l'analyse du mouvement vidéo, gourmande en calculs, est un domaine de recherche actif [KMY86, HL93, LSFC93, STM+95].

L'implantation matérielle sur des circuits intégrés spécifiques (ASIC) offre des performances élevées en terme de calculs et de communications. Les principaux avantages des ASIC analogiques sont la vitesse de traitement, la faible consommation, l'encombrement réduit, une haute densité de pixels [CSS92], alors que les inconvénients sont une précision et une flexibilité limitée, et un coût de développement (conception et fabrication) élevé. Deux approches existent pour la conception de capteurs intelligents : l'approche par modèles mathématiques et l'approche biologique [ECdSM97].

Un survol rapide de l'état de l'art révèlent plusieurs points intéressants. D'abord, la plupart des approches sont inspirées de la biologie [MM88, Mea89, MAL91], et utilisent des techniques basées soit sur le gradient soit sur la corrélation [HBB+92, Del93, AMSB95, MASJ93]. La plupart des circuits réalisés ne font que détecter les contours en mouvement (les transitions), ou estimer une vitesse globale sur toute l'image. En effet, ils reposent sur la détection, dans les cellules excitées, d'impulsions positives ou négatives dans les dérivées temporelles [SBK93, KSK95, STM+95]. Quelques circuits fournissent un flux optique épars ou dense [HKLM88], mais aucune segmentation n'est réalisée. Enfin, beaucoup d'auteurs insistent sur les avantages d'adopter un mode de représentation des signaux en terme de courant électrique (au lieu de l'approche classique avec des tensions) pour réaliser les briques de base des réseaux neuronaux en VLSI analogique [ERVDC+94, RVDCMDR95]. Cependant, ils ne sont pas tous d'accord sur le meilleur mode de fonctionnement des transistors MOS (faible inversion [ABP+91, AMSB95], région linéaire [KS95], ou saturation).

A notre connaissance, aucun circuit ne fournit les masques complets des régions en mouvement. Par ailleurs, notre approche n'est pas inspirée de la biologie. Mais il est intéressant de constater que, partant d'un point de vue mathématique (régularisation statistique par champ de Markov), l'implantation de l'algorithme sur réseau analogique aboutit à la conception d'un circuit qui ressemble beaucoup aux réseaux neuronaux cellulaires (CNN) inspirés de la biologie [CY88, CR93]. Enfin, nous utilisons une approche en courants commutés (SI-technique) pour la réalisation du circuit [HBM89, HMP90, HM93b, HM93a, SLK93, GDG94, KLC97, NB96, NVGS95].

2.2.2 Notre Réseau Cellulaire Analogique

Un processus de minimisation d'énergie peut être efficacement implanté matériellement sur un réseau résistif analogique qu'on laisse relaxer jusqu'à son état d'équilibre (dissipation minimale d'énergie) [CSC97]. Mais une telle approche nécessite d'adapter le modèle initial pour trouver une bonne analogie électrique avec le comportement d'un réseau résistif.

Nous avons défini un modèle qui respecte cette contrainte [LD99]. Les pixels sont matérialisés par les nœuds interconnectés d'un réseau électrique résistif. On remplace le champ d'étiquettes binaires (m, b) par un champ d'étiquettes continues variant de "0" (étiquette b) à "1" (étiquette m). Ces étiquettes continues sont alors représentées par des potentiels électriques variant de la tension de masse à la tension d'alimentation du circuit analogique. En utilisant un voisinage d'ordre 1 (4 voisins spatiaux et 2 voisins temporels) Fig. 2.3.a, des potentiels énergétiques quadratiques (cf. Eq.(1.9)), on montre qu'on aboutit à une expression de l'énergie sous la forme :

$$\begin{aligned} U \;=\; \textstyle\sum_{i,j} \quad & \beta_s(l_{ij}-l_{i+1,j})^2+\beta_s(l_{ij}-l_{i-1,j})^2+\beta_s(l_{ij}-l_{i,j+1})^2+\beta_s(l_{ij}-l_{i,j-1})^2+ \\ & \beta_p(l_{ij}-p_{ij})^2+\beta_f(l_{ij}-f_{ij})^2+K\left[o_{ij}-\Psi(l_{ij})\right]^2 \end{aligned} \tag{2.1}$$

où p_{ij} et f_{ij} sont les potentiels électriques des voisins passé et futur, l_{ij} celui du pixel courant $s=(i,j,t)$, o_{ij} l'observation correspondante et $K = 1/2\sigma^2$ une constante.

Le minimum d'énergie correspond à l'annulation des dérivées partielles par rapport à tous les potentiels électriques l_{ij} en chaque nœud (i,j) du réseau :

$$\forall ij \quad \frac{\partial U}{\partial l_{ij}} = 0 \quad \Leftrightarrow \quad \beta_s \nabla^2 l_{ij} + \beta_f(l_{ij} - f_{ij}) + \beta_p(l_{ij} - p_{ij}) - K\frac{\partial \Psi}{\partial l_{ij}}[o_{ij} - \Psi(l_{ij})] = 0 \qquad (2.2)$$

où $\nabla^2 l_{ij} = 4l_{ij} - l_{i+1,j} - l_{i-1,j} - l_{i,j+1} - l_{i,j-1}$ est une approximation discrète du Laplacien.

En choisissant une fonction $\Psi(l_{ij}) = A_{ij}(l_{ij} - p_{ij})$, avec $A_{ij} = \alpha_2(p_{ij} - f_{ij}) - \alpha_1$ et $\alpha_2 > \alpha_1 > 0$, ce qui permet de bien éliminer l'écho du mouvement [LD99], on obtient finalement une équation dont l'analogie électrique est immédiate :

$$\forall ij \quad C\frac{\partial l_{ij}}{\partial t} = \beta_s \nabla^2 l_{ij} + \beta_f(l_{ij} - f_{ij}) + \beta_a(l_{ij} - p_{ij}) - KA_{ij}o_{ij} \qquad (2.3)$$

où $\beta_a = \beta_p + KA_{ij}^2$. La cellule élémentaire associée à chaque site est donnée Fig. 2.3 b. L'idée

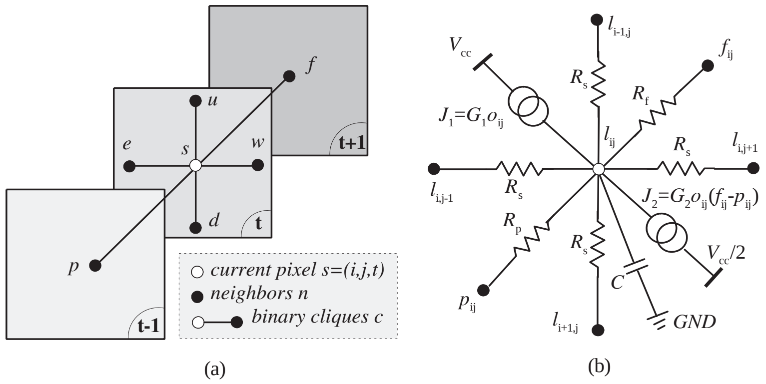


FIG. 2.3 - *(a) Voisinage spatiotemporel η_s du site courant s, avec cliques binaires ; u, d, e, w, p, f dénotent les voisins de s, resp. up, down, east, west, passé et futur. (b) Cellule analogique élémentaire correspondante : $R_s = 1/\beta_s$, $R_f = 1/\beta_f$, $R_p = 1/\beta_a$, $G_1 = K\alpha_1$ et $G_2 = K\alpha_2$.*

principale du fonctionnement de la cellule électrique consiste à injecter ou pomper du courant en chaque nœud du réseau (grâce à deux générateurs de courant commandés en tension par les observations), pour faire monter ou baisser le potentiel électrique du nœud, selon que l'amplitude de l'observation correspondante est forte (pixel mobile) ou faible (pixel fixe).

On peut intégrer sur chaque cellule élémentaire du réseau un photorécepteur, ce qui fait du circuit électrique une caméra "intelligente". Le schéma de la cellule complète est représenté sur la Fig. 2.4.

L'architecture du réseau analogique résultant est analogue à celle d'une caméra CCD classique (Fig. 3.14, page 24). Mais c'est une technologie CMOS qui est utilisée pour la concpetion de la cellule, et non une technologie CCD, dont le processus de fabrication est plus coûteux.

Deux résultats de simulation électrique du réseau sont présentés. Le premier résultat, obtenu sur une séquence synthétique, met en évidence le comportement correct du réseau et son adaptation temporelle (Fig. 2.5).

Le deuxième exemple montre un résultat typique de détection, obtenu sur une séquence réelle acquise avec une caméra (Fig. 2.6).

2.2.3 Bilan

Les avantages pratiques d'une telle mise en œuvre matérielle sont évidents : encombrement réduit du système (taille d'une puce électronique), consommation réduite du circuit, vitesse

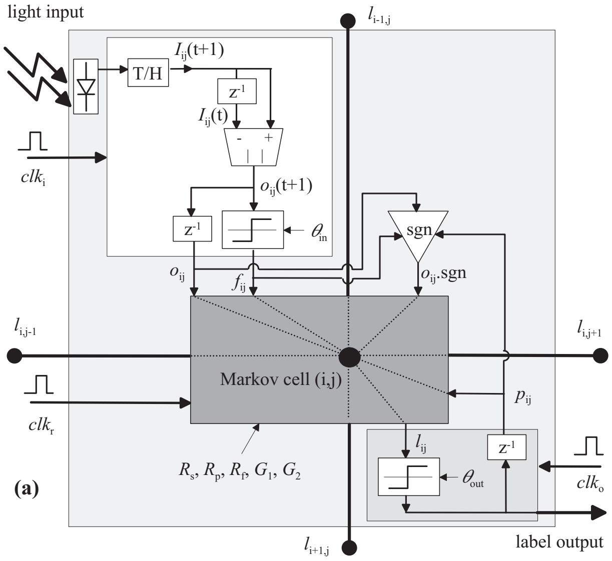


FIG. 2.4 - *Cellule complète intégrant l'acquisition (photorécepteur), le prétraitement (calcul de o_{ij}, f_{ij}), le traitement (relaxation de la cellule de Markov), et le post-traitement (binarisation, et mémoire de sortie). T/H représente le circuit suiveur-bloqueur (Track-and-Hold). L'étage de prétraitement (en blanc sur la Fig.), peut être implanté avec la technique des courants commutés (SI-technique).*

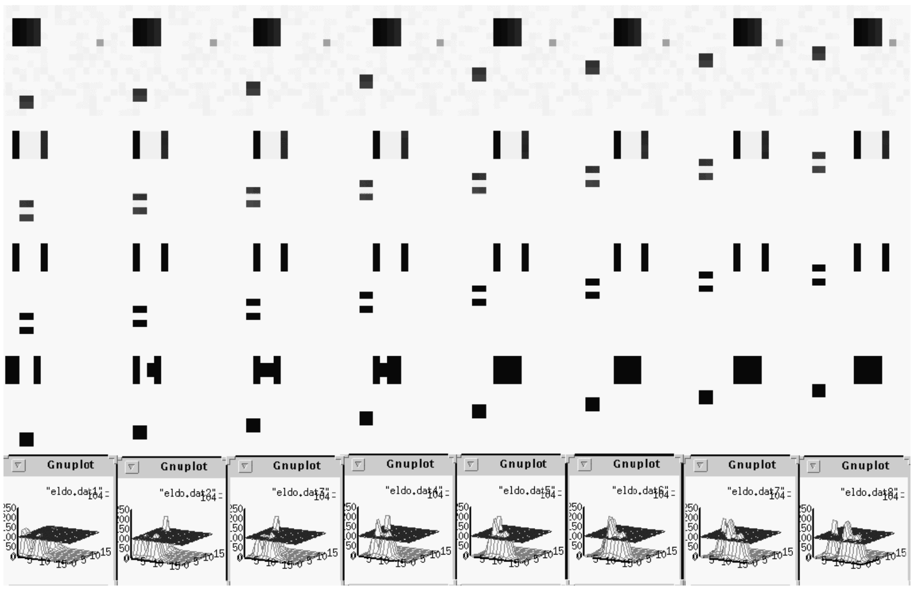


FIG. 2.5 - De haut en bas*: séquence synthétique avec deux carrés mobiles légèrement texturés; séquence des observations; champs d'étiquettes futures; masques de mouvement finaux; représentation 3-D des potentiels électriques après relaxation du réseau. N.B. les champs d'observations et le futur sont décalés d'une image vers la gauche, par rapport à la séquence d'images et à la séquence de masques. Paramètres:* $\theta_{in} = 11, \theta_{out} = 104$, $R_s = 400k\Omega, R_f = 200k\Omega, R_p = 206k\Omega$, $G_1 = 950.10^{-8} Siemens, G_2 = 2015.10^{-8} Siemens.$

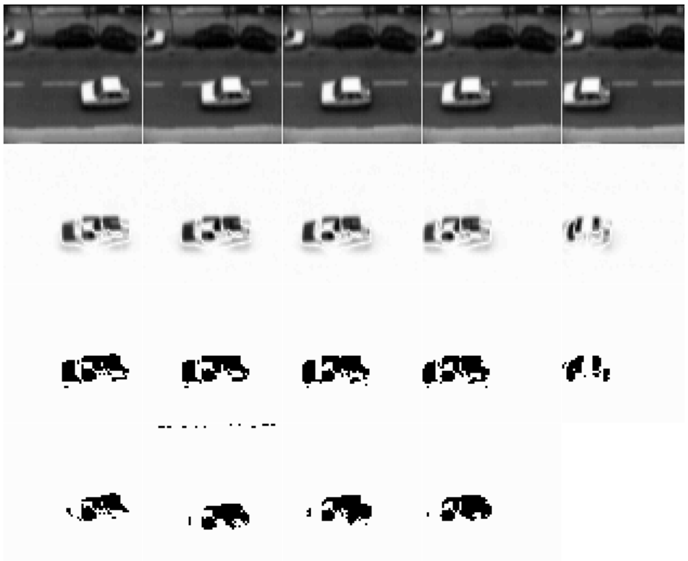


FIG. 2.6 - *Simulation électrique du réseau. De haut en bas : 1) séquence naturelle : scène de rue avec une voiture mobile ; 2) séquence des observations en niveaux de gris ; 3) séquence binaire des champs d'étiquettes initiaux (champs futurs) ; 4) masques binaires détectés. Paramétrage :* $\theta_{in} = 30, \theta_{out} = 130$, $R_s = 400k\Omega, R_f = 200k\Omega, R_p = 216k\Omega$, $G_1 = 700.10^{-8}Siemens, G_2 = 1400.10^{-8}Siemens$.

de traitement rapide (durée de relaxation du réseau de l'ordre de la μs), intégration sur un seul capteur de l'acquisition en parallèle des informations sur tous les pixels, et du traitement simultané de tous ces pixels.

Par ailleurs, ce type de réseau cellulaire présente un avantage théorique notable : en effet, la relaxation du réseau est un processus optimal dans le sens où elle se fait simultanément et en parallèle sur tous les nœuds du réseau. Donc la mise à jour des sites est instantanée et simultanée. Ceci lève de façon élégante les contraintes et défauts classiques des algorithmes de relaxation de type récursif et itératif.

En contrepartie, les inconvénients sont bien sûr le coût élevé de développement d'un prototype (200 000 US$), les problèmes technologiques d'intégration VLSI à résoudre (taille de la cellule, consommation, interconnexions ...), et le manque de flexibilité.

2.3 Composants Programmables Standards

2.3.1 Carte Electronique à Base de DSP et FPGA

Après une étude détaillée de l'algorithme de détection de mouvement, il s'avère possible de proposer une architecture matérielle alternative aux solutions coûteuses présentées plus haut. Elle consiste à implanter l'algorithme sur une carte au format PC à base de composants standards.

L'architecture de la carte est présentée sur la Fig. 2.7.

En effet, le prétraitement (filtrage, différence et seuillage) peut être réalisé avec des circuits logiques programmables (FPGA) et, grâce aux processeurs de signaux numériques (DSP) actuellement disponibles sur le marché, on peut mettre en œuvre le traitement proprement dit (la relaxation énergétique) à moindre frais. Tout ce qui concerne l'acquisition, le prétraitement et la visualisation

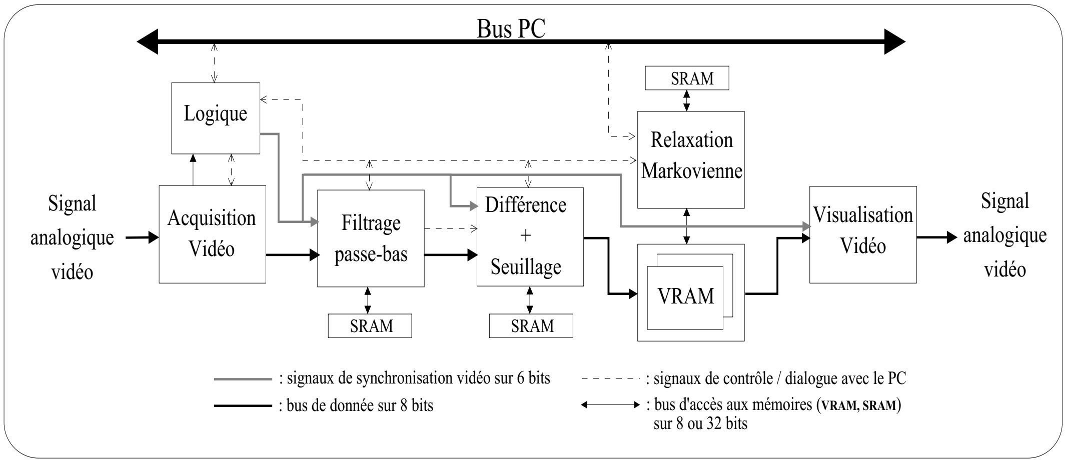


FIG. 2.7 - *Architecture de la carte PCMARKOV.*

(FPGA et DAC/ADC) fonctionne à la cadence pixel (15 MHz). Seule la relaxation markovienne est réalisée en mode asynchrone par un DSP (Motorola 96002), avec une mémoire tampon qui sert d'interface. C'est le point clé de l'architecture proposée.

Une carte opérationnelle, qui traite des images de taille 256×256 à la cadence de 15 images par seconde, a été développée au laboratoire (Fig. 3.15, page 25). Cette carte s'insère sur le bus ISA d'un ordinateur personnel, est fonctionne de manière complètement autonome après chargement de son programme (phase initiale de *boot*). On dispose ainsi d'un prototype complet, peu encombrant et peu coûteux (8000 US$), allant de l'acquisition à la visualisation des masques des objets mobiles, et approchant un fonctionnement en temps réel. En contre-partie, cette solution offre moins de flexibilité que les machines parallèles.

Ce type de matériel est bien adapté à la vision embarquée. Une application typique de la détection de mouvement est le contrôle du trafic routier ou la télésurveillance à l'entrée d'un site grâce à une caméra "intelligente". La Fig. 2.8 présente un résultat de détection automatique du mouvement d'un piéton sur un trottoir, obtenu grâce à la carte à DSP décrite ci-dessus. On constate le bon comportement du système qui élimine les bruits parasites. Notons que l'algorithme détecte également l'ombre du piéton sur le capot de la voiture en stationnement (il ne s'agit pas d'une fausse détection).

2.3.2 Les Processeurs Vidéo Numériques (DVP)

L'apparition récente sur le marché de processeurs vidéo programmables (*Digital Video Processors*), qui intègrent dans leur architecture une certaine dose de parallélisme (TMS320C80, PVP du CNET ...) semble être une solution d'avenir offrant un bon compromis entre complexité de développement, performance de traitement, coût et encombrement. Il s'agit alors d'un parallélisme dit à gros grain, contrairement au parallélisme à grain fin des machines parallèles mentionnées au paragraphe 2.1.

Un exemple de ce type de composant est le processeur PVP développé par le CNET. Son architecture, de type SIMD avec 8 ou 16 PE sophistiqués, est donnée Fig. 2.9.

Ce processeur vidéo a été conçu pour des applications vidéo en télécommmunications (compression MPEG4). En plus de sa puissance de calcul élevée (2 Gops), une de ses caractéristiques principales est sa grande bande passante qui permet des entrées-sorties à haut débit (4Gbits/s).

Nous avons implanté le détecteur de mouvement sur ce type de matériel, et l'on atteint une cadence de traitement impressionnante : 150 images/s pour une taille d'image 256×256

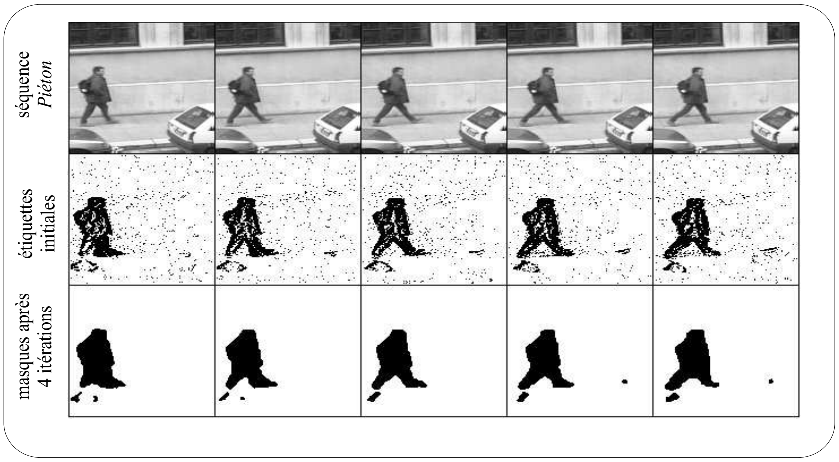


FIG. 2.8 - *Détection de mouvement avec la carte PCMARKOV.* De haut en bas *: Séquence de rue acquise à la cadence de 25 images/seconde (mouvement peu important entre deux images successives); Champ d'étiquettes initiales (fausses détections dues au bruit de la caméra); Masques obtenus après relaxation (convergence atteinte en 4 itérations).*

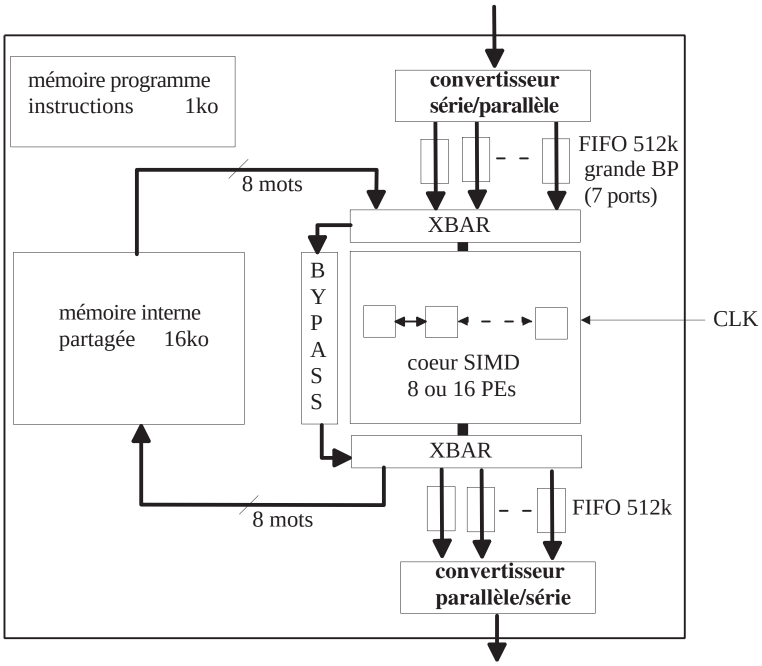


FIG. 2.9 - *Architecture simplifiée du PVP.*

et une fréquence d'horologe de 70MHz. Par contre, sa structure spécifique oblige à un exercice non trivial d'adéquation algorithme/architecture dans le cas de traitements plus complexes : une première évaluation de l'algorithme de segmentation des lèvres sur le simulateur logiciel du PVP nous donne une cadence d'environ 15 images/s.

La conception d'une carte au format PC (bus PCI), basée sur ce processeur vidéo est actuellement à l'étude. Elle devrait comporter une carte-mère avec FPGA et convertisseurs ADC/DAC pour l'acquisition, les prétraitements simples et la restitution vidéo, et une carte-fille mezzanine, avec PVP et mémoires associées, pour les traitements lourds.

L'implantation du détecteur de mouvement sur un autre type de processeur vidéo (TMS320C80) est également à l'étude. Il s'agit cette fois d'une petite structure parallèle de type MIMD à 4 PE (4 DSPs), reliés entre eux par un *crossbar* offrant une grande souplesse d'interconnexion. De plus, un processeur RISC flottant 32 bits permet d'exécuter des opérations mathématiques complexes tout en contrôlant les PE.

La comparaison des performances que l'on arrivera à atteindre sur ces deux types de DVP se révélera certainement riche d'enseignement sur le type architecture le mieux adapté aux algorithmes markoviens.

En parallèle au développement des processeurs vidéo numériques, on peut aussi signaler l'apparition sur le marché de l'instrumentation de nouvelles normes de bus rapide, comme le bus FireWire (norme IEEE 1394) qui rend possible une liaison série directe et haut débit entre une caméra et un PC, ce qui ouvre des perspectives intéressantes pour la conception de cartes de traitement vidéo au format PC ...

Chapitre 3

Perspective de Recherche

3.1 Le Point sur la Situation

Ayant initié au LTIRF l'activité "analyse du mouvement dans les séquences d'images", j'ai constitué et j'anime un petit groupe de chercheurs autour de ce thème : le groupe Mouvement-Markov (M&Ms). Le laboratoire a acquis une compétence désormais reconnue dans ce domaine.

Cette activité a été l'occasion de renforcer les collaborations entre les trois équipes du LTIRF (Vision, Réseaux Neuronaux et Circuits Spécifiques), aussi bien sur des aspects théoriques de traitement (multirésolution, coopération entre segmentation bayésienne et contours actifs, perception du mouvement et de la couleur), que sur des aspects pratiques de mise en œuvre matérielle des algorithmes (Fig. 3.1).

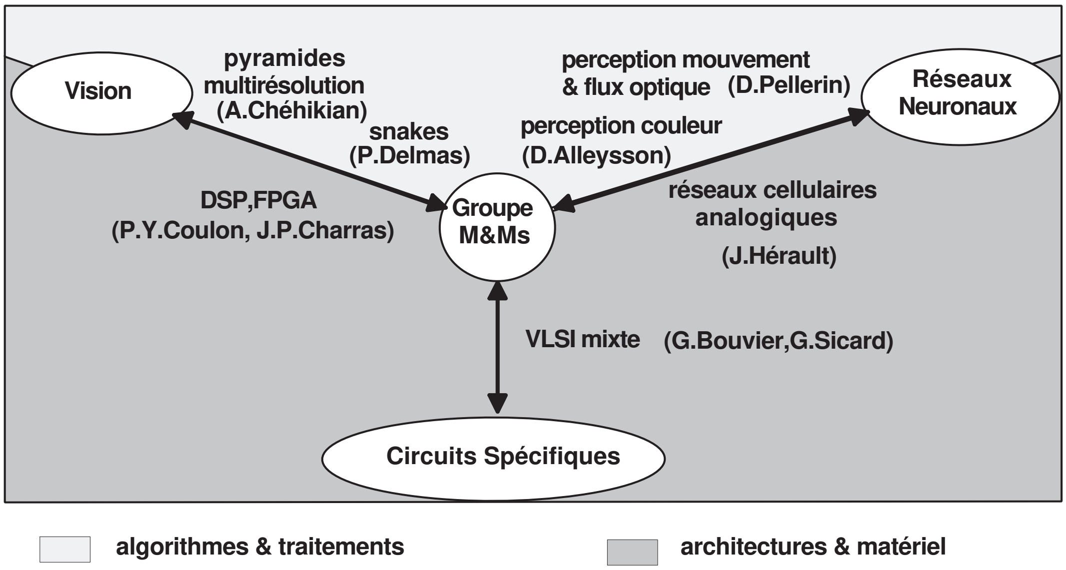


FIG. 3.1 - *Collaborations dans la structure LTIRF.*

L'étude du mouvement dans les images a été pour moi un excellent catalyseur de ma recherche. C'est en quelque sorte l'archétype d'une tâche liée à la vidéo : d'une part, c'est un **problème d'analyse mal posé** qui requiert des méthodes de traitement robustes et performantes (*e.g.* régularisation statistique sous contraintes ...) D'autre part, ses applications effectives nécessitent la prise en compte de la notion de cadence vidéo, donc une implication forte dans les **implantations matérielles temps-réel**. Ces deux constats conduisent donc à aborder de front le traitement du signal et de l'image, l'informatique et l'électronique.

3.2 Valorisation des Acquis à Moyen Terme

J'élargis actuellement ma thématique à la **segmentation spatio-temporelle** de scènes dynamiques, tâche essentielle de la vision artificielle, qui trouve application dans des domaines en pleine expansion comme les **télecommunications audio-visuelles**, où existent une forte demande et une forte potentialité (internet, multimédia, compression, codage, interface homme/machine) et en vision embarquée (télésurveillance, robotique mobile, conduite automobile guidée, contrôle non destructif, diagnostic médical, météorologie ...).

Dans cette optique, étant parti du cas simple de la détection de mouvement avec caméra fixe, j'aborde maintenant la segmentation dans le cas d'une caméra mobile : compensation de mouvement (avec implantation sur processeur vidéo ou DSP), estimateurs robustes (dont certains sont implantables en VLSI analogique), segmentation de visages avec coopération couleur - mouvement - texture et coopération segmentation bayésienne - modèles déformables (snakes).

Ayant déjà pris une orientation vers les réalisations matérielles, je souhaite continuer à m'investir sur les aspects **adéquation algorithme/architecture** (AAA), qui sont un challenge intéressant de la recherche appliquée, les concepts architecturaux ayant une pérennité malgré l'évolution rapide du matériel. Deux stages et un contrat sont prévus en 1999 pour l'implantation de nos algorithmes sur processeurs vidéo de type SIMD (PVP du CNET et TMS320C80). Dans un proche avenir, je prévois de co-encadrer une thèse en micro-électronique (avec Gérard Bouvier), en vue de la réalisation du circuit ASIC VLSI (réseau 100×100) pour la détection de mouvement en temps-réel vidéo.

La Fig. 3.2 tente une synthèse chronologique de mes thèmes d'intérêt et donne l'orientation de ma recherche à moyen terme.

3.3 Perspectives au Sein du LIS (Long Terme?)

Dans le cadre de la fusion des deux laboratoires LTIRF et CEPHAG au sein de la nouvelle entité LIS (Laboratoire des Images et Signaux, structuré en cinq groupes et mixant les compétences en signal et en image), plusieurs perspectives de recherche s'ouvrent à l'horizon 2000 (Fig. 3.3).

Dans la logique de mes travaux antérieurs, j'appartiens aux deux groupes CLS (Circuits et Langages Spécifiques) et API (Analyse-Perception-Interprétation). J'ai notamment pris en charge la valorisation du logiciel MUSTIG pour le prototypage d'architectures matérielles (générateur VHDL pour la programmation de DSP et FPGA). Un projet de stage a été proposé cette année sur ce sujet, en collaboration avec J. Liénard.

J'ai également vocation à participer au groupe Communication, en rapport avec mes travaux sur le projet Labiophone (**télécommunication et compression audiovisuelles**).

Le savoir faire spécifique du LIS-Viallet (ex-LTIRF), et notamment l'apport des méthodes de régularisation statistique que j'ai développées, s'avèrent aptes à traiter certains problèmes de **segmentation** de séquences d'images sonar, spécifiques à l'acoustique sous-marine, problèmes sur lesquels travaille le groupe SIN (Signaux et Images Naturels) du LIS-Campus (ex-CEPHAG), en partenariat avec la DCE (ex-Division des Chantiers Navals) pour la détection et la poursuite de cibles sous-marines.

La segmentation bayésienne d'images temps-fréquence est également un point de convergence potentiel avec les travaux du LIS dans le groupe API.

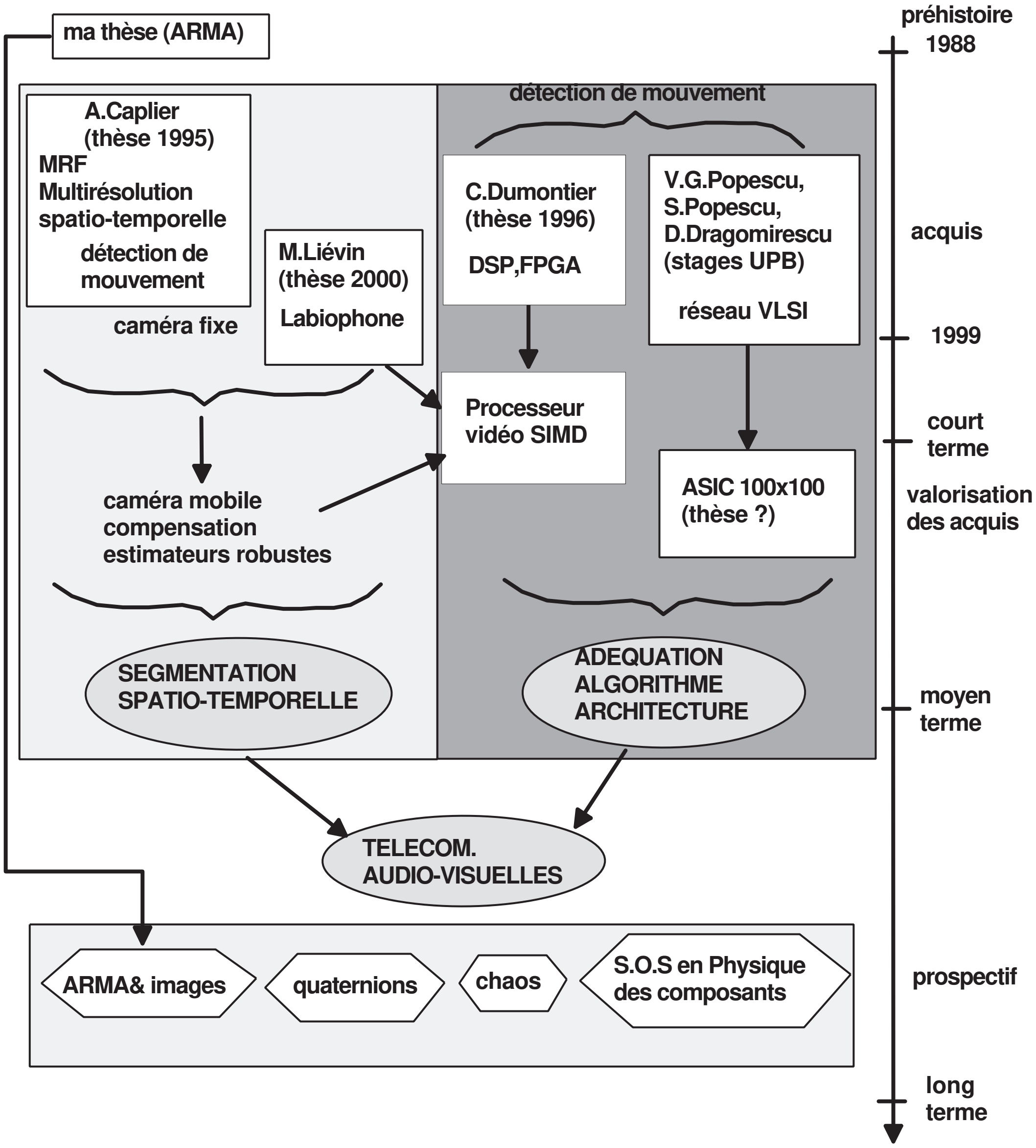


FIG. 3.2 - *Vision chaotique.*

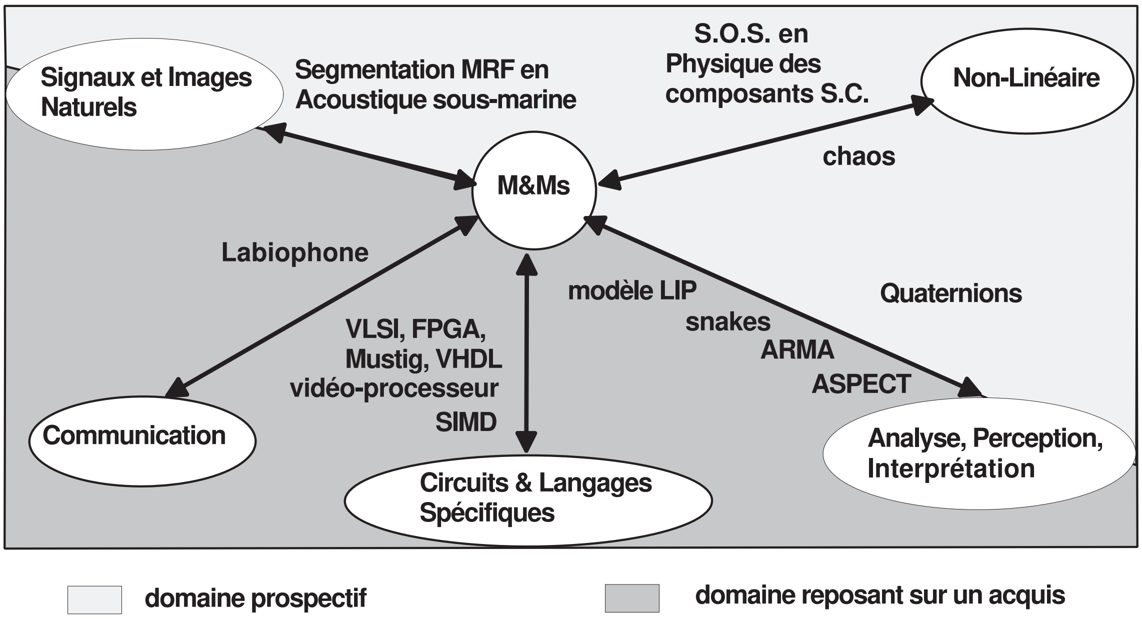


FIG. 3.3 - *Positionnement futur dans la structure LIS.*

Une autre voie de recherche, qui me ferait rebondir sur mon acquis en thèse, concerne d'une part l'analyse spectrale intelligente (projet ASPECT du LIS), et la réduction de bruit, thèmes abordés par le passé, et que j'ai l'opportunité d'approfondir et de valoriser dans les applications au sein du groupe API. Et d'autre part l'application des modèles ARMA aux images.

Une collaboration engagée en 1997 avec le LPCS (Laboratoire de Physique des Composants à Semi-conducteurs) porte sur l'acquisition de signaux non-stationnaires dans les composants à semi-conducteurs, et leur analyse par l'étude de **statistiques d'ordre supérieur** (S.O.S.), qui correspond à une compétence reconnue du LIS [LAC97], thème dans lequel je souhaite m'impliquer pour cette application en physique des composants.

Sur le plan théorique, l'exploration de changements d'espaces de représentation des données manipulées est un champ d'investigation qui m'intéresse. La théorie des **quaternions** (nombres hypercomplexes) en fait partie. Cette théorie n'est pas nouvelle [KS89], mais les tous récents travaux pour leur application au traitement du signal et des images me semblent riches de potentialités : FFT ultra-rapide [Che95], estimation de mouvements 2-D de translation [BS97], gestion de l'espace des couleurs [SE98]. Je compte évidemment poursuivre aussi mon investigation sur le modèle LIP (Logarithmic Image Processing), et les tranformées multi-échelles (ondelettes 3-D).

Enfin, les phénomènes chaotiques sont également attractifs à mes yeux, et leurs applications commencent à se développer (cryptage, conception de circuits et systèmes non-linéaires). L'existence du groupe Non-Linéaire du LIS me donne l'opportunité de développer cet axe passionnant.

"Lecteur,
au sortir de ces landes arides
qu'il a bien fallu traverser,
je souhaite que jeunesse te garde."

([Ala41], p. 260)

Bibliographie

[ABP+91] A. G. Andreou, K. A. Boahen, P. O. Pouliquen, A. Pavasovic, R. E. Jenkins, and K. Strohbehn. Current-mode subthreshold MOS circuits for analog VLSI neural systems. *IEEE Trans. on Neural Networks*, 2(2):205–213, March 1991.

[ADM81] J. K. Aggarwal, L. S. Davis, and W. N. Martin. Correspondence processes in dynamic scene analysis. *Proceedings of the IEEE*, 69(5):562–572, May 1981.

[AKM93] T. Aach, A. Kaup, and R. Mester. Statistical model-based change detection in moving video. *Signal Processing*, 31(2):165–180, March 1993.

[Ala41] Alain. *Eléments de philosophie.* Folio Essais. Gallimard, Paris, 1941.

[AMSB95] A. G. Andreou, R. C. Meitzler, K. Strohbehn, and K. A. Boahen. Analog VLSI neuromorphic image acquisition and pre-processing systems. *Neural Networks*, 8(7/8):1323–1347, 1995.

[Ana89] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision*, 2:283–310, 1989.

[Asf97] A. Asfour. Analyse automatique du mouvement des lèvres d'un locuteur. Rapport Ingénieur ENSPG, LTIRF/INPG, Grenoble, Juin 1997.

[Baa91] N. Baaziz. *Approches d'estimation et de compensation de mouvement multirésolutions pour le codage de séquences d'images.* PhD thesis, Université de Rennes 1, October 1991.

[BC98] S. Berthout and G. Clédat. Etude et réalisation d'une interface analogique/numérique et des circuits de contrtle mémoire et d'arbitrage de bus d'un processeur vidéo. Rapport Ingénieur ENSERG, TCEC, Meylan, Juin 1998.

[Bes86] J. Besag. On the statistical analysis of dirty pictures. *J. R. Statist. Soc. B*, 48(3):259–302, 1986.

[BL93] P. Bouthémy and P. Lalande. Recovery of moving object masks in an image sequence using local spatiotemporal contextual information. *Optical Engineering*, 32(6):1205–1212, June 1993.

[BLBP87] J. Biemond, L. Looijenga, D.E. Boekee, and R.H.J.M. Plompen. A pel-recursive Wiener-based displacement estimation algorithm. *Signal Processing*, 13:399–412, 1987.

[BLMA92] C. Benoît, T. Lallouache, T. Mohamadi, and C. Abry. A set of French visemes for visual speech synthesis. In G. Bailly, C. Benoît, and T.R. Sawallis, editors, *Talking Machines: Theories, Models and Designs*, pages 485–504, Amsterdam, North-Holland, 1992. Elsevier Science Publishers B.V.

[Bon98] D. Bonifaci. Mise en œuvre de traitements statistiques sur des signaux de bruit. DEA Microélectronique, LPCS/INPG, Grenoble, Juin 1998.

[Bou96] X. Boudet. Détermination du flot optique dans une séquence d'images par mise en correspondance hiérarchique. Stage élève-ingénieur ENSERG, LTIRF/INPG, Grenoble, Septembre 1996.

[BP95] B. Benmiloud and W. Pieczynski. Estimation des paramètres dans les chaînes de Markov cachées et segmentation d'images. *Traitement du Signal*, 12(5):433–454, 1995.

[BS97] T. Bülow and G. Sommer. Algeraically extended representations of multi-dimensional signals. In *Proc. of 10th Scandinavian Conf. on Image Analysis*, pages 559–566, Lappeenrenta, Finland, June 9-11 1997.

[BSGG92] G. Bilbro, W. Snyder, S. Garnier, and J. Gault. Mean field annealing: A formalism for constructing GNC-like algorithms. *IEEE Trans. on Neural Networks*, 3:131–138, January 1992.

[Bur84] J. P. Burt. The Laplacien pyramid as a compact image code. *IEEE Trans. on Communications*, 31(4):532–540, 1984.

[Cap95] A. Caplier. *Modèles markoviens de détection de mouvement dans les séquences d'images : Approche spatio-temporelle et mises en œuvre temps réel.* PhD thesis, Institut National Polytechnique, Grenoble, 20 décembre 1995.

[CB90] P. B. Chou and C. M. Brown. The theory and practice of Bayesian image labeling. *International Journal of Computer Vision*, 4:185–210, 1990.

[CC91] A. Chéhikian and J.L. Crowley. Fast computation of optimal semi-octave pyramids. In *Proc. of 7th Scandinavian Conf. on Image Analysis*, pages 18–27, Aalborg, Denmark, 1991.

[CCL97] A. Chéhikian, P. Y. Coulon, and F. Luthon, editors. *Ecole des Techniques Avancées en Signal-Image-Parole (ETASIP'97)*, Grenoble, 9-12 Septembre 1997. LTIRF-INPG. Session 2, 486 pages.

[CDLC96] A. Caplier, C. Dumontier, F. Luthon, and P. Y. Coulon. Algorithme de détection de mouvement par modélisation markovienne. Mise en oeuvre sur DSP. *Traitement du Signal*, 13(2):177–190, 1996.

[Chb98] O. Chbani. Compensation du mouvement d'une caméra par estimation de mouvement sur DSP Motorola 96002. Rapport Ingénieur ENSERG, LIS/INPG, Grenoble, Juin 1998.

[Che92] A. Chehikian. Algorithmes optimaux pour la génération de pyramides d'images passe-bas et laplaciennes. *Traitement du Signal*, 9(4):297–307, 1992.

[Che95] V.M. Chernov. Discrete orthogonal transforms with data representation in composition algebras. In *Proc. of 9th Scandinavian Conf. on Image Analysis*, pages 357–364, Uppsala, Sweden, June 1995.

[Chr96] A. Christoforou. Estimation de mouvement et transformée en ondelettes pour la compression de séquences d'images. DEA SIP, LTIRF/INPG, Grenoble, Juillet 1996.

[CJ93] R. Chellappa and A. Jain, editors. *Markov Random Fields: Theory and Application*. Academic Press, Inc., San Diego, 1993.

[CL93] A. Caplier and F. Luthon. Segmentation du mouvement dans une séquence d'images selon une approche markovienne. Journées thématiques du GDR 134, GT2, Champs de Markov - Segmentation, Telecom Paris, 14 Octobre 1993.

[CL95a] A. Caplier and F. Luthon. A 3D space and time MRF based algorithm for motion detection. In *Mediterranean Conf. on Electronics and Automatic Control (MCEA'95)*, pages 147–152, Grenoble, France, September 13-15, 1995.

[CL95b] A. Caplier and F. Luthon. An MRF-based spatiotemporal multiresolution algorithm for motion detection. In *9th Scandinavian Conf. on Image Analysis (SCIA'95)*, pages 559–566, Uppsala, Sweden, June 6-9, 1995.

[CL95c] A. Caplier and F. Luthon. A new spatiotemporal approach for image analysis. Application to motion detection. In *6th International Conf. on Computer Analysis of Images and Patterns (CAIP'95)*, pages 246–253, Prague, Czech Republic, September 6-8, 1995.

[CL95d] A. Caplier and F. Luthon. Spatiotemporal multiresolution associated to MRF modelling for motion detection. In *5th IEE International Conf. on Image Processing and its Applications (ICIPA'95)*, pages 158–162, Edinburgh, U.K., July 3-6, 1995.

[CL97] A. Caplier and F. Luthon. Approche spatio-temporelle pour l'analyse de séquences d'images. Application en détection de mouvement. *Traitement du Signal*, 14(2):195–208, 1997.

[CLD95] A. Caplier, F. Luthon, and C. Dumontier. Algorithme markovien de détection de mouvement. Mises en oeuvre temps réel? In *15e Colloque GRETSI*, pages 1033–1036, Juan-les-Pins, 18-21 Septembre 1995.

[CLD98] A. Caplier, F. Luthon, and C. Dumontier. Real-time implementations of an MRF-based motion detection algorithm. *Real-Time Imaging*, 4(1):41–54, February 1998.

[CR93] L. O. Chua and T. Roska. The CNN paradigm. *IEEE Trans. on Circuits and Systems-I*, 40(3):147–156, March 1993.

[CSC97] E. Y. Chou, B. J. Sheu, and R. C. Chang. VLSI design of optimization and image processing cellular neural networks. *IEEE Trans. on Circuits and Systems-I*, 44(1):12–20, January 1997.

[CSS92] C. P. Chong, C. A. T. Salama, and K. C. Smith. Image-motion detection using analog VLSI. *IEEE Journal of Solid-State Circuits*, 27(1):93–96, January 1992.

[CTS97] M. M. Chang, A. M. Tekalp, and M. I. Sezan. Simultaneous motion estimation and segmentation. *IEEE Trans. on Image Processing*, 6(9):1326–1333, September 1997.

[CY88] L. O. Chua and L. Yang. Cellular neural networks: Theory and applications. *IEEE Trans. on Circuits and Systems*, 35(10):1257–1290, October 1988.

[Dau88] I. Daubechies. Orthonormal basis of compactly supported wavelets. *Comm. Pure Applied Math.*, 41:909–996, 1988.

[Dau92] I. Daubechies. *Ten lectures on wavelets*, volume 61 of *CBMS-NSF regional conference series in Applied Mathematics.* SIAM, Philadelphia, Pennsylvania, 1992.

[DBG91] J. P. Dérutin, B. Besserer, and J. Gallice. A parallel vision machine : Transvision. In B. Zavidovique and P.L. Wendel, editors, *Proc. of Computer Architecture for Machine Perception*, pages 241–251, Dec. 1991.

[Del93] T. Delbrück. Silicon retina with correlation-based, velocity-tuned pixels. *IEEE Trans. on Neural Networks*, 4(3):529–541, May 1993.

[Dja94] K. Djafarian. Etude d'une architecture systolique pour la détection de mouvement. Rapports Ingénieur ENSERG et DEA, Laboratoire de microélectronique, Université Catholique de Louvain, Belgique, Septembre 1994.

[DLC96a] C. Dumontier, F. Luthon, and J. P. Charras. Implantation sur carte à DSP d'un algorithme de détection de mouvement par approche markovienne. In *3e Journées Adéquation Algorithme Architecture en Traitement du Signal et des Images (AAA'96)*, pages 255–262, CNES, Toulouse, 17-19 Janvier 1996.

[DLC96b] C. Dumontier, F. Luthon, and J. P. Charras. Real time implementation of an MRF-based motion detection algorithm on a DSP board. In *7th IEEE Digital Signal Processing Workshop (DSPW'96)*, pages 183–186, Loen, Norway, September 1-4, 1996.

[DLC97] C. Dumontier, F. Luthon, and J. P. Charras. Une architecture alternative pour la mise en oeuvre temps réel d'un algorithme markovien de détection de mouvement . In *16e Colloque GRETSI*, pages 1193–1196, Grenoble, 15-19 Septembre 1997.

[DLC98] C. Dumontier, F. Luthon, and J. P. Charras. Architecture pour le prototypage rapide d'applications embarquées de vision. In *11e Congrès Reconn. Formes Intell. Artific. (RFIA'98)*, pages 337–345, Clermont-Ferrand, 20-22 Janvier 1998.

[DLC99] C. Dumontier, F. Luthon, and J. P. Charras. Real-time DSP implementation for MRF-based video motion detection. *IEEE Trans. on Image Processing*, 8(10), October 1999.

[DLR77] M. M. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39:1–38, 1977.

[DP98] G. Deng and J.-C. Pinoli. Differentiation-based edge detection using the logarithmic image processing model. *Journal of Mathematical Imaging and Vision*, 8:161–180, 1998.

[Dra96] D. Dragomirescu. Etude de l'implantation sur circuit intégré analogique d'un algorithme de détection de mouvement dans les séquences d'images suivant un modèle markovien. Stage de fin d'études UPB, LTIRF/INPG, Grenoble, Août 1996.

[Dum96] C. Dumontier. *Etude et mise en œuvre temps réel d'un algorithme de détection de mouvement par approche markovienne.* PhD thesis, Institut National Polytechnique, Grenoble, 15 novembre 1996.

[Dum97] A. Dumitrescu. Conception d'un didacticiel destiné à un enseignement de travaux pratiques pour élèves-ingénieurs. Stage de fin d'études UPB, TP2A/ENSERG/INPG, Grenoble, Août 1997.

[ECdSM97] R. Etienne-Cummings, J. Van der Spiegel, and P. Mueller. A focal plane visual motion measurement sensor. *IEEE Trans. on Circuits and Systems-I*, 44(1):55–66, January 1997.

[ElA98] A. ElArmouzi. Contribution à la conception d'un circuit intégré analogique pour la détection de mouvement. DEA Microélectronique, LIS/INPG, Grenoble, Juin 1998.

[ERVDC+94] S. Espejo, A. Rodriguez-Vazquez, R. Dominguez-Castro, J. L. Huertas, and E. Sanchez-Sinencio. Smart-pixel cellular neural networks in analog current-mode CMOS technology. *IEEE Journal of Solid-State Circuits*, 29(8):895–905, August 1994.

[Fau92] E. Faure. Amélioration de la résolution spatiale du logiciel de reconstruction 3D en géométrie conique TOMOCONIC. Rapport Ingénieur ENSERG, LETI/CENG/CEA, Grenoble, Juin 1992.

[Fra94] O. Franchet. Calcul de flot optique par une approche différentielle. Stage élève-ingénieur ENSERG, LTIRF/INPG, Grenoble, Septembre 1994.

[Fri95] B. Frichot. Data compression of remote sensing instrument. Data compression with wavelets. Rapport Ingénieur ENSERG, Daimler-Benz Aerospace, Munich, Allemagne, Septembre 1995.

[GDG94] W. Guggenbühl, J. Di, and J. Goette. Switched-current memory circuits for high-precision applications. *IEEE Journal of Solid-State Circuits*, 29(9):1108–1116, September 1994.

[GG84] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 6(6):721–741, November 1984.

[Gje98] E. Gjelsvik. Implementation of image segmentation algorithms on a simulator for a programmable video processor. Rapport Ingénieur ENSERG, LIS/INPG, Grenoble, Juin 1998.

[Gua95] J. Guano. Qualité image et erreurs de transmission. Rapport Ingénieur ENSERG, Aérospatiale, Cannes La Bocca, Juin 1995.

[HBB+92] T. Horiuchi, W. Bair, B. Bishofberger, A. Moore, C. Koch, and J. Lazzaro. Computing motion using analog VLSI vision chips: An experimental comparison among different approaches. *International Journal of Computer Vision*, 8(3):203–216, 1992.

[HBM89] J. B. Hughes, N. C. Bird, and I. C. Macbeth. Switched-currents. A new technique for analog sampled-data signal processing. In *Proc. IEEE Int. Symp. Circuits and Systems*, pages 1584–1587, Portland, May 1989.

[Hee87] D. J. Heeger. Model for the extraction of image flow. *J. Opt. Soc. Am. A*, 4(8):1455–1471, August 1987.

[Hen93] C. Hennebert. Etude de plusieurs approches multirésolutions associées à un modèle markovien pour la détection du mouvement dans les séquences d'images. DEA SIP, LTIRF/INPG, Grenoble, Juillet 1993.

[HKLM88] J. Hutchinson, C. Koch, J. Luo, and C. Mead. Computing motion using analog and binary resistive networks. *Computer*, 21:52–63, March 1988.

[HL93] J. M. Hakkarainen and H. S. Lee. A 40×40 CCD/CMOS absolute-value-of-difference processor for use in a stereo vision system. *IEEE Journal of Solid-State Circuits*, 28(7):799–807, July 1993.

[HM93a] J. B. Hughes and K. W. Moulding. S^2I: A switched-current technique for high performance. *Electronics Letters*, 29(16):1400–1401, August 1993.

[HM93b] J. B. Hughes and K. W. Moulding. Switched-current signal processing for video frequencies and beyond. *IEEE Journal of Solid-State Circuits*, 28(3):314–322, March 1993.

[HMP90] J. B. Hughes, I. C. Macbeth, and D. M. Pattullo. Switched-current system cells. In *Proc. IEEE Int. Symp. Circuits and Systems*, pages 303–306, May 1990.

[HNR84] Y. Z. Hsu, H. H. Nagel, and G. Reckers. New likelihood test methods for change detection in image sequences. *Computer Vision, Graphics, and Image Processing*, 26:73–106, 1984.

[HS81] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[Jea95] R. Jeandel. Implémentation du standard de décompression audio DOLBY AC-3 sur le codeur de DSP D-950. Rapport Ingénieur ENSERG, SGS-Thomson Microelectronics, Grenoble, Septembre 1995.

[JP95] M. Jourlin and J.C. Pinoli. Image dynamic range enhancement and stabilization in the context of the logarithmic image processing model. *Signal Processing*, 41:225–237, 1995.

[Jul97] A. Jullian. Reconnaissance d'images simples par modhles de Markov cachés et programmation dynamique 2D. Rapports Ingénieur ENSERG et DEA, DGA/DCE/ETC5, Arcueil, Juin 1997.

[KGV83] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, May 1983.

[KLC97] T. H. Kuo, S. Y. Lee, and K. D. Chen. Offset current cancellation based on a multiple-path feedback compensation (MPFC) technique for switched-current circuits and systems. *IEEE Trans. on Circuits and Systems-II*, 44(4):299–309, April 1997.

[Kle95] G. Kleine. La télévision numérique avec compression de données MPEG-2. *Elektor*, 207:42–48, Septembre 1995.

[KMY86] C. Koch, J. Marroquin, and A. Yuille. Analog "neuronal" networks in early vision. In *Proc. Nat. Acad. Sci. USA, Biophysics 83*, pages 4263–4267, 1986.

[KN95] A. Kurianski and M. Nieniewski. A model of the MRF with three observation sources for obtaining the masks of moving objects. In *Proc. of 9th Scandinavian Conf. on Image Analysis*, pages 931–940, Uppsala, Sweden, June 1995.

[KS89] I.L. Kantor and A.S. Solodovnikov, editors. *Hypercomplex numbers. An elementary introduction to algebras.* Springer-Verlag, New-York, 1989.

[KS95] P. Kinget and M. S. J. Steyaert. A programmable analog cellular neural network CMOS chip for high speed image processing. *IEEE Journal of Solid-State Circuits*, 30(3):235–243, March 1995.

[KSK95] J. Kramer, R. Sarpeshkar, and C. Koch. An analog VLSI velocity sensor. In *Proc. IEEE Symposium on Circuits and Systems*, pages 413–416, Seattle, May 1995.

[Lab98] G. Laborde. Implémentation d'algorithmes de traitement d'images sur processeur vidéo spécialisé et démonstration. Rapport Ingénieur ENSERG, CNS/CNET, Meylan, Juin 1998.

[LAC97] J.L. Lacoume, P.O. Amblard, and P. Common. *Statistiques d'ordre supirieur pour le traitement du signal.* Traitement du Signal. Masson, Paris, 1997.

[Lal90] P. Lalande. *Détection du mouvement dans une séquence d'images selon une approche markovienne. Application à la robotique mobile.* PhD thesis, Université de Rennes I, 19 mars 1990.

[LB90] P. Lalande and P. Bouthémy. A statistical approach to the detection and tracking of moving objects in an image sequence. In *Proc. of Eusipco*, pages 947–950, Barcelona, Spain, May 1990.

[LBDA89] F. Luthon, R. Blanpain, M. Decorps, and J. P. Albrand. Parametric spectrum analysis of 2D NMR signals. Application to in vivo J spectroscopy. *Journal of Magnetic Resonance*, 81:538–551, 1989.

[LC93] F. Luthon and A. Caplier. Motion detection and segmentation in image sequences using Markov Random Field modelling. In *4th Eurographics Animation and Simulation Workshop (EG'93)*, pages 265–275, Barcelona, Spain, September 4-5, 1993.

[LCL99] F. Luthon, A. Caplier, and M. Liévin. Spatiotemporal MRF approach to video segmentation: application to motion detection and lip segmentation. *Signal Processing*, 76(1):61–80, July 1999.

[LD99] F. Luthon and D. Dragomirescu. A cellular analog network for MRF-based video motion detection. *IEEE Trans. on Circuits and Systems-I*, 46(2):281–293, February 1999.

[LDC+99] M. Liévin, P. Delmas, P.Y. Coulon, F. Luthon, and V. Fristot. Automatic lip tracking: Bayesian segmentation and active contours in a cooperative scheme. In *IEEE Int. Conf. on Multimedia Computing and Systems (ICMCS'99)*, Florence, Italy, June 7-11, 1999.

[Lie96a] M. Lievin. Détection du mouvement des lèvres par analyse dynamique de séquences d'images. DEA SIP, LTIRF/INPG, Grenoble, Sept 1996.

[Lie96b] M. Lievin. Détection du mouvement des lèvres par analyse dynamique de séquences d'images. Analyse et recherche de paramètres dynamiques caractéristiques sur lèvres non maquillées. Rapport Ingénieur ENSERG, LTIRF/INPG, Grenoble, Juin 1996.

[Lie00] M. Lievin. *Analyse automatique du mouvement des lèvres d'un locuteur.* PhD thesis, National Polytechnic Institute, Grenoble, France, October 2000.

[LL97a] M. Liévin and F. Luthon. Analyse automatique du mouvement des lèvres d'un locuteur. Rencontres ELESA, ENSIEG, Grenoble, 14 Novembre 1997.

[LL97b] F. Luthon and M. Liévin. Lip motion automatic detection. In *10th Scandinavian Conference on Image Analysis (SCIA'97)*, pages 253–260, Lappeenranta, Finland, June 9-11, 1997.

[LL98a] M. Liévin and F. Luthon. Détection du mouvement des lèvres. In *4e Journées Compression et Représentation des Signaux Audiovisuels (CORESA'98)*, pages 151–158, CNET, Lannion, 9-10 Juin 1998.

[LL98b] M. Liévin and F. Luthon. Lip features automatic extraction. In *IEEE Int. Conf. on Image Processing (ICIP'98)*, pages 168–172, Chicago, Illinois, October 4-7, 1998. Vol. 3.

[LL99] M. Liévin and F. Luthon. Unsupervised lip segmentation under natural conditions. In *IEEE Int. Conf. on Acoustic, Speech and Signal Processing (ICASSP'99)*, pages 3065–3068, Phoenix, Arizona, March 15-19, 1999. Vol. 6.

[LP95] F. Luthon and S. Popescu. Un circuit intégré analogique pour la détection de mouvement. In *Journées sur les nouvelles techniques pour la compression et la représentation des signaux audiovisuels (CORESA'95)*, pages 105–112, CCETT, Rennes, 12-13 Janvier 1995.

[LPC94a] F. Luthon, G. V. Popescu, and A. Caplier. Analog resistive networks for motion detection. In V. I. Vlad, editor, *4th Conference in Optics (ROMOPTO'94)*, pages 299–301, Bucharest, Romania, September 5-8, 1994. Proc. SPIE 2461.

[LPC94b] F. Luthon, G. V. Popescu, and A. Caplier. An MRF-based motion detection algorithm implemented on analog resistive network. In J. O. Eklundh, editor, *3rd European Conference on Computer Vision (ECCV'94)*, pages 167–174, Stockholm, Sweden, May 2-6, 1994. Springer-Verlag. LNCS Vol. 800.

[LS96] F. Luthon and S. Simon. Transformée en ondelettes et compensation de mouvement dans les séquences d'images. In *Journées Compression et Représentation des Signaux Audiovisuels (CORESA'96)*, pages 2–9, CNET, Meylan, 15-16 Février 1996.

[LSFC93] J. C. Lee, B. J. Sheu, W. C. Fang, and R. Chellappa. VLSI neuroprocessors for video motion detection. *IEEE Trans. on Neural Networks*, 4(2):178–191, March 1993.

[Lut85] F. Luthon. Analyseur vectoriel de réseaux microondes - Méthodes générales de calibrage gérées par calculateur. DEA Systèmes Electroniques, ENSERG, Grenoble, Septembre 1985. 70 pages.

[Lut88a] F. Luthon. *Filtrage et analyse spectrale paramétriques. Applications en spectroscopie et magnétométrie RMN. Eléments de synthèse pour un analyseur paramétrique.* PhD thesis, INPG, Grenoble, Septembre 1988. 298 pages.

[Lut88b] F. Luthon. *Filtrage et analyse spectrale paramétriques. Notice d'utilisation des logiciels.* L.E.T.I., Grenoble, Novembre 1988. Rapport technique, 116 pages.

[Lut89a] F. Luthon. IT4 : le programme national suédois de développement industriel en technologies de l'information. Rapport d'Ambassade no. SC 6 ST/I, Ambassade de France, Stockholm, Suède, Juin 1989. 47 pages.

[Lut89b] F. Luthon. Le programme spatial suédois. Rapport d'Ambassade no. SC 8 ST/I, Ambassade de France, Stockholm, Suède, Décembre 1989. 47 pages.

[Lut90a] F. Luthon. Filtrage et analyse spectrale paramétriques, Exemples d'applications. In *Traitement du Signal*, chapter 7. INSTN, CEA-CENG, Grenoble, 1990. Document de cours.

[Lut90b] F. Luthon. Le centre de recherche nucléaire de Studsvik. Rapport d'Ambassade no. SC 1 ST/I, Ambassade de France, Stockholm, Suède, Mai 1990. 50 pages.

[Lut93] F. Luthon. Implémentation de la segmentation markovienne du mouvement dans des séquences d'images. Journées thématiques du GDR 134, GT2, Champs de Markov - Segmentation, Telecom Paris, 14 Octobre 1993.

[Lut96] F. Luthon. Mathématique de l'ingénieur. Reprographie ENSERG, Grenoble, 1996. Document de cours, 100 pages.

[Lut97] F. Luthon. Détection de mouvement dans les séquences d'images. Mises en oeuvre temps réel. In A. Chéhikian, P. Y. Coulon, and F. Luthon, editors, *Ecole des Techniques Avancées en Signal-Image-Parole (ETASIP'97)*, pages 181–201. INPG, Grenoble, 9-12 Septembre 1997.

[Mal89] S. G. Mallat. Multifrequency channel decompositions of images and wavelet models. *IEEE Trans. on ASSP*, 37(12):2091–2110, December 1989.

[MAL91] C. A. Mead, X. Arreguit, and J. Lazzaro. Analog VLSI model of binaural hearing. *IEEE Trans. on Neural Networks*, 2(2):230–236, mar 1991.

[MASJ93] R. C. Meitzler, A. G. Andreou, K. Strohbehn, and R. E. Jenkins. A sampled-data motion chip. In *36th Midwest Symp. Circuits and Systems, Vol.1*, pages 288–291, Detroit, 1993.

[MB87] D. W. Murray and B. F. Buxton. Scene segmentation from visual motion using global optimization. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 9(2):220–228, March 1987.

[MB96] A. Mitiche and P. Bouthémy. Computation and analysis of image motion: A synopsis of current problems and methods. *International Journal of Computer Vision*, 19(1):29–55, 1996.

[Mea89] C. A. Mead. *Analog VLSI and Neural Systems*. Addison-Wesley, Reading, MA, 1989.

[MM88] C. A. Mead and M. A. Mahowald. A silicon model of early visual processing. *Neural Networks*, 1:91–97, 1988.

[Mon96a] M. Montrichard. Compression vidéo type MPEG et quantification vectorielle. Rapport Ingénieur ENSERG, IRISA/INRIA, Rennes, Juin 1996.

[Mon96b] M. Montrichard. Compression vidéo type MPEG et quantification vectorielle algébrique et arborescente : allocation binaire et analyse du codage. Rapport de DEA, IRISA/INRIA, Rennes, Septembre 1996. Option SIP.

[MPJ96] F. Mayet, J.C. Pinoli, and M. Jourlin. Justifications physiques et applications du modèle LIP pour le traitement des images obtenues en lumière transmise. *Traitement du Signal*, 13(3):245–256, 1996.

[Nag87] H. H. Nagel. On the estimation of optical flow: relations between different approaches and some new results. *Artificial Intelligence*, 33:299–324, 1987.

[NB96] D. G. Nairn and A. Biman. A comparative analysis of switched-current circuits. *IEEE Trans. on Circuits and Systems-II*, 43(11):733–743, November 1996.

[NE86] H. H. Nagel and W. Enkelmann. An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8(5):565–593, September 1986.

[NR79] A. N. Netravali and J. D. Robbins. Motion-compensated television coding: Part 1. *Bell System Technical Journal*, 58(3):631–670, March 1979.

[NVGS95] J. Nedved, J. Vanneuville, D. Gevaert, and J. Sevenhans. A transistor-only switched current sigma-delta A/D converter for a CMOS speech CODEC. *IEEE Journal of Solid-State Circuits*, 30(7):819–822, July 1995.

[OB94] J. M. Odobez and P. Bouthémy. Detection of multiple moving objects using multiscale MRF with camera compensation. In *Proc. 1st IEEE Int. Conf. on Image Processing*, pages 257–261, Austin, 1994.

[OB95] J. M. Odobez and P. Bouthémy. Robust multiresolution estimation of parametric motion models. *Journal of Visual Communication and Image Representation*, 6(4):348–365, 1995.

[Odo94] J. M. Odobez. *Estimation, ditection et segmentation du mouvement: une approche robuste markovienne.* PhD thesis, Universiti de Rennes I, December 1994.

[PB95] H. Palus and D. Bereska. The comparison between transformations from RGB colour space to IHS colour space, used for object recognition. In *5th IEE International Conference on Image Processing and its Applications*, pages 825–827, Heriot-Watt University, Edinburgh, UK, July 1995.

[PC92] A. Pascal-Caplier. Détection d'objets en mouvement dans une séquence d'images selon une approche markovienne. DEA SIP, LTIRF/INPG, Grenoble, Juillet 1992.

[PHS95] A. J. Pritchard, R. E. N. Horne, and S. J. Sangwine. Achieving brigthness-insensitive measurements of colour saturation for use in object recognition. In *5th IEE International Conference on Image Processing and its Applications*, pages 791–795, Heriot-Watt University, Edinburgh, UK, July 1995.

[Pie94] W. Pieczynski. Champs de Markov cachés et estimation conditionnelle itérative. *Traitement du Signal*, 11(2):141–153, 1994.

[Pin97] J. C. Pinoli. The logarithmic image processing model: Connections with human brightness perception and contrast estimators. *Journal of Mathematical Imaging and Vision*, 7:341–358, 1997.

[PL95] S. Popescu and F. Luthon. The design of an efficient VLSI circuit for motion detection at high image repetition rate. In *2nd Advanced Training Course: Mixed Design of VLSI Circuits (MixVLSI'95)*, pages 361–366, Krakow, Poland, May 29-31, 1995.

[Poi68] H. Poincaré. *La science et l'hypothhse.* Champs. Flammarion, Paris, 1968.

[Pop93] V. G. Popescu. Etude de l'implantation sur circuit intégré analogique d'un algorithme de détection de mouvement dans les séquences d'images suivant un modèle markovien. Stage de fin d'études UPB, LTIRF/INPG, Grenoble, Août 1993.

[Pop94] S. Popescu. Etude de l'implantation sur circuit intégré analogique d'un algorithme de détection de mouvement dans les séquences d'images. Stage de fin d'études UPB, LTIRF/INPG, Grenoble, July 1994.

[PTK85] T. Poggio, V. Torre, and C. Koch. Computational vision and regularization theory. *Nature*, 317:314–319, September 1985.

[Riq96] D. Riquet. Etude de modèles de mouvements dans une séquence d'images dans le cas d'une caméra fixe ou mobile. Stage élève-ingénieur ENSERG, LTIRF/INPG, Grenoble, Septembre 1996.

[Ros84] A. Rosenfeld, editor. *Multiresolution Image Processing and Analysis*. Springer-Verlag, New-York, 1984.

[Run92] Y. Rungsunseri. *Systhme temps réel dédié à la description d'image par segments de droite*. PhD thesis, Institut National Polytechnique, Grenoble, 24 novembre 1992.

[RVDCMDR95] A. Rodriguez-Vazquez, R. Dominguez-Castro, F. Medeiro, and M. Delgado-Restituto. High resolution CMOS current comparators: Design and applications to current-mode function generation. *Analog Integrated Circuits and Signal Processing*, 7:149–165, March 1995.

[SBK93] R. Sarpeshkar, W. Bair, and C. Koch. Visual motion computation in analog VLSI using pulses. In D. S. Touretzky, editor, *Advances in Neural Information Processing Systems 5*, pages 781–788, San Monteo, 1993.

[SE98] S. J. Sangwine and T.A. Ell. The discrete Fourier transform of a colour image. In *2nd IMA Conference on Image Processing Mathematical Methods, Algorithms and Applications*, De Montfort University, Leicester, UK, September 1998.

[SH96] D. G. Stork and M. E. Hennecke, editors. *Speechreading by Humans and Machines*, volume 150 of *NATO ASI Series F: Computer and Systems Sciences*. Springer-Verlag, Berlin, 1996.

[Sic99] G. Sicard. *De la biologie au silicium : Une rétine bio-inspirée analogique pour un capteur de vision "intelligent" adaptatif*. PhD thesis, Institut National Polytechnique, Grenoble, 20 janvier 1999.

[Sil96] P. L. Silsbee. Motion in deformable templates. In *Speechreading by Humans and Machines, D. G. Stork and M. E.Hennecke Ed.*, pages 489–496, NATO ASI Series F: Computer and Systems Sciences, Vol.150, Springer-Verlag,Berlin, 1996.

[Sim95] S. Simon. Mise en œuvre de la transformation en ondelettes pour l'estimation du mouvement dans les séquences d'images. DEA SIP, LTIRF/INPG, Grenoble, Juillet 1995.

[SLK93] M. Song, Y. Lee, and W. Kim. A clock feedthrough reduction circuit for switched-current systems. *IEEE Journal of Solid-State Circuits*, 28(2):133–137, February 1993.

[Sou98] C. M'Barek Soussi. Développement d'applications interactives en télévision numérique. Réalisation du module TELETEL en OpenTV 1.1. Rapport Ingénieur ENSERG, T.P.S., Issy-les-Moulineaux, Juin 1998.

[Spi98] A. Spinei. *Estimation du mouvement par triades de filtres de Gabor. Application au mouvement d'objets transparents*. PhD thesis, Institut National Polytechnique de Grenoble, 14 octobre 1998.

[STM+95] A. Simoni, G. Torelli, F. Maloberti, A. Sartori, S. E. Plevridis, and A. N. Birbas. A single-chip optical sensor with analog memory for motion detection. *IEEE Journal of Solid-State Circuits*, 30(7):800–806, July 1995.

[TL94] G. Tziritas and C. Labit. *Motion Analysis for Image Sequence Coding*, volume 4 of *Advances in Image Communication.* Elsevier Science B. V., Amsterdam, The Netherlands, 1994.

[WR84] D. R. Walker and K. R. Rao. Improved pel-recursive motion compensation. *IEEE Trans. on Communications*, 32(10):1128–1134, October 1984.

[Yao94] Q. Yao. Détermination du flot optique dans une séquence d'images par une approche fréquentielle. DEA SIP, LTIRF/INPG, Grenoble, Juillet 1994.

[ZC93] J. Zerubia and R. Chellappa. Mean field annealing using compound Gauss-Markov random fields for edge detection and image estimation. *IEEE Trans. on Neural Networks*, 4(4):703–709, July 1993.

Appendix A

Curriculum Vitae

Nom, Prénom :	LUTHON, Franck
Date, Lieu de naissance :	13 janvier 1963 à Rambouillet, France
Situation de famille :	marié, 2 enfants
Grade :	Maître de Conférences 1ère classe, Section 61
Enseignement :	Ecole Nat. Sup. d'Electronique et Radioélectricité, Grenoble (ENSERG)
Recherche :	Laboratoire des Images et des Signaux (LIS/INPG)

Formation Universitaire et Diplômes

1980	Baccalauréat C, *mention Bien*, Lycée de Rambouillet.
1980-82	Maths Sup, Maths Spé P', Lycée Chaptal, Paris.
1984	First Certificate in English, délivré par l'Université de Cambridge.
1985	Certificate of Proficiency in English, délivré par l'Université de Cambridge.
1985	Diplôme d'ingénieur électronicien, *mention Très Bien*, ENSERG.
1985	DEA d'électronique, *mention Bien*, ENSERG.
1988	Doctorat de l'INPG, spécialité Systèmes électroniques, *mention Très Honorable.* *Lauréat du Prix de Thèse de l'INPG*
1989	Certificats de suédois, niveaux 1 et 2, Ecole pour étrangers, Stockholm
1991	Licence de philosophie, Université Pierre Mendès-France, Grenoble.
Langues	Anglais, Suédois : couramment Allemand, Italien : bonnes notions

Cursus Professionnel

1985-88	Contrat de Formation par la Recherche (CFR), LETI/CENG/CEA, Grenoble
1985-88	Vacations de Maths (Formation Continue INPG et CUEFA/CNAM), Grenoble
1988-90	Attaché Scientifique à l'Ambassade de France, chargé de l'électronique. Service National de Coopération, Stockholm, Suède (18 mois).
Depuis 1990	Maître de Conférences à l'ENSERG, affecté au laboratoire TIRF (LIS).
Depuis 1992	Responsable de la plateforme de Travaux Pratiques de 2ème année à l'ENSERG.
1988-94	Conférencier en Traitement de Signal : stages bisannuels de formation professionnelle organisés par l'INSTN/CENG/CEA.
1989-94	Expert du Comité de Lecture de la revue Traitement du Signal.
Depuis 1995	Membre du Comité de Rédaction de la revue Traitement du Signal.
Depuis 1995	Expert du Comité de Lecture du journal Signal Processing.

Appendix B

Bilan des Encadrements

3 Thèses

J'ai obtenu l'agrément doctoral pour être Directeur de trois thèses SIP (encadrement à 100%) :

- A. Caplier :
  Modèles markoviens de détection de mouvement dans les séquences d'images ;
  Thèse INPG soutenue en décembre 1995 [Cap95].
  Jury : J.L. Lacoume, P.Bouthémy, J.P. Dérutin, G. Privat, F.Luthon.
  Statut actuel : MCF à l'ENSERG.

- C. Dumontier :
  Mise en œuvre temps réel sur DSP de la détection de mouvement;
  Thèse INPG soutenue en novembre 1996 [Dum96].
  Jury : J.M. Dolmazon, M. Paindavoine, D. Barba, F. Heitz, J.P. Charras, F.Luthon.
  Statut actuel : salarié chez Intermec ST (Labège, 31676).

- M. Liévin :
  Segmentation spatio-temporelle de séquences vidéo.
  Application à l'analyse automatique du mouvement des lèvres d'un locuteur ;
  Soutenance prévue en octobre 2000 [Lie00].
  Statut actuel : boursier MENRT.

8 DEA

- 6 DEA de l'option Signal-Image-Parole que j'ai encadrés seul :
  sur la détection de mouvement par MRF [PC92], la multirésolution pyramidale [Hen93] et les ondelettes [Sim95, Chr96], l'approche fréquentielle de calcul du flux optique [Yao94] et le labiophone [Lie96a].

- 2 DEA de Microélectronique co-encadrés :
  l'un sur la conception d'un circuit VLSI de détection de mouvement (avec Gérard Bouvier du LIS) [ElA98], l'autre sur l'analyse statistique (ordres supérieurs) du bruit dans les composants (avec Jean Brini du LPCS) [Bon98].

11 Stages Ingénieurs

J'ai encadré 4 stages de fin d'études et de recherche (6 mois) d'étudiants de l'UPB (Université Polytechnique de Bucarest) :

- 3 stages sur la conception et la simulation d'un circuit intégré VLSI analogique pour la détection de mouvement par MRF [Pop93, Pop94, Dra96],
- un stage sur l'instrumentation programmée et la conception de didacticiels pour Travaux Pratiques d'électronique (utilisation de Labwindows, LabView et ToolBook) [Dum97].

J'ai également encadré 4 stages de fin d'études d'ingénieurs INPG autour du projet Labiophone [Lie96b, Asf97] et de la mise en œuvre des algorithmes sur un processeur vidéo (PVP) [Gje98, Lab98].

Enfin, j'ai encadré 3 stages d'élèves-ingénieurs de 2ème année de l'ENSERG. Ces stages courts (2 mois) sont l'occasion de programmer des méthodes de traitement proposées dans la littérature, en vue de les comparer à nos propres algorithmes : calcul du flux optique par méthode différentielle [Fra94] ou par mise en correspondance hiérarchique [Bou96], compensation et segmentation de mouvement dans le cas d'une caméra mobile [Riq96].

9 Tutorats de stage

Toujours sur des thèmes proches de mon activité de recherche, ces tutorats de stages de fin d'études sont l'occasion de nouer des contacts et de se tenir au courant des préoccupations du monde industriel, en plus de l'occasion de promouvoir l'ENSERG :

- télévision numérique et processeurs vidéo : TPS [Sou98], TCEC [BC98]
- compression, codage, MPEG, ondelettes : IRISA/INRIA [Mon96a, Mon96b], Daimler-Benz Aerospace [Fri95], Aérospatiale [Gua95], SGS-Thomson [Jea95]
- segmentation par MRF : DGE/ETCA [Jul97]
- architecture systolique VLSI pour la vision : UCL Belgique [Dja94]
- reconstruction 3D en tomographie : LETI/CEA [Fau92]

Liste des Travaux Encadrés

[Asf97] A. Asfour. Analyse automatique du mouvement des lèvres d'un locuteur. Rapport Ingénieur ENSPG, LTIRF/INPG, Grenoble, Juin 1997.

[BC98] S. Berthout and G. Clédat. Etude et réalisation d'une interface analogique/numérique et des circuits de contr le mémoire et d'arbitrage de bus d'un processeur vidéo. Rapport Ingénieur ENSERG, TCEC, Meylan, Juin 1998.

[Bon98] D. Bonifaci. Mise en œuvre de traitements statistiques sur des signaux de bruit. DEA Microélectronique, LPCS/INPG, Grenoble, Juin 1998.

[Bou96] X. Boudet. Détermination du flot optique dans une séquence d'images par mise en correspondance hiérarchique. Stage élève-ingénieur ENSERG, LTIRF/INPG, Grenoble, Septembre 1996.

[Cap95] A. Caplier. *Modèles markoviens de détection de mouvement dans les séquences d'images: Approche spatio-temporelle et mises en œuvre temps réel.* PhD thesis, Institut National Polytechnique, Grenoble, 20 décembre 1995.

[Chr96] A. Christoforou. Estimation de mouvement et transformée en ondelettes pour la compression de séquences d'images. DEA SIP, LTIRF/INPG, Grenoble, Juillet 1996.

[Dja94] K. Djafarian. Etude d'une architecture systolique pour la détection de mouvement. Rapports Ingénieur ENSERG et DEA, Laboratoire de microélectronique, Université Catholique de Louvain, Belgique, Septembre 1994.

[Dra96] D. Dragomirescu. Etude de l'implantation sur circuit intégré analogique d'un algorithme de détection de mouvement dans les séquences d'images suivant un modèle markovien. Stage de fin d'études UPB, LTIRF/INPG, Grenoble, Août 1996.

[Dum96] C. Dumontier. *Etude et mise en œuvre temps réel d'un algorithme de détection de mouvement par approche markovienne.* PhD thesis, Institut National Polytechnique, Grenoble, 15 novembre 1996.

[Dum97] A. Dumitrescu. Conception d'un didacticiel destiné à un enseignement de travaux pratiques pour élèves-ingénieurs. Stage de fin d'études UPB, TP2A/ENSERG/INPG, Grenoble, Août 1997.

[ElA98] A. ElArmouzi. Contribution à la conception d'un circuit intégré analogique pour la détection de mouvement. DEA Microélectronique, LIS/INPG, Grenoble, Juin 1998.

[Fau92] E. Faure. Amélioration de la résolution spatiale du logiciel de reconstruction 3D en géométrie conique TOMOCONIC. Rapport Ingénieur ENSERG, LETI/CENG/CEA, Grenoble, Juin 1992.

[Fra94] O. Franchet. Calcul de flot optique par une approche différentielle. Stage élève-ingénieur ENSERG, LTIRF/INPG, Grenoble, Septembre 1994.

[Fri95] B. Frichot. Data compression of remote sensing instrument. Data compression with wavelets. Rapport Ingénieur ENSERG, Daimler-Benz Aerospace, Munich, Allemagne, Septembre 1995.

[Gje98] E. Gjelsvik. Implementation of image segmentation algorithms on a simulator for a programmable video processor. Rapport Ingénieur ENSERG, LIS/INPG, Grenoble, Juin 1998.

[Gua95] J. Guano. Qualité image et erreurs de transmission. Rapport Ingénieur ENSERG, Aérospatiale, Cannes La Bocca, Juin 1995.

[Hen93] C. Hennebert. Etude de plusieurs approches multirésolutions associées à un modèle markovien pour la détection du mouvement dans les séquences d'images. DEA SIP, LTIRF/INPG, Grenoble, Juillet 1993.

[Jea95] R. Jeandel. Implémentation du standard de décompression audio DOLBY AC-3 sur le codeur de DSP D-950. Rapport Ingénieur ENSERG, SGS-Thomson Microelectronics, Grenoble, Septembre 1995.

[Jul97] A. Jullian. Reconnaissance d'images simples par mod les de Markov cachés et programmation dynamique 2D. Rapports Ingénieur ENSERG et DEA, DGA/DCE/ETC5, Arcueil, Juin 1997.

[Lab98] G. Laborde. Implémentation d'algorithmes de traitement d'images sur processeur vidéo spécialisé et démonstration. Rapport Ingénieur ENSERG, CNS/CNET, Meylan, Juin 1998.

[Lie96a] M. Lievin. Détection du mouvement des lèvres par analyse dynamique de séquences d'images. DEA SIP, LTIRF/INPG, Grenoble, Sept 1996.

[Lie96b] M. Lievin. Détection du mouvement des lèvres par analyse dynamique de séquences d'images. Analyse et recherche de paramètres dynamiques caractéristiques sur lèvres non maquillées. Rapport Ingénieur ENSERG, LTIRF/INPG, Grenoble, Juin 1996.

[Lie00] M. Lievin. *Analyse automatique du mouvement des lèvres d'un locuteur.* PhD thesis, National Polytechnic Institute, Grenoble, France, October 2000.

[Mon96a] M. Montrichard. Compression vidéo type MPEG et quantification vectorielle. Rapport Ingénieur ENSERG, IRISA/INRIA, Rennes, Juin 1996.

[Mon96b] M. Montrichard. Compression vidéo type MPEG et quantification vectorielle algébrique et arborescente : allocation binaire et analyse du codage. Rapport de DEA, IRISA/INRIA, Rennes, Septembre 1996. Option SIP.

[PC92] A. Pascal-Caplier. Détection d'objets en mouvement dans une séquence d'images selon une approche markovienne. DEA SIP, LTIRF/INPG, Grenoble, Juillet 1992.

[Pop93] V. G. Popescu. Etude de l'implantation sur circuit intégré analogique d'un algorithme de détection de mouvement dans les séquences d'images suivant un modèle markovien. Stage de fin d'études UPB, LTIRF/INPG, Grenoble, Août 1993.

[Pop94] S. Popescu. Etude de l'implantation sur circuit intégré analogique d'un algorithme de détection de mouvement dans les séquences d'images. Stage de fin d'études UPB, LTIRF/INPG, Grenoble, July 1994.

[Riq96] D. Riquet. Etude de modèles de mouvements dans une séquence d'images dans le cas d'une caméra fixe ou mobile. Stage élève-ingénieur ENSERG, LTIRF/INPG, Grenoble, Septembre 1996.

[Sim95] S. Simon. Mise en œuvre de la transformation en ondelettes pour l'estimation du mouvement dans les séquences d'images. DEA SIP, LTIRF/INPG, Grenoble, Juillet 1995.

[Sou98] C. M'Barek Soussi. Développement d'applications interactives en télévision numérique. Réalisation du module TELETEL en OpenTV 1.1. Rapport Ingénieur ENSERG, T.P.S., Issy-les-Moulineaux, Juin 1998.

[Yao94] Q. Yao. Détermination du flot optique dans une séquence d'images par une approche fréquentielle. DEA SIP, LTIRF/INPG, Grenoble, Juillet 1994.

Bilan des Publications

Cette annexe récapitule mon activité en matière de publications. Elle est suivie de la liste intégrale de mes rapports et publications. Enfin, les six documents marqués d'une étoile (échantillons représentatifs de mon activité de recherche) sont annexés à la fin de ce document.

5 Articles de Revues Internationales

(parus ou acceptés)

- (*) Journal of Magnetic Resonance (1989) : analyse spectrale paramétrique (travail de thèse) [LBDA89]
- (*) IEEE Trans. on Circuits and Systems-I (1998) : circuit VLSI analogique de détection de mouvement [LD99]
- (*) IEEE Trans. on Image Processing (1999) : mise en œuvre sur DSP [DLC98b]
- (*) Signal Processing (1999) : segmentation spatio-temporelle par approche 3D [LCL99]
- (*) Real-Time Imaging (1998) : mises en œuvre matérielles de la détection de mouvement [CLD98]

2 Articles de Revue Nationale

- Traitement du Signal (1997) : approche spatio-temporelle 3D [CL97]
- Traitement du Signal (1996) : détection de mouvement et mises en œuvre matérielles [CDLC96]

10 Conférences ou Workshops Internationaux

(avec actes et comité de lecture)

- (*) IEEE Int. Conf. Image Processing (ICIP'98) : application Labiophone [LL98b]
- IEEE Int. Conf. ICASSP'99 : application labiophone [LL99]
- IEEE Digital Signal Processing Workshop (DSPW'96) : implantation sur DSP du détecteur de mouvement [DLC96b]
- IEE Int. Conf. Image Processing and Applications (ICIPA'95) [CL95d]

- Europ. Conf. Computer Vision (ECCV'94) [LPC94b]
- Scandinav. Conf. Image Analysis (SCIA'95, SCIA'97) [CL95b, LL97b]
- Int. Conf. Computer Analysis of Images and Patterns (CAIP'95) [CL95c]
- Eurographics Animation and Simulation Workshop (EG'93) [LC93]
- IEEE Int. Conf. Multimedia Computing and Systems ICMCS'99: suivi automatique de lèvres par coopération MRF/contours actifs (papier soumis [LDC+99]).

10 Conférences Nationales ou Restreintes

(avec actes et comité de lecture)

- Mediterranen Conf. Electronics Automatic Control (MCEA'95) [CL95a]
- SPIE Conf. in Optics (ROMOPTO'94) [LPC94a]
- 2nd Advanced Training Course: Mixed Design of VLSI Circuits (MixVLS'I95 [PL95])
- GRETSI (95, 97) [CLD95, DLC97]
- Journées du GDR Adéquation Algorithme Architecture (AAA'96) [DLC96a]
- Congrès Reconnaissance Formes Intelligence Artificielle (RFIA'98) [DLC98a]
- Journées Compression Représentation Signaux Audio Visuels (CORESA'95, CORESA'96, CORESA'98) [LP95, LS96, LL98a]

Divers

- Actes Ecole d 'été des Techniques Avancées Signal Image Parole (ETASIP'97) [CCL97]
- 3 Rapports Techniques d'Ambassade [Lut89a, Lut89b, Lut90b]
- 1 Rapport Technique LETI/CEA [Lut88b]
- 2 Séminaires du GDR [CL93, Lut93]
- 1 Poster aux Rencontres ELESA [LL97a]
- 3 Documents de cours (analyse spectrale, math, ETASIP'97) [Lut90a, Lut96, Lut97]
- 1 Mémoire de thèse [Lut88a]
- 1 Rapport de DEA [Lut85]

Liste de Publications

[CCL97] A. Chéhikian, P. Y. Coulon, and F. Luthon, editors. *Ecole des Techniques Avancées en Signal-Image-Parole (ETASIP'97)*, Grenoble, France, 9-12 Septembre 1997. LTIRF-INPG. Session 2.

[CDLC96] A. Caplier, C. Dumontier, F. Luthon, and P. Y. Coulon. Algorithme de détection de mouvement par modélisation markovienne. Mise en oeuvre sur DSP. *Traitement du Signal*, 13(2):177–190, 1996.

[CL93] A. Caplier and F. Luthon. Segmentation du mouvement dans une séquence d'images selon une approche markovienne. Journées thématiques du GDR 134, GT2, Champs de Markov - Segmentation, Telecom Paris, 14 Octobre 1993.

[CL95a] A. Caplier and F. Luthon. A 3D space and time MRF based algorithm for motion detection. In *Mediterranean Conf. on Electronics and Automatic Control (MCEA'95)*, pages 147–152, Grenoble, France, September 13-15, 1995.

[CL95b] A. Caplier and F. Luthon. An MRF-based spatiotemporal multiresolution algorithm for motion detection. In *Proc. 9th Scandinavian Conf. on Image Analysis (SCIA'95)*, pages 559–566, Uppsala, Sweden, June 6-9, 1995.

[CL95c] A. Caplier and F. Luthon. A new spatiotemporal approach for image analysis. Application to motion detection. In *Proc. of 6th International Conf. on Computer Analysis of Images and Patterns (CAIP'95)*, pages 246–253, Prague, Czech Republic, September 6-8, 1995.

[CL95d] A. Caplier and F. Luthon. Spatiotemporal multiresolution associated to MRF modelling for motion detection. In *Proc. 5th IEE International Conf. on Image Processing and its Applications (ICIPA'95)*, pages 158–162, Edinburgh, U.K., July 3-6, 1995.

[CL97] A. Caplier and F. Luthon. Approche spatio-temporelle pour l'analyse de séquences d'images. Application en détection de mouvement. *Traitement du Signal*, 14(2):195–208, 1997. Abstract in English.

[CLD95] A. Caplier, F. Luthon, and C. Dumontier. Algorithme markovien de détection de mouvement. Mises en oeuvre temps réel? In *15e Colloque GRETSI*, pages 1033–1036, Juan-les-Pins, France, 18-21 Septembre 1995.

[CLD98] A. Caplier, F. Luthon, and C. Dumontier. Real-time implementations of an MRF-based motion detection algorithm. *Real-Time Imaging*, 4(1):41–54, February 1998. Special Issue on Real-Time Motion Analysis.

[DLC96a] C. Dumontier, F. Luthon, and J. P. Charras. Implantation sur carte à DSP d'un algorithme de détection de mouvement par approche markovienne. In *3e Journées Adéquation Algorithme Architecture en Traitement du Signal et des Images (AAA'96)*, pages 255–262, CNES Toulouse, France, 17-19 Janvier 1996.

[DLC96b] C. Dumontier, F. Luthon, and J. P. Charras. Real time implementation of an MRF-based motion detection algorithm on a DSP board. In *Proc. 7th IEEE Digital Signal Processing Workshop (DSPW'96)*, pages 183–186, Loen, Norway, September 1-4, 1996.

[DLC97] C. Dumontier, F. Luthon, and J. P. Charras. Une architecture alternative pour la mise en oeuvre temps réel d'un algorithme markovien de détection de mouvement . In *Actes 16e Colloque GRETSI*, pages 1193–1196, Grenoble, France, 15-19 Septembre 1997.

[DLC98a] C. Dumontier, F. Luthon, and J. P. Charras. Architecture pour le prototypage rapide d'applications embarquées de vision. In *Actes 11e Congrès Reconn. Formes Intell. Artific. (RFIA'98)*, pages 337–345, Clermont-Ferrand, France, 20-22 Janvier 1998.

[DLC98b] C. Dumontier, F. Luthon, and J. P. Charras. Real-time implementation for MRF-based motion detection in image sequences. *IEEE Trans. on Image Processing*, 1998. Accepted for publication.

[LBDA89] F. Luthon, R. Blanpain, M. Decorps, and J. P. Albrand. Parametric spectrum analysis of 2D NMR signals. Application to in vivo J spectroscopy. *Journal of Magnetic Resonance*, 81:538–551, 1989.

[LC93] F. Luthon and A. Caplier. Motion detection and segmentation in image sequences using Markov Random Field modelling. In *4th Eurographics Animation and Simulation Workshop (EG'93)*, pages 265–275, Barcelona, Spain, September 4-5, 1993.

[LCL99] F. Luthon, A. Caplier, and M. Liévin. Spatiotemporal MRF approach to video segmentation: application to motion detection and lip segmentation . *Signal Processing*, 76(1), 1999.

[LD99] F. Luthon and D. Dragomirescu. A cellular analog network for MRF-based video motion detection. *IEEE Trans. on Circuits and Systems-I*, 46(2), February 1999.

[LDC+99] M. Liévin, P. Delmas, P.Y. Coulon, F. Luthon, and V. Fristot. Automatic lip tracking: Bayesian segmentation and active contours in a cooperative scheme. In *Proc. of IEEE Int. Conf. on Multimedia Computing and Systems (ICMCS99)*, Florence, Italy, June 7-11 1999. submitted.

[LL97a] M. Liévin and F. Luthon. Analyse automatique du mouvement des lèvres d'un locuteur. Rencontres ELESA, ENSIEG, Grenoble, 14 Novembre 1997.

[LL97b] F. Luthon and M. Liévin. Lip motion automatic detection. In *Proc. 10th Scandinavian Conference on Image Analysis (SCIA'97)*, pages 253–260, Lappeenranta, Finland, June 9-11, 1997.

[LL98a] M. Liévin and F. Luthon. Détection du mouvement des lèvres. In *Actes Journées Compression et Représentation des Signaux Audiovisuels (CORESA'98)*, CNET Lannion, France, 9-10 Juin 1998.

[LL98b] M. Liévin and F. Luthon. Lip features automatic extraction. In *Proc. of IEEE Int. Conf. on Image Processing (ICIP'98)*, pages 168–172, Chicago, Illinois, October 4-7 1998. Vol. 3.

[LL99] M. Liévin and F. Luthon. Unsupervised lip segmentation under natural conditions. In *Proc. of IEEE Int. Conf. on Acoustic, Speech and Signal Processing (ICASSP'99)*, Phoenix, Arizona, March 15-19 1999. Accepted for publication.

[LP95] F. Luthon and S. Popescu. Un circuit intégré analogique pour la détection de mouvement. In *Journées sur les nouvelles techniques pour la compression et la représentation des signaux audiovisuels (CORESA'95)*, pages 105–112, CCETT, Rennes, France, 12-13 Janvier 1995.

[LPC94a] F. Luthon, G. V. Popescu, and A. Caplier. Analog resistive networks for motion detection. In V. I. Vlad, editor, *4th Conference in Optics (ROMOPTO'94)*, pages 299–301, Bucharest, Romania, September 5-8, 1994. Proc. SPIE 2461.

[LPC94b] F. Luthon, G. V. Popescu, and A. Caplier. An MRF-based motion detection algorithm implemented on analog resistive network. In J. O. Eklundh, editor, *Proc. 3rd European Conference on Computer Vision (ECCV'94)*, pages 167–174, Stockholm, Sweden, May 2-6, 1994. Springer-Verlag. LNCS Vol. 800.

[LS96] F. Luthon and S. Simon. Transformée en ondelettes et compensation de mouvement dans les séquences d'images. In *Journées Compression et Représentation des Signaux Audiovisuels (CORESA'96)*, pages 2–9, CNET, Meylan, France, 15-16 Février 1996.

[Lut85] F. Luthon. Analyseur vectoriel de réseaux microondes - Méthodes générales de calibrage gérées par calculateur. DEA Systèmes Electroniques, ENSERG, Grenoble, Septembre 1985. 70 pages.

[Lut88a] F. Luthon. *Filtrage et analyse spectrale paramétriques. Applications en spectroscopie et magnétométrie RMN. Eléments de synthèse pour un analyseur paramétrique.* PhD thesis, INPG, Grenoble, Septembre 1988. 298 pages.

[Lut88b] F. Luthon. *Filtrage et analyse spectrale paramétriques. Notice d'utilisation des logiciels.* L.E.T.I., Grenoble, France, Novembre 1988. Rapport technique, 116 pages.

[Lut89a] F. Luthon. IT4 : le programme national suédois de développement industriel en technologies de l'information. Rapport d'Ambassade no. SC 6 ST/I, Ambassade de France, Stockholm, Suède, Juin 1989. 47 pages.

[Lut89b] F. Luthon. Le programme spatial suédois. Rapport d'Ambassade no. SC 8 ST/I, Ambassade de France, Stockholm, Suède, Décembre 1989. 47 pages.

[Lut90a] F. Luthon. Filtrage et analyse spectrale paramétriques, Exemples d'applications. In *Traitement du Signal*, chapter 7. INSTN, CEA-CENG, Grenoble, 1990. Document de cours.

[Lut90b] F. Luthon. Le centre de recherche nucléaire de Studsvik. Rapport d'Ambassade no. SC 1 ST/I, Ambassade de France, Stockholm, Suède, Mai 1990. 50 pages.

[Lut93] F. Luthon. Implémentation de la segmentation markovienne du mouvement dans des séquences d'images. Journées thématiques du GDR 134, GT2, Champs de Markov - Segmentation, Telecom Paris, 14 Octobre 1993.

[Lut96] F. Luthon. Mathématique de l'ingénieur. Reprographie ENSERG, Grenoble, 1996. Document de cours, 100 pages.

[Lut97] F. Luthon. Détection de mouvement dans les séquences d'images. Mises en oeuvre temps réel. In A. Chéhikian, P. Y. Coulon, and F. Luthon, editors, *Ecole des Techniques Avancées en Signal-Image-Parole (ETASIP'97)*, pages 181–201. INP Grenoble, France, 9-12 Septembre 1997. Session 2.

[PL95] S. Popescu and F. Luthon. The design of an efficient VLSI circuit for motion detection at high image repetition rate. In *2nd Advanced Training Course: Mixed Design of VLSI Circuits (MixVLSI'95)*, pages 361–366, Krakow, Poland, May 29-31 1995.

Echantillons de Publications