

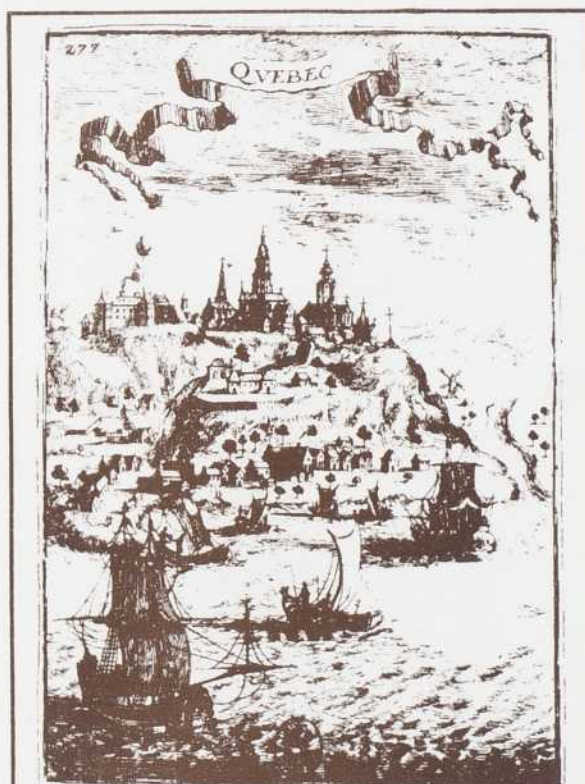
L'intelligence artificielle dans les technologies de l'information

*Sous la direction de
Richard Lepage et Rita Noumeir*



LES CAHIERS SCIENTIFIQUES

90



Bibliothèque Nationale du Québec



L

SYN

Les techniques d'intelligence artificielle appliquées aux technologies de l'information

*Réflexions sur les approches neuroniques,
symboliques et numériques appliquées à la vision,
l'écrit, la parole et le biomédical*

**Actes du colloque multidisciplinaire
*L'intelligence artificielle dans les technologies de
l'information***

**tenu dans le cadre du congrès de l'Acfas
à Montréal en mai 1996**

**publiés sous la responsabilité de
Richard Lepage et Rita Noumeir**

**Laboratoire d'imagerie, de vision et d'intelligence artificielle
Département de génie électrique
École de technologie supérieure**

Remerciements

Nous avons appris que la préparation d'un colloque et surtout du manuscrit qui immortalise ses présentations constitue une tâche immense qui nécessite la collaboration de nombreux intervenants.

Le colloque a été organisé dans le cadre d'une subvention de trois ans du programme FODAR (développement de la recherche) du réseau de l'Université du Québec, «Développement d'un réseau concerté multidisciplinaire de recherche-formation dans le domaine de l'intelligence artificielle appliquée à l'imagerie, la vision et la parole». Le réseau comprend les huit chercheurs du Laboratoire d'imagerie, de vision et d'intelligence artificielle de l'École de technologie supérieure qui agit comme constituante leader, deux chercheurs du département de géographie de l'UQAM, un chercheur du département de psychologie de l'UQAM, un chercheur du département de design graphique de l'UQAM, deux chercheurs du département de sciences appliquées de l'UQAC, un chercheur de l'UQTR et un chercheur de l'INRS-Télécommunications. Le colloque a pu voir le jour grâce à l'enthousiasme des chercheurs du réseau et des étudiants gradués impliqués dans ces projets. Nous tenons à remercier M. Jean Lebel, Directeur du programme FODAR, pour son appui financier à la tenue du colloque et à la publication des Actes. Nos remerciements vont aussi à M. Christian Lardinois, Doyen des études avancées et de la recherche de l'ÉTS, pour son complément de subvention et ses encouragements constants.

Dès le début, il fut décidé d'ouvrir le colloque à la communauté scientifique francophone. Un appel de participation fut lancé, et la réponse fut on ne peut plus enthousiaste. Nous remercions très chaleureusement nos confrères de l'extérieur du réseau de l'Université du Québec et même du réseau universitaire qui nous ont fait partager leur expertise et leur intérêt dans ce domaine de recherche en pleine effervescence. Un merci tout à fait spécial à notre confrère français, notre invité au colloque, M. Jean-Pierre Tubach, qui a su synthétiser les principaux enjeux de la recherche dans ce vaste domaine et qui a animé avec beaucoup de compétence la table ronde à la fin du colloque.

Q
334
C65
1996

Table des matières

Remerciements

Table des matières.....	i
--------------------------------	----------

Préface.....	1
---------------------	----------

Richard Lepage

Résumé.....	1
Introduction.....	1
Les enjeux et les techniques.....	3
La vision artificielle	4
Le traitement du document et de la parole.....	5
L'imagerie biomédicale	6

Les enjeux et les techniques	7
---	----------

Intelligence Artificielle et Technologies de l'Information	9
---	----------

Jean-Pierre Tubach

Résumé.....	9
Introduction.....	9
Technologies de l'Information	9
Qu'est-ce qu'IA ?	10
Le contexte actuel	10
IA et monde réel: mise en perspective des niveaux.....	12
Image.....	13
Imagerie médicale.....	14
«Vision» et fusion de données	15
Autres approches récentes en analyse et compréhension d'images....	16
Interaction personne - système.....	16
Traitement de Document, Traitement de l'Écrit.....	17
Document.....	18
Reconnaissance d'écriture	19
Signatures	19
Parole	20
Traitement du langage naturel	23
Deux jolis problèmes pour terminer	24
La téléphonie interprétée	24
L'indexation automatique de séquences video	24

Acquisition des connaissances de spécifications des logiciels à l'aide d'un collecticiel: modélisation et prototypage26

Ghislain Lévesque, Guy Massicotte, Alain Albert

Résumé.....	26
Introduction.....	26
Définition et contexte d'application du Collecticiel	28
Modèle générique du Collecticiel	29
Premier niveau du modèle structural	30
Le modèle évolutif d'une Séquence de travail	31
Le modèle structural détaillé	32
Conclusion	34
Références.....	35

La segmentation d'images par système à base de connaissances36

Daniel Crevier, Abdenour Bouzouane, Guturu Parthasarathy,

Richard Lepage

Résumé.....	36
Introduction.....	36
Description de la méthode	38
Calcul de l'adjacence des segments aux régions	39
Prise en compte de l'organisation perceptuelle	41
Mécanisme de contrôle	43
Résultats	43
Conclusion	44
Bibliographie.....	44

Le traitement optique en visionique47

Roland Schaefer et Denis Gingras

Résumé.....	47
Introduction.....	47
Structures optiques pour le traitement analogique en visionique	48
La transformée optique de Fourier	49
Le corrélateur à plan fréquentiel.....	49
Le corrélateur à transformée jointe.....	51
Performances du traitement optique analogique.....	52
Transformation de coordonnée et autres transformées	54
Exemples de traitement optique en visionique	55
Quel sera l'avenir des processeurs optiques?.....	57
Conclusion	58
Références bibliographiques.....	58

Approche connexionniste au problème de l'extraction de connaissances terminologiques à partir de textes62

Jean-Guy Meunier et Georges Nault

Résumé.....	62
Le problème	63
Extraction et classification.....	64
Les approches émergentistes.....	65
Le cadre théorique et l'hypothèse	66
Le traitement connexionniste du lexique d'un texte	68
L'expérimentation.....	69
Exemple	71
Interprétation et validation des résultats	73
Conclusion	73
Recherches et développements futurs.....	74
Références.....	75

La reconnaissance de la parole en français77

Pierre Dumouchel, Rivarol Vergin, Douglas O'Shaughnessy et Jean Rouat

Résumé.....	77
Introduction.....	77
Système de reconnaissance	78
Vue d'ensemble	78
Modèle de langage	79
<i>Modèle de langage déterministe</i>	79
Modèle acoustico-phonétique	80
<i>Les modèles HMMs</i>	81
<i>Principe de reconnaissance</i>	82
Transcription phonétique	83
Tests et résultats	84
Conclusion	85
Références.....	85

Vision artificielle 87

L'effet du désapprentissage dans les modèles autoassociatifs de la catégorisation.....89

Robert Proulx, Nicolas Ledoux et Jean Bégin

Résumé.....	89
Introduction.....	89
Le modèle EIDOS.....	91
Architecture	91
Transmission.....	92
Apprentissage	92
Propriétés du modèle	92
La filtration du bruit et le désapprentissage.....	93
Méthode de simulation.....	94
Stimuli.....	94
Procédure d'apprentissage, de désapprentissage et de rappel	94
Résultats	95
Effet du désapprentissage	96
Effet de la fréquence	97
Effet du patron de désapprentissage	98
Conclusion	99
Références.....	99

Revue des principaux modèles de réseaux neuroniques en vision artificielle préattentive.....100

Richard Lepage

Résumé.....	100
Introduction.....	100
Modèle computationnel du système de vision numérique.....	101
Extraction des caractéristiques de l'image.....	103
Détection directionnelle des arêtes	103
Détection isotrope du gradient.....	105
Passage par zéro de la dérivée seconde	106
Détection des arêtes par apprentissage supervisé	106
Détection multirésolution	108
L'esquisse	109
Stéréoscopie	109
Ombrage	109
Description et classification 3D	111
Conclusion	111
Références.....	112

Reconnaissance des linéaments à partir d'images acquises par satellite.

Discours sur la méthode 116

Desjardins, R., Iris, S. et St-Onge, B. A.

Roy, D.W.

Résumé..... 116

Introduction..... 116

Problématique générale..... 117

Objectif de la recherche 117

État de la question..... 117

 Les méthodes analogiques 117

 Les méthodes numériques..... 118

Les filtres directionnels..... 119

 Étape 1: Les prétraitements 120

 Étape II: Détection et extraction des linéaments 120

 Étape III: La cartographie 122

Les réseaux de neurones 122

 Étape I: Les prétraitements 123

 Étape II: Phase de traitement 123

 Étape III: Cartographie 123

Conclusion 123

Bibliographie..... 124

Traitement du document et de la parole. 129

Lecture automatique d'écriture cursive utilisant des concepts

perceptifs..... 131

M. Côté, E. Lecolinet, M. Cheriet, C. Y. Suen

Résumé..... 131

Introduction..... 131

Modèle de McClelland & Rumelhart..... 132

Méthode proposée..... 133

Activation..... 136

Étiquetage 137

Simulations et résultats préliminaires 138

Conclusion 139

Références..... 140

Une approche perceptuelle pour la définition d'une représentation adaptée au problème de la vérification hors-ligne des signatures manuscrites 141

Robert Sabourin, Ph.D., ing.

Résumé.....	141
Introduction.....	141
Critères pour la définition d'une représentation adaptée à la signature.....	142
Une représentation de la signature inspirée de la perception visuelle ...	143
Le spectre de forme comme facteur de forme local.....	144
Définition des principaux opérateurs morphologiques.....	144
Définition d'une représentation locale de la signature	146
Expérimentation.....	148
Prototypage	148
Évaluation de la performance des classificateurs	149
Conclusion	150
Références.....	150

Éditeur de schémas de composants électroniques dessinés sur ardoise électronique..... 151

Dave Aimé-Désiré Kodjo & Jean-Jules Brault

Résumé.....	151
Introduction.....	151
Problématiques.....	152
Description fonctionnelle de l'éditeur	153
Symboles électroniques	153
Commandes	153
Caractérisation et reconnaissance des symboles.....	154
Prétraitements	154
Caractérisation des symboles.....	154
Reconnaissance des symboles.....	155
Restauration du schéma	155
Évaluation de l'éditeur.....	156
Recommandations.....	157
Conclusion	157
Références.....	157

Utilisation d'un réseau de neurones pour localiser les informations pertinentes distribuées sur un signal manuscrit.....162

Jean-Jules Brault

Résumé.....	162
Problématique	162
Contexte	162
Taxinomie des systèmes existants en vérification des signatures (VdS) ..	163
Les lacunes des approches symboliques en VdS	163
Approche connexionniste ou distribuée.....	164
Réseau de concentration temporelle proposé par Tank & Hopfield	165
Adaptation du réseau de Tank & Hopfield à la VdS	166
Choix des détecteurs	166
Décomposition fréquentielle.....	166
Décomposition en déplacement tangentiel	167
Paramètres du réseau à fixer	167
Expérimentation et discussion	167
Signaux synthétiques et signatures réelles déformées	167
Deux classes de signatures réelles et leurs imitations	168
Recommandations supplémentaires pour un système complet de VdS....	168
Conclusion	169
Remerciements.....	169
Références.....	170

Un modèle formel pour le traitement de données visuelles sur des formulaires d'affaires et documents manuscrits.....172

M. Cheriet, J. N. Said, C. Y. Suen

Résumé.....	172
Introduction.....	173
Seuillage dynamique par l'approche étendue de Otsu.....	174
Le modèle en bref	175
Résultats expérimentaux et conclusion.....	178
Références.....	179

Détection des lignes de référence de mots cursifs à l'aide de l'entropie 184

M. Côté, M. Cheriet, C. Y. Suen, E. Lecolinet

Résumé.....	184
Introduction.....	184
Méthode proposée.....	185
Extraction des points du contour.....	186
Génération d'histogrammes.....	186
Choix du meilleur histogramme selon l'entropie.....	187
Extraction des lignes de référence.....	188
Simulations, résultats et discussion.....	189
Conclusions.....	193
Références.....	193

Modélisation du langage pour la reconnaissance de la parole: une approche stochastique non-déterministe à contexte de longueur variable..... 194

Marie-Hélène Talon, Pierre Dumouchel, Douglas O'Shaughnessy

Résumé.....	194
Introduction.....	194
Pourquoi un modèle de langage.....	194
Le modèle n-gramme.....	195
Le modèle d'interpolation.....	197
Le modèle "back-off".....	197
Mesure de la qualité d'un modèle de langage.....	198
Le modèle VNSA.....	199
Principe de base.....	199
Estimation des probabilités et des paramètres a.....	201
Calcul de la perplexité.....	203
Expériences et résultats.....	203
Base de données.....	203
Entraînement.....	203
Test.....	204
Conclusion.....	204
Références.....	205

Imagerie biomédicale 207

Numérisation d'images radiologiques par segmentation morphologique.....209

J. Potvin, J.A. de Guise, Y. Ross, J. Dansereau

Introduction.....	209
Méthodologie	210
Méthode de segmentation	210
Modèles Statistiques	212
Évaluation de la méthode.....	214
Résultats.....	215
Discussion et conclusion.....	216
Références.....	217

Un système structurel de vision pour la reconnaissance d'objet dans des images radiographiques.....219

Hail Mallouche, Jacques A. de Guise et Yves Goussard

Résumé.....	219
Introduction.....	219
Modélisation géométrique	221
Représentation des connaissances.....	222
Critère à trois interactions.....	223
Implantation	225
Résultats expérimentaux	225
Discussion	227
Références.....	229

Recalage 3D automatique d'images SPECT du cerveau.....230

J. Meunier, B. Imbert, C. Janicki, A. Guimond, J-P Soucy

Résumé.....	230
Introduction.....	230
Méthode	231
Étape 1: Recalage linéaire	231
Étape 2: Recalage non-linéaire	232
Résultats.....	235
Conclusions.....	237
Références.....	238

Reconstruction du mouvement d'une ellipsoïde à partir de ses projections.....240

Rita Noumeir

Résumé.....	240
Introduction.....	240
Méthode	242
Projection orthographique	242
Reconstruction tomographique.....	243
Rotation tridimensionnelle.....	244
Détection de mouvement	245
Quantification du mouvement	246
Résultats	247
Conclusion	249
Remerciements.....	250
Références.....	250

Table ronde 251

Table ronde.....253

Participants du colloque

Présidée par Jean-Pierre Tubach, ENST, Paris

Compilé par Richard Lepage, ETS

Résumé.....	253
Discussions	253

Préface

Richard Lepage

*Laboratoire d'Imagerie, de Vision et d'Intelligence Artificielle
École de technologie supérieure, Montréal (Québec), CANADA.
e-mail: lepage@ele.etsmtl.ca,*

Résumé

Le développement de la société moderne est érigé essentiellement sur le savoir, succédant ainsi à une société axée sur l'exploitation des matières premières. On constate donc une nette recrudescence du développement des technologies de l'information, qui regroupent tant les télécommunications que les sciences de l'informatique. Les technologies de l'information entrent dans une nouvelle phase d'évolution explosive qui se caractérise d'abord par un foisonnement de développements technologiques convergents: entre autres, le traitement parallèle, les mémoires optiques, les réseaux numériques, la simulation, la visualisation, les réseaux neuroniques, les industries de la langue et les systèmes experts. Les travaux de recherche sont centrés sur l'intégration des médias aux processus de traitement de l'information, la génération et la compréhension du langage naturel et le traitement des connaissances et de l'information.

L'intelligence artificielle, pour sa part, connaît des progrès remarquables actuellement. L'intelligence artificielle essaie de mimer les processus cognitifs de l'être humain au moyen de l'ordinateur. L'intelligence artificielle trouve des applications dans plusieurs domaines, dont en technologie de l'information.

Le colloque visait à procurer aux participants un survol des activités de recherche dans l'application des techniques d'intelligence artificielle dans le domaine des technologies de l'information. Il avait comme principal objectif de rassembler des chercheurs francophones pour présenter les principaux travaux de recherche dans le sous-ensemble des technologies de l'information comprenant la vision par ordinateur, le traitement du document écrit et de la parole et l'imagerie biomédicale.

1. Introduction

Le colloque *L'intelligence artificielle dans les technologies de l'information* s'est déroulé les 15 et 16 mai 1996 à l'Université McGill dans le cadre du 64^e congrès de l'Acfas. Vingt présentations étaient au programme et il fut décidé avant même la tenue du colloque que les présentations seraient appuyées par des articles révisés par les pairs. Ces Actes du colloque sont le résultat de cette volonté commune de perpétuer par écrit quelques résultats de recherche dans ce domaine de recherche en pleine croissance.

Les Actes se veulent un reflet de la recherche en intelligence artificielle appliquée aux technologies de l'information qui s'est déroulée récemment, au Québec plus particulièrement. Les techniques d'intelligence artificielle retenues pour le colloque sont:

1. Les réseaux neuroniques artificiels. Le relativement nouveau paradigme des réseaux neuroniques s'inspire, bien imparfaitement encore, du fonctionnement massivement parallèle et associatif du cerveau des vertébrés pour offrir des tentatives de solution parallèle et adaptative à des problèmes réputés complexes pour une machine de calcul de type séquentiel.
2. Les systèmes symboliques à base de connaissance. Une approche Haut → Bas est préconisée dans laquelle le contenu sémantique de la scène observée infère sur les traitements de bas niveau responsables d'extraire les caractéristiques utilisées pour l'identification et la classification. Les connaissances sont emmagasinées sous forme de règles ou d'arbres de décision.
3. Les algorithmes numériques. Les connaissances permettant de résoudre le problème complexe à l'étude sont directement emmagasinées dans la structure de l'algorithme choisi et dans le choix des paramètres de fonctionnement de ce dernier. Par exemple, en imagerie biomédicale, le choix d'un algorithme d'appariement stéréoscopique des contours actifs présents sur les deux radiographies d'entrée constitue une approche numérique de solution de la reconstitution tridimensionnelle de la structure vertébrale observée.

Le développement de la société moderne est érigé essentiellement sur le savoir, succédant ainsi à une société axée sur l'exploitation des matières premières. On constate donc une nette recrudescence du développement des technologies de l'information, qui regroupent tant les télécommunications que les sciences de l'informatique. Les technologies de l'information entrent dans une nouvelle phase d'évolution explosive qui se caractérise d'abord par un foisonnement de développements technologiques convergents: entre autres, le traitement parallèle, les mémoires optiques, les réseaux numériques, la simulation, la visualisation, les réseaux neuroniques, les industries de la langue et les systèmes experts. Les travaux de recherche sont centrés sur l'intégration des médias aux processus de traitement de l'information, la

génération et la compréhension du langage naturel et le traitement des connaissances et de l'information. Le colloque s'est concentré sur les technologies de l'information axées sur les sciences de l'informatique, soit plus particulièrement:

1. La vision artificielle.
2. Le traitement du document et de la parole.
3. L'imagerie biomédicale.

2. Les enjeux et les techniques

Cette première section des Actes regroupe différents articles qui résument la recherche francophone dans les différents domaines retenus pour le colloque. Les articles sont de nature plus exhaustives et procurent un bon survol des techniques d'intelligence artificielle appliquées à la vision et au traitement de l'écrit et de la parole.

Un premier article, de notre conférencier invité Jean-Pierre Tubac, ENST Paris, brosse un tableau général des différents thèmes abordés durant le colloque. On y précise les notions d'«intelligence artificielle» et de «technologies de l'information» pour ensuite en restreindre la portée pour les besoins spécifiques du colloque.

Le second article, de Ghislain Lévesque *et al.*, explore le développement d'un collecticiel informatique de spécifications de logiciels qui permet de rassembler et documenter les connaissances des experts du domaine. Le collecticiel est un progiciel de travail en groupe à distance offrant le soutien pour les communications électroniques et l'accès partagé à la documentation nécessaire à la réalisation de ce travail.

L'article de Daniel Crevier *et al.* montre les avantages d'utiliser un système à base de connaissances pour aider à résoudre un problème difficile en vision artificielle, la segmentation des différents objets ou parties d'objet présents dans la scène visuelle observée. Les connaissances *a priori* sont emmagasinées sous forme de règles *si-alors* et sont fondées en grande partie sur l'organisation perceptuelle du système de vision humain.

L'article de Roland Schaefer et Denis Gingras porte sur l'utilisation du

traitement optique de l'information visuelle. Un des grands avantages du traitement optique est qu'il permet un traitement en temps réel entièrement parallèle de l'information visuelle. L'article explore les différents outils optiques qui sont maintenant disponibles à la recherche tels les corrélateurs et les réseaux neuroniques à base d'hologramme.

Une approche connexioniste (par réseaux neuroniques) pour explorer et extraire des termes dans un corpus de documents écrits est proposée dans l'article de Jean-Guy Meunier et Georges Nault. Le problème de base abordé par cet article réside dans la dynamique et la plasticité de l'information textuelle emmagasinée faisant l'objet de la recherche de termes.

La reconnaissance de la parole en français constitue le dernier grand thème abordé dans cette section. Les difficultés propres à la langue française y sont abordées pour montrer l'ampleur du problème à résoudre. Une approche algorithmique axée sur un modèle statistique discret de chaîne de Markov cachée y est utilisée.

3. La vision artificielle

La vision artificielle consiste à traiter des images synthétisées ou provenant de senseurs variés (caméras 2D ou 3D, imageurs médicaux) afin d'en interpréter le contenu ou de reconstituer la scène tridimensionnelle originale.

Un premier article de cette section, de Robert Proulx *et al.*, propose l'utilisation d'une mémoire associative réalisée par réseau neuronique pour la reconnaissance de forme (en particulier, pour la reconnaissance de caractères alphanumériques manuscrits). Le modèle neuronique permet une reconnaissance robuste et parallèle de caractères bruités. L'article explore également la notion de désapprentissage afin d'améliorer les performances de reconnaissance du réseau.

L'article de Richard Lepage passe en revue les principaux modèles de réseaux neuroniques utilisés en vision artificielle préattentive. L'article propose d'abord un modèle computationnel d'un système de vision artificielle pour ensuite se concentrer sur les divers algorithmes neuroniques de bas niveau dont le rôle est d'extraire les caractéristiques de la scène observée en vue d'une éventuelle identification et localisation des objets présents.

Le dernier article de cette section, par Robert Desjardins *et al.*, donne un aperçu des applications de la vision artificielle en télémétrie pour l'interprétation d'images satellitaires. Une approche numérique et une autre neuronique sont proposées pour détecter les linéaments dans de telles images. Les linéaments correspondent à des événements géologiques telles des failles dans l'écorce terrestre et leur détection peut aider par exemple à la prédiction de séismes naturels.

4. Le traitement du document et de la parole

Le traitement automatique des documents vise à identifier et à authentifier des éléments d'une scène résultant de l'utilisation de l'écrit. La technologie de la reconnaissance de l'écrit joue un rôle primordial dans l'industrie du traitement des données. Depuis les dernières années, la recherche est focalisée sur le problème stimulant de la reconnaissance de l'écriture et du script cursif.

Un premier article, de Myriam Côté *et al.*, propose un réseau neuronique basé sur des concepts perceptifs pour la reconnaissance hors-ligne de mots cursifs isolés. Robert Sabourin, pour sa part, s'inspire également des concepts de perception visuelle humaine pour la vérification hors-ligne de signatures manuscrites.

La vérification en-ligne du signal manuscrit généré par une tablette de numérisation constitue le sujet des deux articles suivants. Un premier article, par Dave Aimé-Désiré Kodjo et Jean-Jules Brault, compare diverses techniques de reconnaissance de forme, dont deux neuroniques, dans une application d'éditeur de schémas dessinés à la main de composants électroniques. Le second article, de Jean-Jules Brault, explore les possibilités d'un réseau neuronique à minimisation d'énergie interne pour la reconnaissance d'un signal manuscrit de signature.

Mohamed Cheriet *et al.* proposent une approche numérique basée sur la morphologie mathématique afin d'extraire les divers champs d'intérêt sur un formulaire d'affaire. Plusieurs algorithmes de vision et de morphologie sont mis à contribution pour construire un modèle formel de traitement de données apparaissant sur un formulaire ou un document manuscrit. D'autre part, Myriam Côté *et al.* proposent d'utiliser l'entropie sur l'histogramme de densité pour détecter les lignes de base de mots cursifs, ce qui contribue à mieux identifier le contenu des divers champs d'intérêt.

Le dernier article de cette section, de Marie-Hélène Talon *et al.*, propose une approche numérique stochastique non-déterministe pour l'important problème de la reconnaissance de la parole à contenu variable.

5. L'imagerie biomédicale

La dernière section de ces Actes est consacrée à l'imagerie biomédicale. Ce dernier champ de recherche s'intéresse aux techniques d'acquisition, de traitement et d'analyse des images issues d'imageurs médicaux variés (radiographie simple et multiplanare, tomодensitométrie par rayons X et résonance magnétique, médecine nucléaire, ultrasons). Plusieurs approches numériques y sont proposées.

Une première approche, basée sur la segmentation morphologique, est présentée par Julie Potvin *et al.* pour la numérisation d'images radiologiques. L'article porte plus spécifiquement sur la reconstruction tridimensionnelle de la colonne vertébrale de patients scoliotiques. Hail Mallouche *et al.* proposent pour leur part l'utilisation de techniques de vision artificielle, plus spécifiquement l'extraction des contours dynamiques, afin de faciliter la reconstruction tridimensionnelle de structures biologiques comme la colonne vertébrale.

Jean Meunier *et al.*, présentent une méthode basée sur le flux optique pour comparer les images SPECT du cerveau d'un patient prises à divers moments afin d'en étudier l'évolution. Enfin, Rita Noumeir propose une approche géométrique pour l'étude du mouvement du coeur en tomographie cardiaque. Le coeur est modélisé comme une ellipsoïde et ses différentes projections sont analysées afin de reconstituer la forme originale.

Les enjeux et les techniques

Le premier aspect de l'analyse de l'impact de l'IA sur la société est celui de l'impact sur l'emploi. L'IA a le potentiel de remplacer certains emplois, mais elle peut aussi en créer de nouveaux. L'impact net dépendra de la façon dont la société gère la transition.

3.2. L'impact sur l'éducation

La deuxième partie de l'analyse de l'impact de l'IA sur la société est celle de l'impact sur l'éducation. L'IA peut être utilisée pour personnaliser l'apprentissage, adapter le rythme et le contenu de l'enseignement à chaque élève. Cela peut améliorer les résultats scolaires et réduire les inégalités éducatives. Cependant, l'IA peut aussi réduire le rôle des enseignants, qui doivent être formés pour utiliser ces outils efficacement.

Enfin, l'IA a un impact sur la vie privée et la sécurité. Les données personnelles sont collectées et analysées à grande échelle, ce qui pose des questions sur la confidentialité et le contrôle de ces données. De plus, l'IA peut être utilisée pour développer des armes autonomes, ce qui est une préoccupation majeure pour la communauté internationale. Ces enjeux sont essentiels et doivent être traités avec attention.

En conclusion, l'IA est une technologie puissante qui a le potentiel de transformer notre société. Cependant, elle présente aussi des défis importants. Il est crucial de développer une approche équilibrée qui maximise les avantages de l'IA tout en minimisant ses risques. Cela implique une coopération internationale, une régulation stricte et une éducation continue pour garantir que l'IA serve le bien commun.

Intelli

Résumé

Cet article explore les impacts de l'IA sur la société, en particulier sur l'emploi, l'éducation, la vie privée et la sécurité.

Nous abordons les problèmes exhaustifs de l'IA, du doute au l'IA n'y

1. Intro

Il parait évident que la technologie

1.1 Tec

Le développement de la technologie biomédicale automatisée

Certains

peuvent penser que les résultats de l'information particulière diffèrent de l'influence de l'IA n'y

Intelligence Artificielle et Technologies de l'Information

Jean-Pierre Tubach

*Ecole Nationale Supérieure des Télécommunications,
(CNRS Unité de Recherche Associée 820), Paris, FRANCE
E-mail: tubach@ds.enst.fr*

Résumé

Cet article se propose de «survoler» les quelques grands domaines de recherche qui vont faire l'objet de communications au cours du Colloque, et de donner quelques idées directrices, après avoir précisé dans quel sens nous entendons les mots «Intelligence Artificielle» et «Technologies de l'Information».

Nous y citons un certain nombre de travaux qui paraissent bien illustrer des problématiques ou approches actuelles. Mais nous n'avons évidemment pas cherché à être exhaustifs, et il ne s'agit nullement de dresser un tableau d'honneur! Cela permettra sans doute au lecteur de ne pas (trop) nous reprocher de nous être appuyés très souvent sur nos Collègues de l'ENST de Paris et de l'ETS de Montréal.

1. Introduction

Il paraît nécessaire dans cette introduction de préciser un peu ce que l'on entend par «Intelligence Artificielle» (IA), et dire quelles sont les Technologies de l'Information considérées.

1.1 Technologies de l'Information

Le programme du Colloque indique que l'on pense, en matière de technologies de l'information, à la vision artificielle, à l'imagerie biomédicale, à l'infographie et à la synthèse d'image, au traitement automatique de la parole et de documents.

Certes, au sens le plus habituel de «technologies de l'information», on pense aussi, voire surtout, plus explicitement à des choses qui tournent autour des réseaux à haut débit et du multimédia (donc des «autoroutes de l'information»). Cela comprend des aspects très «techniques», avec en particulier les technologies de transmission, les protocoles de réseaux à différents niveaux, l'administration et l'interconnexion des réseaux, l'influence du haut débit sur les systèmes d'exploitation des ordinateurs, etc.; l'IA n'y a guère sa place (quoique, pour la validation des protocoles ...).

En fait, le sens de «technologies de l'information» adopté ici n'est pas celui des fondements techniques des réseaux, mais bien plutôt celui des recherches scientifiques et techniques nécessaires pour parvenir à des services innovants sur ces réseaux, pour le bénéfice de l'utilisateur final.

1.2 Qu'est-ce qu'IA ?

Le terme IA a au moins deux sens différents :

- un sens restrictif: utiliser des techniques dites habituellement «d'intelligence artificielle», où une machine utilise des connaissances qui lui ont été fournies sous une forme explicite, pour représenter la compétence d'un expert humain; ou peut-être ces «connaissances» ont-elles été «appprises» à partir d'ensembles de données.

et

- un sens extensif: faire accomplir par une machine une tâche «intelligente», complexe, traditionnellement du ressort de l'homme (exemple: compréhension de langage naturel écrit ou parlé), ou accomplir un tâche complexe en coopération avec elle (exemple: prise de décision en environnement complexe).

Dans le premier sens (restrictif), on pense surtout aux systèmes experts (SE), systèmes à base de connaissances (KBS), où un formalisme informatique (règles,...) permet de rendre disponible des «connaissances» fournies par un expert humain, et de les mettre en oeuvre dans un processus d'inférence. On tend également à inclure dans cette rubrique les méthodes et réseaux connexionistes; il est plus difficile déjà d'y introduire d'autres méthodes utilisant des formalismes probabilistes (modèles de Markov...).

Nous pensons personnellement qu'il n'est pas vraiment intéressant de parler d'intelligence artificielle au sens des techniques qu'elle propose (en concurrence ou complémentarité avec d'autres), mais bien plutôt au sens des tâches accomplies, des applications et des usages, donc. Nous retenons le deuxième sens, le sens extensif.

1.3 Le contexte actuel

L'intelligence artificielle n'a pas partout la cote ces temps-ci; on entend exprimer des impressions d'échec, au moins partiel, de désillusion. Contre

coup sans doute des rêves d'une autre époque, et des problèmes rencontrés face à des questions très difficiles!

En fait, c'est surtout là opinion de chercheurs. D'où l'intérêt de parfois lire la grande presse, pour avoir d'autres sons de cloche: on y trouve certes des sujets de légitime agacement, mais parfois aussi de justes remises en perspective.

Comme le rappelle le journal *Business Week* du 17 juillet 95 (sous le bien mauvais titre «Computers that think are almost here»), il apparaît que l'IA a des milliers d'applications dans le commerce, l'industrie, pour les professionnels; 70% des 500 plus grandes compagnies américaines en sont utilisatrices, les ventes de logiciels d'IA ont atteint le milliard de dollars US en 1994 etc... Quelques exemples sont cités, du domaine médical (aide au diagnostic, à la prescription, examen d'électro-cardiogrammes...) au secteur bancaire (détection d'anomalies dans l'utilisation de cartes de crédit...).

Cela nous rappelle que les systèmes experts, en particulier, ont cessé d'être un domaine pour la recherche, mais ont de nombreuses applications, le plus souvent non visibles par le public; et des méthodes connexionistes sont aussi utilisées dans le même contexte.

On trouve dans le même article la phrase suivante, provocante mais plutôt juste: «quand quelque chose marche, on cesse de l'appeler IA, ça devient autre chose, ça s'appelle bases de données, reconnaissance de mots etc...» On pourrait ajouter que les techniques «orientées objets», nées ailleurs, reprises en IA sous la forme de «frames», «acteurs»,... ont surtout trouvé leur fortune en génie logiciel, où elles constituent un apport très important. Autrement dit, l'IA est un ensemble de techniques diffusantes en informatique!

Mais si le grand public est confronté à des machines supposées être plus ou moins «intelligentes», l'impression est «qu'elles ne sont pas si intelligentes que ça, les machines !»; il faut s'adapter à elles, alors qu'on voudrait le contraire. On trouvera des exemples dans les services téléphoniques avec reconnaissance vocale (avec l'automatisation des appels à frais virés de Nortel. il faut répondre par oui ou non à une question bien définie¹); dans les

1. Et on sait que reconnaître de façon très fiable les différentes façons qu'a le grand public de dire «oui» ou «non», au téléphone, n'a pas du tout été une tâche aisée en reconnaissance de parole!

dialogues permis par nombre de bornes d'information interactives, d'automates de distribution de billets de train, de services sur le Minitel, qui sont plutôt rigides.

On trouve là, et c'est une bonne chose, une justification pour l'ambition des chercheurs d'aller vers des systèmes plus «intelligents», vraiment utilisables par tout un chacun. Cela nécessite encore beaucoup de travail, pluridisciplinaire à l'évidence. C'est de cela que nous allons parler durant ce Colloque.

Au sens extensif que nous donnons à «Intelligence Artificielle», il apparaît souhaitable de ne pas se limiter strictement aux thèmes du programme, mais de s'intéresser aussi au traitement du langage naturel, et à l'interaction personne - système.

Au traitement automatique du langage naturel (TALN), pour plusieurs raisons: y a-t-il tâche plus «intelligente», plus caractéristique de l'intelligence humaine? De plus il intervient en reconnaissance de l'écriture et de la parole, pour pallier les imperfections de la reconnaissance d'unités élémentaires; enfin la Traduction automatique (TA) ou Traduction assistée par ordinateur (TAO) est un sujet d'importance et d'excellence au Canada.

A l'interaction personne - système, au moins comme fournissant un cadre, un objectif (parmi d'autres), un contexte d'utilisation pour les techniques qui font l'objet de nos recherches.

2. IA et monde réel: mise en perspective des niveaux.

Un système «intelligent» a un environnement (le monde réel, dont fait partie l'utilisateur); il interface avec lui par des signaux, analogiques, en entrée comme en sortie.

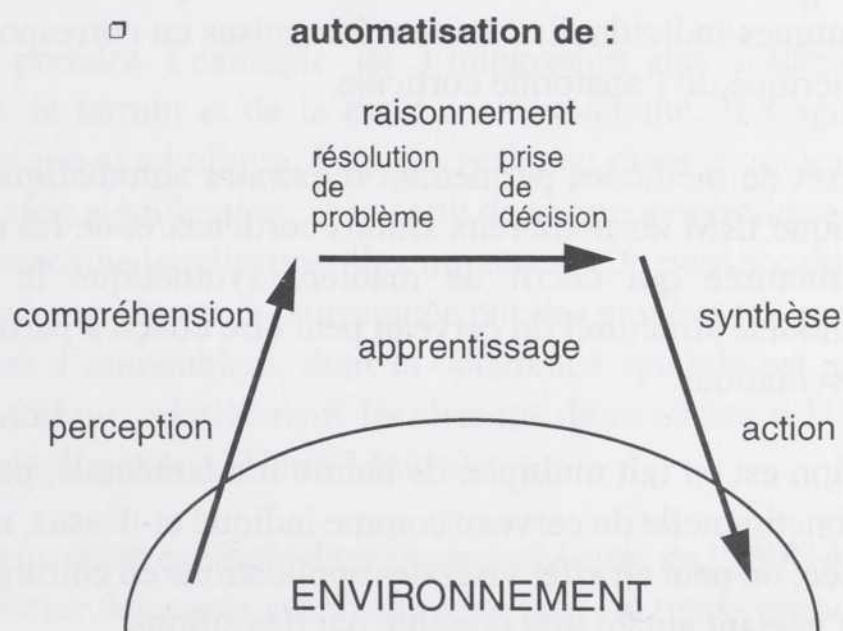
Le système comprend en entrée une phase ascendante, à travers la numérisation, le traitement du signal, la reconnaissance des formes, un premier niveau de compréhension, permettant le passage du continu au discret, du numérique au symbolique. C'est dans cette phase qu'actuellement des modèles de Markov, ou/et des réseaux neuromimétiques ont le plus de chances d'être utilisés.

Il y a ensuite une phase «horizontale» permettant sans doute un degré plus abstrait de compréhension, et assurément «raisonnement» en vue de résolution de problème, de prise de décision. C'est là que peuvent principalement intervenir des systèmes experts, la logique floue, etc.

On termine par une phase «descendante», pour la synthèse de commandes ou de messages, en vue de l'action.

Le tout peut bien sûr être itéré dans un contexte d'interaction, de dialogue.

Intelligence Artificielle



□ **modèles :**
- symboliques
- neuro-mimétiques

D'après Alain Grumbach (ENST)

3. Image

D'une façon générale, qu'il s'agisse d'imagerie médicale ou de vision artificielle (au sens large), un aspect important est la fusion (intelligente) de

données de sources différentes se rapportant au même «objet», et apportant des «visions» complémentaires

3.1 *Imagerie médicale*

On prendra d'abord l'exemple de l'imagerie cérébrale, domaine très actif actuellement.

Les neurologues disposent aujourd'hui d'un grand nombre de modalités d'imagerie permettant potentiellement d'accéder au fonctionnement du cerveau (tomographie d'émission de positons, magnéto-encéphalographie, etc...). Afin de réaliser une étude approfondie des processus physiologiques décrits par ces images, les informations fonctionnelles doivent être corrélées avec les structures anatomiques sous-jacentes, mises en évidence par l'imagerie par résonance magnétique (IRM) par exemple. Pour comparer ou fusionner les informations fonctionnelles issues de plusieurs individus, les images anatomiques individuelles doivent être mises en correspondance avec un modèle générique de l'anatomie corticale.

D'où l'intérêt de méthodes permettant d'extraire automatiquement d'une image anatomique IRM les nombreux sillons corticaux et de les agencer sous une forme structurée qui décrit de manière synthétique la topographie corticale. Un modèle structurel du cerveau peut être conçu à partir d'une base de telles représentations.

La motivation est en fait multiple: de nature fondamentale, parvenir à une cartographie fonctionnelle du cerveau comme indiqué ci-dessus, mais aussi de nature appliquée; on peut en effet viser des applications en chirurgie: atteindre une tumeur en passant autant que possible par des sillons.

Voir par exemple les travaux menés autour d'Isabelle Bloch à l'ENST, avec la thèse de JF Mangin

La restitution sur écran des images anatomiques obtenues après acquisition(s) et traitement a une grande importance pour les praticiens, dans les domaines de l'imagerie diagnostique, de la planification chirurgicale, de la morphométrie quantitative, et de l'assistance chirurgicale. C'est là qu'interviennent les techniques de l'infographie et de l'image de synthèse.

Voir par exemple les travaux de Jacques de Guise, pour l'orthopédie et en

particulier l'imagerie de la colonne vertébrale (ETS / Ecole Polytechnique / Hôpital Ste-Justine, Montréal).

3.2 «*Vision*» et fusion de données

Il arrive, par exemple en imagerie satellitaire ou aérienne, qu'une source d'information soit suffisante pour extraire l'information recherchée.

Par exemple, Rita Noumeir et Richard Lepage à l'ETS trouvent des linéaments (des segments de droites reliés) sur des images satellitaires. Ces linéaments indiquent généralement une présence de faille (tests sur la région de Charlevoix).

Mais bien souvent la fusion de données de types différents revêt une importance particulière.

On peut prendre l'exemple de l'intégration des informations d'une photographie de terrain et de la carte correspondante. Il s'agit d'un cas de fusion numérique-symbolique, l'image aérienne étant purement numérique, précise mais sans signification, et la carte de nature symbolique, faisant appel à des codes, avec une localisation plus imprécise. On peut aboutir par exemple pour une ville à une structure représentée par des graphes (de rues, de pâtés de maisons, voire d'immeubles), dont la cohérence spatiale est assurée par de champs de Markov, relativement localement donc (thèse d'H. Moissinac à l'ENST sous la direction d'Henri Maître).

On trouve un autre exemple dans le projet Messie de l'INRIA (Giraudon et al.). Pour détecter des ponts sur image aérienne, on tient compte de ce qu'ils sont sur des rivières ou des routes, que l'on sait détecter et suivre sur certaines images, les candidats-ponts étant plus facilement détectables sur d'autres. On utilise donc des connaissances a priori sur ce qu'on cherche, et fait intervenir une architecture de tableau noir, de système multi-agents; certains cas s'y prêtent bien; la même démarche en reconnaissance de parole dans les années 70 avait rencontré plus de difficultés (système Hearsay).

Ce type d'architecture (objets, agents, acteurs...), de fait, fournit un environnement logiciel commode pour expérimenter, mais pas des concepts ou outils nouveaux en reconnaissance des formes.

3.3 Autres approches récentes en analyse et compréhension d'images

Pendant longtemps, toute compréhension d'image ou de séquence d'images partait nécessairement d'analyses fines, avec détection précise de contours, etc.; on finissait par conclure qu'un objet était devant un autre, parce que situé à 3,84 mètres de l'objectif de prise de vue, l'autre étant à 4,21 mètres.

On travaille maintenant aussi à privilégier le qualitatif sur le quantitatif (Rosenfeld): raisonnement qualitatif, perception qualitative, en utilisant des ensembles flous, du raisonnement incertain pour des conclusions robustes et peu détaillées, par exemple justement pour estimer rapidement «c'est devant / c'est derrière». Autres travaux à rapprocher de cette approche: la détermination approximative des lignes de fuite dans une image, pour s'y situer.

Voir à titre d'illustration l'article de Schulman et Fermüller «Qualitative depth and focus of expansion information via smoothed flow».

Est-ce de l'IA? Oui et non; oui si l'on admet qu'il y a là certains aspects de la perception humaine; non si l'on ne regarde que l'aspect purement algorithmique des méthodes employées.

4. Interaction personne - système

Sans rappeler les concepts généraux de l'interaction personne système (utilisateur occasionnel et utilisateur habituel; modèle de l'utilisateur, modèle de la tâche...), nous prendrons l'exemple de l'ordinateur à stylo comme un cas d'interaction orientée vers l'utilisateur, et potentiellement multimodale, d'après l'article de Claudie Faure (ENST), dans la revue *La Recherche de Mars* 1996.

La métaphore du papier-stylo participe au courant qui tend à rapprocher l'interaction personne - système de la communication humaine. L'objectif d'une métaphore est de faciliter l'interaction en remplaçant l'utilisateur dans une situation familière. Il peut ainsi utiliser ce qu'il connaît de cette situation, l'interaction devient plus intuitive et le temps d'apprentissage de l'outil informatique est réduit.

Des actions sont possibles au delà de la reconnaissance de caractères (qui n'est pas ce qui marche le mieux...): barrer un mot ou un signe dans la

correction de document; obtenir une image «idéale» d'un dessin géométrique à main levée (ce que les américains appellent «beautification»). La capacité de réaction de l'ordinateur aux signaux d'entrée constitue une extension de la métaphore papier stylo, qui débouche sur des concepts d'assistance et de partage des tâches entre l'utilisateur et la machine.

L'utilisation de la modalité «voix» en complément au stylo, par exemple pour spécifier un mode d'utilisation, est possible avec des systèmes de reconnaissance de mots parlés disponibles commercialement et assez élémentaires, et ouvre des perspectives ergonomiques en cours d'exploration.

5. Traitement de Document, Traitement de l'Écrit

Faisons remarquer d'abord que le terme «traitement de document» est ambigu: des chercheurs en traitement du langage naturel penseront avant tout à l'automatisation de la compréhension, du résumé, de l'indexation, de l'accès à des documents en langue naturelle («informatique du contenu»). D'autres, et c'est ainsi que nous nous situons ici, envisagent plutôt l'ensemble des traitements nécessaires face à des documents matériels réels, en particulier lorsqu'ils ont été produits «à la main». La première interprétation est un (petit?) sous-ensemble de la seconde!

Le document est l'expression visuelle d'une pensée à communiquer («signaux intentionnels de la communication humaine»), dont une partie utilise la langue naturelle.

Il s'agit de déterminer dans l'image d'un document complexe ce qui est texte (imprimé, manuscrit), dessin etc..., de détecter des lignes d'écriture... Après quoi se pose en particulier le problème de la reconnaissance de l'écriture (reconnaissance de «caractères», reconnaissance d'écriture manuelle spontanée; cas particuliers des formules mathématiques, des signatures, etc...). Ce n'est qu'ensuite que l'on peut se soucier de compréhension, d'indexation etc...

Cette vue est à vrai dire trop schématique, des traitements linguistiques devant être utilisés pour remédier aux imperfections de la reconnaissance de caractères, par exemple (pour le cursif spontané, il ne saurait de toutes façons guère en être autrement!)

Quoi qu'il en soit, le traitement de l'écrit offre un terrain privilégié pour élaborer des modèles, puis des systèmes, intégrant perception artificielle et connaissances, ou autrement dit numérique et symbolique.

5.1 Document

Comme indiqué ci-dessus, la structuration perceptive du document est un préalable à sa lecture dans une chaîne complète de traitement.

Le traitement d'image est nécessaire, mais tout à fait insuffisant pour accéder à une représentation symbolique, conceptuelle, assez complète d'un document. Il faut de plus des traitements linguistiques, mais aussi une connaissance des modalités d'expression et autres «façons de faire» usuelles, un modèle du scripteur, donc.

En effet, le contenu du document, même en se limitant au texte, est fonction de l'expression linguistique («ce qui est écrit») mais aussi de la présentation visuelle («comment c'est écrit»): il faut par exemple tenir compte de la mise en valeur par des attributs typographiques et positionnels; la présentation ajoute des composantes extra-linguistiques au contenu.

Différents niveaux de structures doivent être étudiés dans un document (L. Likforman à l'ENST) :

- structure physique (groupements, alignements)
- structure logique (étiquetage des composantes de la structure physique, par exemple en titres, notes de bas de pages, corps du texte...)
- structure textuelle: mots, phrases, paragraphes, ordre de lecture¹...

On pourra se reporter à l'article de Liebowitz-Taylor et al. (Unisys Corporation) «Integrating Natural Language Understanding with Document Structure Analysis».

1. Ce dernier aspect n'est pas aussi simple qu'on le croit: que l'on songe à un journal quotidien, le texte y est en colonnes, elles sont parfois coupées par des photos ou illustrations; déterminer l'ordre de lecture (où aller à la fin d'une colonne) fait appel au moins à une syntaxe locale, parfois à la sémantique. C'est une illustration de l'équation lecture = perception + connaissances !

5.2 Reconnaissance d'écriture

Comme on l'a indiqué, la perception de l'écriture fait appel à la combinaison du visuel et du linguistique.

La prise en compte de contraintes linguistiques est nécessaire, elle peut se traduire par l'utilisation de lexiques généraux ou spécifiques, par la prise en compte de contraintes positionnelles ou fonctionnelles de la syntaxe ou par l'identification de champs sémantiques.

La reconnaissance de l'écriture cursive diffère de celle de caractères isolés (OCR) en ce sens que l'information y est parcellisée, difficile à localiser et généralement dégradée et lacunaire. D'où la nécessité de faire appel à l'information contextuelle. Pour certains cas (reconnaissance d'adresses postales...), utiliser la connaissance a priori d'un vocabulaire peut faciliter les différentes étapes de la reconnaissance. Cela peut conduire à des systèmes «descendants», partant du vocabulaire et de ses propriétés pour arriver aux pixels (E. Lecolinet à l'ENST), ou à la combinaison d'un amorçage perceptif et d'un renforcement des hypothèses de mots par guidage lexical (thèse de M. Côté à l'ETS et l'ENST avec M. Cheriet et E. Lecolinet)

D'autres travaux utilisent des méthodes d'apprentissage automatique, qui peuvent être proches de celles utilisées en reconnaissance de la parole: voir le travail de John Mahkoul présenté à ICASSP'94 sous le titre très révélateur: «On-line cursive handwriting recognition using speech recognition methods»! (Il s'agit de modèles de Markov cachés, avec un modèle de langage).

5.3 Signatures

Le cas particulier de la vérification automatique de signatures manuscrites comporte certaines spécificités.

Il faut distinguer deux cas bien différents en fonction des données disponibles :

- celui d'une saisie au stylo, où l'on dispose du tracé «ordonné», mais aussi des vitesses, accélérations, pressions, plus difficiles à contrefaire que le simple tracé. Le domaine est celui de l'authentification «en ligne» ;

- celui, beaucoup plus fréquent, où l'on ne dispose que de l'image de la signature, donc du tracé non ordonné seulement (encore faut-il l'avoir préalablement localisé); c'est la vérification «hors ligne».

Signatures Nord-Américaines et Européennes ont des styles très différents (et il en existe encore d'autres types); les types de faux sont multiples, de la «copie servile» au «faux aléatoire», et il est donc bien difficile d'envisager une méthodologie unique s'appliquant à tous les cas.

On peut être amené à faire coopérer différentes approches, utilisant des «mesures» différentes (locales, globale), à organiser la coopération de classifieurs pour parvenir à un système plus robuste: voir par exemple Sabourin et Genest (ETS) «Définition et évaluation d'une famille de représentations pour la vérification hors ligne des signatures».

6. Parole

Il y a des analogies entre parole et écriture manuscrite: elles sont chacune une forme d'expression de la langue, se présentent comme un signal difficile à segmenter en unités de base et sujet à une variabilité inter et intra individuelle.

On connaît les facteurs multiples qui conditionnent la difficulté d'une tâche de reconnaissance de parole:

- le mode d'élocution: mots isolés, mots connectés, parole continue, parole spontanée ;
- la dépendance ou l'indépendance au locuteur ;
- la taille du vocabulaire: systèmes à petit, grand, très grand vocabulaire suivant que cette taille se mesure en dizaines, en milliers ou en dizaines de milliers de mots ;
- les conditions acoustiques: studio d'enregistrement, milieu peu bruyant (bureau), téléphone, milieu très bruyant (atelier, véhicule)...

Si l'on sait effectuer pratiquement sans erreurs les tâches de reconnaissance les plus faciles (mots isolés, un seul locuteur, petit vocabulaire, pas de bruit), les performances se dégradent avec la complexité

(typiquement, actuellement, 12% d'erreurs au niveau des mots pour 20 000 mots en parole continue indépendamment du locuteur est déjà un beau résultat).

Jetons un coup d'oeil aux systèmes disponibles commercialement, qui représentent autant de points d'équilibres possibles entre la technologie disponible et les contraintes ou avantages d'applications particulières, bien différentes.

Télécommunications: reconnaissance d'un nombre limité de mots, au téléphone, par le grand public (donc sans apprentissage). Nortel à Montréal est leader dans ce domaine, à partir des recherches menées autour de Matthew Lennig (d'abord le célèbre système «oui-non», déjà mentionné, puis l'automatisation partielle des services 411 (systèmes ABDA, ABDPlus), et les services boursiers (STOCKTalk)... En France la situation est différente, même si la technologie est aussi disponible grâce aux travaux du CNET (Denis Jouvét et al.): le Minitel, très répandu et maintenant familier à tous, ne donne pas un contexte favorable au développement de services vocaux !

Dictée automatique, ou plutôt assistée, qui a l'inconvénient de n'être actuellement proposée qu'en mots isolés! La prise en compte d'une tâche précise dans un domaine bien défini peut rendre cette contrainte moins gênante (exemple de la confection de rapports médicaux où l'on appelle des paragraphes standard, et n'entre que les éléments variables).

Applications industrielles: par exemple la Poste américaine, et Federal Express, utilisent la reconnaissance de mots pour le tri des colis.

On pourra bien sûr citer d'autres applications, pour les handicapés, ou dans l'industrie du jouet et des jeux.

On va s'intéresser surtout ici aux systèmes les plus ambitieux, qui font toujours l'objet d'un gros effort de recherche dans beaucoup de laboratoires: typiquement reconnaissance grand vocabulaire, parole continue, visant la reconnaissance, c'est-à-dire la dictée¹

1. Peu de systèmes s'intéressent à la compréhension nécessaire au dialogue, si ce n'est en considérant le résultat de la reconnaissance comme l'entrée d'un système de compréhension de langage naturel!

La similarité des systèmes de reconnaissance des différents groupes de recherche est remarquable. Le traitement consiste d'abord à extraire des paramètres toutes les 10 millisecondes (coefficients cepstraux et leur première et seconde dérivée temporelle). Puis il y a comparaison de formes à plusieurs niveaux (acoustique, lexical, syntaxique). La modélisation acoustique est faite par un modèle de Markov caché, et les niveaux supérieurs sont représentés par une machine à états finis probabiliste. Les systèmes diffèrent surtout par les variations apportées au modèle de Markov caché et à la machine à états finis, ainsi que par les algorithmes de programmation dynamique employés afin de chercher la meilleure séquence de mots. Voir par exemple l'article de Rivarol Vergin dans ce Cahier pour la description d'un tel système.

On voit que l'intelligence artificielle a peu de rapport avec les techniques actuellement dominantes. On peut rappeler que ce sont les années 70 qui ont été l'époque reine de l'IA en reconnaissance de la parole, avant que les techniques d'apprentissage automatique de type Markov, étudiées à partir de 76 (Baker, Jelinek) ne s'imposent dans les années 80.

Il y avait alors des systèmes à base de connaissances, utilisant en particulier l'expertise de lecteurs de spectrogrammes; il y avait des architectures logicielles à l'époque nouvelles (tableau noir...). On peut dire qu'à cette époque finalement les chercheurs en parole ont apporté des innovations en IA... plutôt que le contraire.

Une caractéristique importante de cette communauté de recherche dans ce domaine, surtout en Amérique du Nord, mais aussi en Europe, est son adhésion au «paradigme de l'évaluation», sous l'influence en particulier des projets DARPA du Département de la Défense américain, avec la mise à disposition de corpus d'apprentissage et de tests communs à toutes les équipes impliquées. Des systèmes Européens participent aux sessions d'évaluation (LIMSI, Université de Cambridge...) et remportent de bons succès.

Il faut se féliciter de l'adoption de la même démarche pour les systèmes utilisant la langue française, grâce à l'action de recherche concertée «Linguistique, informatique et corpus oraux» de l'AUFELF UREF (réseau FRANCIL).

7. Traitement du langage naturel

Nous nous limiterons en fait aux domaines de la Traduction automatique, et de la Traduction assistée.

Si la traduction automatique est un sujet de recherche actif depuis les années soixante (par exemple à Grenoble, autour du regretté Bernard Vauquois, puis de Christian Boitet), on sait bien que le plein succès ne peut être atteint que dans des domaines restreints (dont les bulletins de la météorologie canadienne constituent un exemple fort connu). Pour des textes non contraints, on ne vise actuellement que la traduction assistée par ordinateur, en particulier autour du concept de poste de travail pour traducteur. Le Canada est en pointe dans ce domaine, la motivation étant forte en raison de son bilinguisme officiel: voir le poste de travail pour traducteur PTT3 du CITI (Pierre Isabelle et al.).

Une nouvelle génération de méthodes de traduction est apparue, s'appuyant sur l'alignement de textes homologues en deux langues différentes; on emploie le terme bi-textualité, qui fait référence à un ensemble de techniques permettant d'aligner des textes en langue source et en langue cible, de manière à rendre explicites les liens entre les segments correspondants dans les deux langues du texte. Cette base de données bi-textuelle pourra ensuite être mise au service de la production de nouvelles traductions au coeur d'une nouvelle génération d'outils d'aide au traducteur.

On dispose déjà de quelques grands ensembles de textes déjà traduits lisibles par machine; le plus connu de ces grands ensembles est celui des actes du Parlement canadien, qui donne deux textes parallèles en français et en anglais.

Deux approches complémentaires sont envisagées: d'une part méthodes statistiques (plus des règles linguistiques, en particulier pour la question des désinences en français); d'autre part recherche d'exemples dans des traductions déjà existantes.

Cette approche n'est rendu réaliste que par l'extraordinaire augmentation des capacités de traitement et de stockage d'information fournie par l'évolution de la technologie informatique.

A la jonction de la reconnaissance de la parole et de la traduction assistée, il est intéressant de mentionner l'originalité des travaux conjoints du CRIM (Roland Kuhn et al.) et du CITI (Pierre Isabelle et al.) à Montréal, pour permettre à un traducteur de dicter sa traduction plutôt que de la taper. Le texte de départ, en cours de traduction, est rendu accessible au système et sert de source d'information complémentaire au signal acoustique en langue d'arrivée. Le projet s'appelle TransTalk (voir article de Roland Kuhn à Eurospeech'95: «French Speech Recognition in an Automatic Dictation System for Translators: the TransTalk Project»).

Dans bien des domaines du traitement du langage naturel, le message est actuellement simple, comme pour la traduction: il faut plus de corpus de textes vastes et disponibles à tous (même situation qu'en parole, donc).

Complétons cette section en indiquant qu'un ensemble de contributions sur le traitement du langage naturel, rassemblées sous le titre «Survey of the State of the Art of Human Language Technology» est consultable sur «la toile» (le web) à l'adresse <http://www.cse.ogi.edu/CSLU/HLTsurvey/>.

8. Deux jolis problèmes pour terminer

Pour dire, si l'on en doutait, que les chercheurs ont encore du pain sur la planche, nous voudrions évoquer deux problèmes encore loin de leur solution générale: la téléphonie interprétée, et l'indexation automatique de séquences vidéo.

8.1 La téléphonie interprétée

Ce service, qui serait fort utile, doit permettre à deux personnes parlant deux langues différentes de converser au téléphone. Il s'agit donc «simplement» de mettre bout à bout la reconnaissance de parole (*continue, spontanée, au téléphone !*), la traduction automatique (*sans limitation de domaine a priori*) et la synthèse de parole à partir du texte (ou de concepts) (*en préservant les caractéristiques personnelles et émotionnelles du locuteur !*)

ATR au Japon, avec divers partenaires en Amérique et en Europe est l'acteur le plus avancé dans cette recherche (démonstration de dialogue interprété pour l'inscription à une conférence).

8.2 L'indexation automatique de séquences video

La diffusion numérique permettra de disposer à domicile de centaines et de centaines de programmes de télévision; il ne sera plus envisageable d'avoir un information imprimée sur ces programmes. Pour pouvoir interroger un serveur sur tous les programmes, pour pouvoir se faire faire des propositions en fonction d'un profil d'intérêts, il faut que tout ce qui sera diffusé soit indexé. Pour que tout le soit, il serait utile que ce soit automatiquement !

Or c'est un problème très difficile: à partir du son et des images, comment déterminer ne serait-ce que de quel type de programme il s'agit ?

Remerciements

A mes collègues qui m'ont fait profiter de leurs analyses et opinions sur ces questions, et qui souvent ont bien voulu mettre à ma disposition des éléments de texte :

à l'ENST, Alain Grumbach (intelligence artificielle), Henri Maître, Isabelle Bloch (image), Claudie Faure (interaction personne système, analyse de document), Martin Rajman (traitement du langage naturel)

à l'ETS, Jacques de Guise (image), R. Sabourin (signatures), Pierre Dumouchel (parole).

ACQUISITION DES CONNAISSANCES DE SPÉCIFICATIONS DES LOGICIELS À L'AIDE D'UN COLLECTICIEL: MODÉLISATION ET PROTOTYPAGE

par: Ghislain Lévesque, Département d'Informatique, UQAM,
Guy Massicotte, Directeur, Bureau de la Planification, Université du Québec,
Alain Albert, Département de Sciences administratives, UQAH

Résumé

Le développement des spécifications de logiciels est une tâche ardue qui demeure souvent incomplète à cause des méthodes et des moyens disponibles pour rassembler et documenter les connaissances des experts du domaine. L'utilisation d'une technique semi-formelle comme la modélisation orientée-objet nous enseigne maintenant qu'un modèle satisfaisant du problème à résoudre ne peut être obtenu qu'après un nombre élevé d'itérations mettant à contribution les experts des principaux domaines concernés par l'application à construire. Pour soutenir le travail de ce groupe d'expert, nous avons entrepris les premières étapes du développement d'un collectif de soutien à la documentation de logiciels selon une architecture client-serveur dans un contexte de réseau étendu du type Internet. Par le moyen de groupes de tâches électroniques, en mode asynchrone et synchrone, nous considérons qu'il sera possible d'augmenter sensiblement la qualité des spécifications par la réalisation d'un nombre plus grand d'itérations sur les scénarios d'utilisation et les modèles, et la mise à contribution d'un éventail plus large d'experts de façon à obtenir un modèle comportemental du système plus élaboré et détaillé. Dans cet atelier, nous présenterons le modèle de ce collectif et les stratégies applicables pour rencontrer les objectifs recherchés. Nous prévoyons faire les premiers tests d'un tel système avec les étudiants d'un laboratoire d'analyse de systèmes.

1. Introduction

Le développement de spécifications de logiciels qui rencontrent les exigences et caractéristiques de qualité de la norme IEEE 830-1993 [1] s'avèrent un travail long et ardu. La mise à contribution de plusieurs acteurs différents, dont des experts du domaine et des différentes méthodes et normes en usage, est souvent difficile à réaliser. De plus, un nombre élevé d'itérations dans les spécifications formulées est nécessaire pour assurer la cohérence et l'exactitude du document final à produire.

L'objectif d'une telle spécification est de décrire le problème à résoudre, d'en présenter un modèle conceptuel et de proposer les exigences qui définiront le cadre d'opération du système à implanter. Il est donc essentiel à la fois de formuler une vue globale du problème, mais aussi d'en illustrer tous les détails dont il sera important de tenir compte pour faire la conception du système. En étudiant le processus de développement de logiciel en vue de proposer une taxonomie des

méthodes, Blum [2] a confirmé les prétentions de Brooks [3] à l'effet que la plus grande difficulté à surmonter dans ce processus résidait dans la formulation d'un modèle conceptuel correct pour supporter les phases de conception et de test du logiciel. Brooks qualifie d'ailleurs ce modèle de «*hautement précis et richement détaillé*»¹. Dans une étude sur les difficultés d'apprentissage propres à chaque type de méthode de modélisation (objet, données, fonctions), Vessey [4] a également confirmé cette difficulté de proposer un modèle conceptuel simple et cohérent.

Or, deux catégories d'erreur sont généralement relevées dans ces spécifications: des erreurs dues à un manque de connaissances sur les besoins à combler et des erreurs de spécification, dues au fait de pas savoir comment spécifier adéquatement les besoins. Selon Davis [5] qui propose une telle catégorisation, la diminution du premier type d'erreur peut être obtenue par des activités de prototypage mais il arrive que certaines erreurs ne puissent être détectées avant le déploiement du système; les erreurs de spécification sont toutefois difficilement pardonnables. Dans ce même texte, il propose alors, à partir d'une revue de la littérature, la définition de vingt-quatre caractéristiques de qualité du logiciel et des approches et moyens pour en faire une mesure. Le développement d'un système d'aide, en réseau étendu, fondé sur l'élaboration progressive d'un document par un groupe de tâche, avec un outil que nous appelons un 'Collecticiel', pourrait éventuellement apporter une autre forme de réponse aux deux types de problèmes identifiés par Davis.

Dans l'exposé qui va suivre, le terme 'collecticiel' sera défini brièvement et un modèle des objets qui le constituent en sera présenté. Il deviendra ainsi possible de comprendre comment cet outil de travail collectif pourrait apporter un soutien efficace au processus d'acquisition des connaissances mené dans cette démarche de spécification. Ce modèle servira à la construction d'un prototype à tester avec des groupes d'étudiants dans le cadre d'un cours laboratoire portant sur la préparation d'une spécification de système; le prototype pourra également servir à mener des exercices de jeu de rôles propres au travail de spécification. Enfin, après avoir présenté le modèle du Collecticiel de Soutien à la Documentation de Logiciel (CSDL) nous concluons sur les possibilités de mettre en place des agents dont le rôle serait de surveiller l'application des caractéristiques de qualité de la spécification du document à produire.

¹ «highly precise and richly detailed» dans le texte anglais.

2. Définition et contexte d'application du Collecticiel

Un collecticiel est un progiciel de travail en groupe à distance offrant le soutien pour les communications électroniques asynchrone et synchrone et l'accès partagé à la documentation nécessaire à la réalisation d'un projet. Le courrier électronique constitue la base des mécanismes de communication et une architecture du type client-serveur rend possible, avec les outils de navigation connus pour Internet, la consultation des documents de référence à l'ensemble des participants, dans une équipe. Éventuellement, la visio-conférence assistée ajoutera à ces mécanismes de communication initiaux une plus grande convivialité pour le travail à deux ou plusieurs participants à la fois. Les outils de travail individuels comme le traitement de texte, le chiffrier, et les outils graphiques servant à la préparation des documents demeurent, pour le moment, à l'extérieur du collecticiel; le partage de ces informations se fait par échange de fichiers. La mise en place d'une norme pour assurer «l'interopérabilité» des systèmes devient alors une exigence de base d'une telle application. La discussion et la révision des documents se fera via le courrier électronique et d'autres mécanismes mis en place qui seront explicités plus loin.

Dans ce contexte, un CSDL centré sur la production d'un document d'analyse du problème à résoudre devient un moyen beaucoup plus souple que les moyens traditionnels employés jusqu'à maintenant (principalement des réunions en face à face) pour consulter les experts de chaque domaine concerné. De plus, cet outil assurera la conservation des documents écrits constituant les différentes versions auxquels s'ajouteront les communications échangées relativement à chacune des étapes du travail pour compléter les dossiers de justification. Ainsi, il deviendra plus facile d'expliquer les modifications apportées d'une version à l'autre du document. Sur ce plan d'ailleurs, la construction d'une structure représentant le document et les étapes de sa préparation devient un pas additionnel dans l'établissement d'une mémoire organisationnelle cohérente et structurée pour les aspects documentaires.

Par ailleurs, la mise en place d'un tel collecticiel requiert l'examen plus approfondi de toutes les phases et étapes nécessaires pour supporter le processus de spécifications. Des séquences de travail particulières (workflow) pourront y être définies en vue de structurer davantage les divers cheminements qui sont suivis pour rassembler les résultats de cette analyse. L'établissement de chacune de ces séquences rendra possible l'insertion de mécanismes appropriés, des agents par exemple,

visant à supporter la mise en application des caractéristiques de qualité auxquelles nous avons fait référence plus tôt; à titre d'exemple un agent chargé d'augmenter la traçabilité de la spécification serait sûrement apprécié. Pour des informations complémentaires sur le concept de collectif, de séquences de travail et de moteur d'activation de ces séquences, nous suggérons la consultation des références [6] et [7].

La première étape visant à proposer un modèle conceptuel de ce progiciel a été réalisée récemment en produisant une esquisse générique d'un collectif de production d'un document dont les activités de travail en groupe se réalisent par un mécanisme de réunion qui reproduit les approches courantes de travail en mode face-à-face. D'autres séquences plus spécifiques sont actuellement en voie d'exploration et pourront y être substituées ou considérées complémentaires. La réalisation de cette esquisse a été faite selon une modélisation orientée-objet, d'après la méthode OMT [8]. Les modèles obtenus vont maintenant être présentés.

3. Modèle générique du Collectif

Le processus de spécification d'un logiciel peut être arbitrairement décomposé en quatre phases: préliminaire, spécifications fonctionnelles, modélisation et prototypage, spécifications non-fonctionnelles. A la phase préliminaire, le client produit une description générale du problème à résoudre et identifie, de son point de vue, les acteurs concernés par le projet et les contraintes générales dont il faudra tenir compte. Le chef de projet et l'analyste désigné questionnent cette description pour la préciser, clarifier les frontières du système à construire, identifier les principales fonctionnalités recherchées et y désigner les acteurs requis pour entreprendre la phase de spécifications fonctionnelles.

Cette phase peut débiter par la description des fonctionnalités attendues ou le développement de scénarios représentatifs de l'utilisation prévue. Le travail en mode itératif commence réellement à cette phase comme un processus de va-et-vient entre l'analyste et les experts du domaine jusqu'à ce que les fonctionnalités soient suffisamment décrites sous la forme de spécifications pour mettre en marche le processus d'inspection. La contribution d'une gamme variée de personnes peut alors être requise et obtenue: responsable d'assurance de qualité, expert des méthodes employées, architecte de système, concepteur, responsable de tests, notamment. Le processus itératif se poursuit alors jusqu'à la stabilisation de ces spécifications fonctionnelles.

Vient ensuite la phase de modélisation et prototypage. A cette phase, il s'agit d'extraire des spécifications formulées les objets du domaine à représenter et d'amorcer la construction du modèle du problème. La construction des modèles offre une première possibilité de simuler le problème. Les interfaces personnes-machines sont alors prototypées; certaines fonctionnalités, qui sont difficiles à comprendre et représenter, ou dont la réalisation semble problématique, le sont également. La discussion sur les modèles et leur représentation s'engage ensuite jusqu'à ce que les modèles soient stabilisés. On cherche à obtenir l'interprétation qu'en font chaque personne (client, inspecteur, expert) pour en arriver à une représentation unique des spécifications fonctionnelles et à leur validation.

Enfin, la phase d'élaboration des exigences non-fonctionnelles et de l'évaluation des coûts de développement est réalisée en dernier lieu. Il est donc probable que chacune de ces phases donne lieu à des séquences de travail spécifiques et particulières. Voyons maintenant le modèle de ce «collecticiel».

3.1 Premier niveau du modèle structural

Au début de la modélisation, trois objets génériques permettent de structurer le problème; nous les appelons les objets de premier niveau. Ces objets nous amènent à proposer un modèle global du problème représenté à la figure 1.



Figure 1: Modèle global du Collecticiel

L'objectif final de toute cette activité d'analyse du problème est de produire un *DOCUMENT* de spécifications; c'est le résultat ultime de toute cette démarche. Le processus pour produire ce document est généralement organisé sous la forme d'un *PROJET*; ce dernier comporte un plan de projet, un échéancier et des ressources qui lui sont affectées. La démarche de production du document est constituée de *SÉQUENCES DE TRAVAIL* organisées en fonction des étapes du projet et du contenu du document à produire. Cette démarche, représentant la partie évolutive des activités, comporte deux types de travaux: la préparation individuelle de chaque énoncé ou proposition de spécification et sa

discussion en groupe jusqu'à l'obtention d'une version qui soit l'objet d'un consensus.

3.2 Le modèle évolutif d'une *Séquence de travail*

Le modèle d'une discussion en groupe est présenté à la figure 2. Pour arriver à comprendre et représenter une séquence de travail, nous avons d'abord construit des scénarios descriptifs. Ceux-ci décrivent le processus qui est mené par l'intermédiaire de réunions successives dont l'objectif est de revoir chacune des spécifications formulées pour les critiquer, les compléter, et les améliorer jusqu'à l'obtention d'une spécification définitive. Pour comprendre les aspects dynamiques de la situation en cause, nous avons défini les objets par lesquels une discussion prend forme dans une réunion et les avons représentés comme des objets-événements pour produire un diagramme de transition d'états.

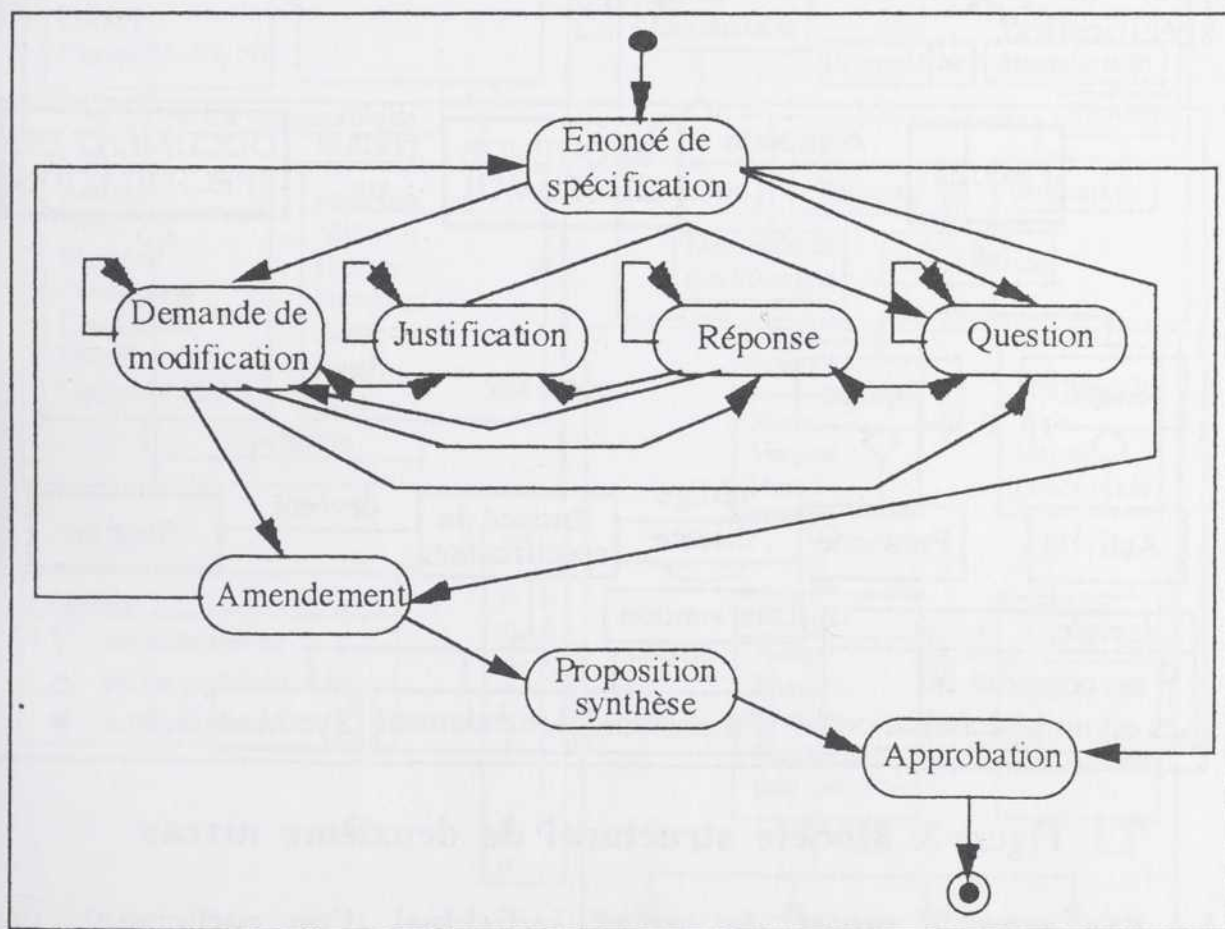


Figure 2: **Modèle évolutif d'une Séquence de travail**

Une séquence de travail débute lorsque qu'un énoncé de spécification est soumis à la discussion. Cette proposition, préparée par l'un des participants au projet, entraîne des réactions des autres participants,

analystes, experts, et inspecteurs, qui prennent la forme de question, de réponse, de justification et de demande de modification. Lorsque le débat évolue, des amendements sont proposés et une recherche de consensus est entreprise. Les amendements proposés entraînent la formulation d'un énoncé révisé, ce qui relance les discussions jusqu'à l'obtention d'une entente; lorsque le consensus est atteint, une proposition synthèse en résulte et devient la version retenue de la spécification. Par le fait même, une section du document final à produire est constituée, ce sera la première version de cette section. Au cours du processus itératif de spécification, de nouvelles versions seront produites selon le même processus.

3.3 Le modèle structural détaillé

La figure 3 montre d'une façon plus détaillée les liens entre les objets de premier niveau et les principaux objets qui structurent le problème pour la préparation d'une proposition et la production d'une spécification.

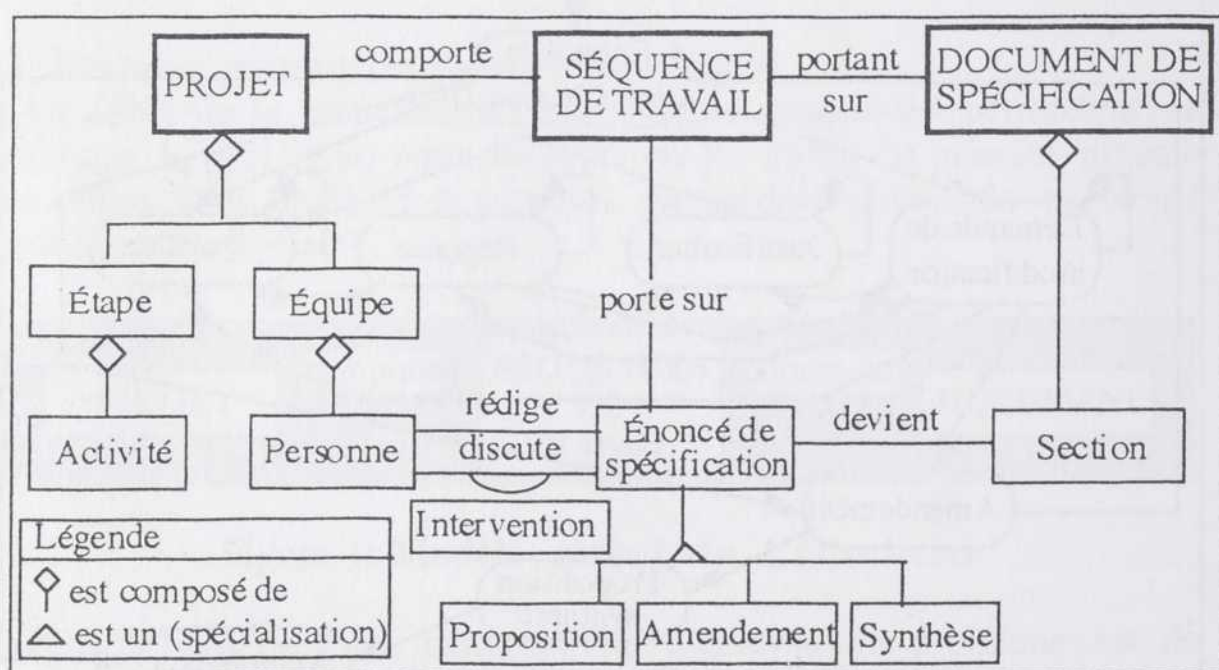


Figure 3: **Modèle structural de deuxième niveau**

Le lien entre le projet, le travail individuel d'un participant, une séquences de travail et le document se fait par l'intermédiaire des objets *Personne*, *Énoncé de spécification* et *Section*.. Quand une proposition de spécification rédigée par une personne est prête, elle devient publique et le point de départ d'une séquence de travail. Ainsi elle est soumise à la discussion des participants concernés qui interviennent,

selon le modèle de la figure 2, jusqu'à ce qu'elle soit l'objet d'un consensus et puisse devenir une section du document.

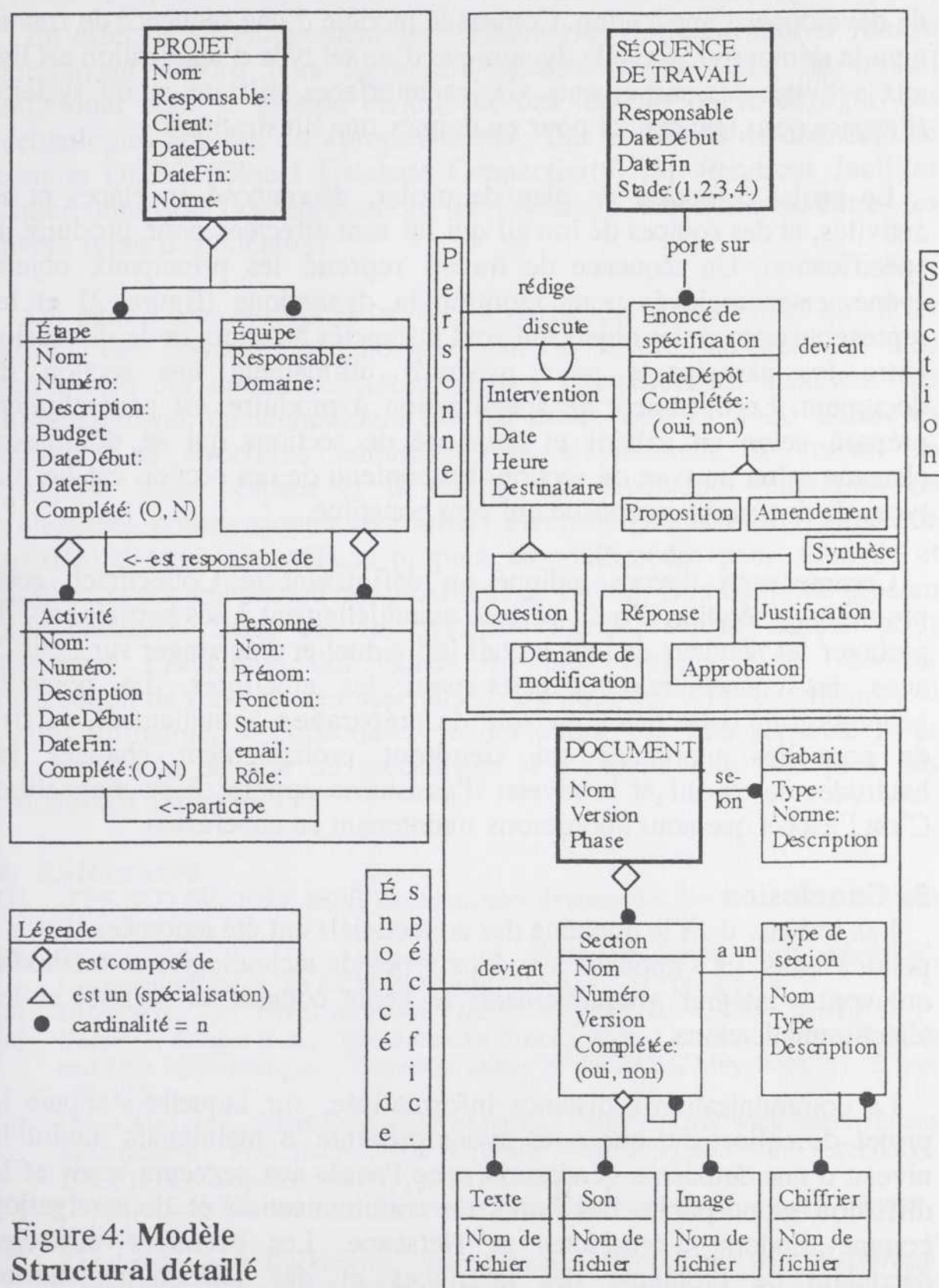


Figure 4: Modèle Structural détaillé

La figure 4 présente le détail des objets: *PROJET*, *SÉQUENCE DE TRAVAIL*, et *DOCUMENT*. Les principaux liens qui s'établissent entre les

objets y sont représentés. Les méthodes propres à chaque objet n'ont pas été incluses pour éviter de surcharger les modèles. La décomposition de chaque objet permet de comprendre la structure plus fine qui permettra de développer l'application. Comme le modèle d'une séquence de travail a pu le démontrer, toute la dynamique d'un tel type d'application est liée aux activités des participants via les interfaces-utilisateurs du système (l'espace nous manque ici pour en donner une illustration).

Le projet comporte un plan de projet, décomposé en étapes et en activités, et des équipes de travail qui lui sont affectées pour produire la spécification. Une séquence de travail reprend les principaux objets-événements employés pour montrer la dynamique (figure 2) et les représente comme les objets qui sont instanciés au cours de la discussion entre les participants pour produire ultimement une section du document. Le document de spécification à produire est généralement préparé selon un gabarit et composé de sections qui se définissent chacune selon un type de section. Le contenu de ces sections est lié aux types de document composite qui sera constitué.

Comme nous l'avons indiqué en définissant le Collecticiel, cette première génération d'outil permet essentiellement à ses participants de partager les résultats de leur travail individuel et d'échanger sur ceux-ci avec des conseillers et experts pour les améliorer. La nouvelle génération de collecticiel, en voie de préparation actuellement, offrira de nouvelles approches qui viendront probablement changer les habitudes de travail et le niveau d'assistance apporté à sa réalisation. C'est l'aspect que nous aborderons maintenant en conclusion.

5. Conclusion

Les travaux dans le domaine des collecticiels ont été amorcées il y a à peine 15 ans. Ils s'appuient sur deux types de technologies et méthodes qui vont s'intégrer graduellement: le génie cognitif et logiciel et les télécommunications.

La communication à distance informatisée, sur laquelle s'appuie le projet de collecticiel que nous avons présenté, a maintenant atteint le niveau d'une deuxième génération avec l'accès aux serveurs www et la diffusion grand-public des outils de communication et de navigation comme Eudora, LotusNotes et Netscape. Les premiers obstacles découlant de problème des interfaces et des télécommunications informatisés sont donc résolus. Le génie cognitif et logiciel, par le développement d'interfaces graphiques de plus en plus conviviales a contribué grandement à la levée de ces obstacles. Les nouvelles

génération de collecticiels viseront maintenant une plus grande intégration des outils individuels pour la réalisation de tâches en groupe.

En effet, le développement des technologies Intranet/Internet rendra dorénavant encore plus facilement accessible les outils de travail individuel pour le groupe en offrant des moyens de se dégager des technologies fermées ou «propriétaires». Les standards de connectivité comme ODBC (Object Database Connectivity) [9] viendront faciliter l'interconnexion des applications et l'échange des données entre les divers types d'application.

On peut d'ores et déjà entrevoir, à partir de là, les effets qu'ont et auront ces nouvelles technologies sur le travail en général. Les entreprises ont commencé récemment à récupérer les bénéfices des investissements en technologie fait au cours des 15 à 20 dernières années. Les structures administratives s'aplatissent, le travail se virtualise dans l'espace et le temps. La prochaine génération de collecticiel, pourra ajouter au courrier électronique, la visio-conférence assistée, et tous les artéfacts propres au multimédia pour soutenir et faciliter la documentation des processus de travail. Des outils de plus en plus puissants, basés sur les systèmes agents, seront conçus pour supporter l'exécution des tâches les plus complexes et faciliter l'exécution de travaux très spécialisés. On peut dès lors se demander si l'outil lui-même, le collecticiel ne deviendra pas indispensable pour répondre à la pénurie de spécialistes et d'experts que l'on peut dès maintenant anticiper.

4. Références

- [1] IEEE STD 830-1993, IEEE Recommended Practice for Software Requirements Specifications,
- [2] Blum Bruce I., 'A Taxonomy of Software Development Methods', Communications of the ACM, November 1994, 37, 11, pp. 82-94.
- [3] Brooks, F.P. Jr, 'No Silver Bullet', IEEE Computer, 20, 4, 1987, pp. 10-19.
- [4] Vessey I, Conger S.A., 'Requirements Specification: Learning objects, Process and Data Methodologies', Communications of the ACM, May 1994, 37, 5, pp. 102-113.
- [5] Davis Alan et al., 'Identifying and Measuring Quality in a Software Requirements Specification', First International Software Metrics Symposium Proceedings, 1993, IEEE Computer Society, pp. 141-152.
- [6] Coleman C. et Khanna R., 'Groupware Technology and Applications', Prentice-Hall, 1995, 576 p.
- [7] Khoshafian S et Buckiewicz, 'Introduction to Workflow, and Workgroup Computing', Wiley, 1995, 376 p.
- [8] Rumbaugh J., Blaha M. Premerlani W., Eddy F. et Lorensen W., 'OMT Modélisation et conception orientées objet', Masson, Paris, 1995, 515 p.
- [9] Jepson B., World Wide Web Database Programming for Windows™ NT, Woley, N.Y., 1996, 352 p.

La segmentation d'images par système à base de connaissances

Daniel Crevier, Abdenour Bouzouane, Guturu Parthasarathy,
Richard Lepage

*Laboratoire d'Imagerie, de Vision et d'Intelligence Artificielles (LIVIA)
École de Technologie Supérieure,
Université du Québec, Montréal*

Résumé

Nous exposons une méthode de segmentation d'images basée sur l'exploitation de connaissances formulées à l'aide de règles si-alors. Ces connaissances permettent d'effectuer la segmentation en subdivisant l'image en contours fermés qui correspondent à des objets ou parties d'objets. Il sera alors possible au système à base de connaissances, dans une étape ultérieure, de les comparer à une description symbolique et de déterminer la nature de l'objet. Cet article porte sur la segmentation proprement dite. Dans un premier temps, on extrait des segments de contour consistant en arcs de cercle et lignes droites, qui sont ensuite regroupés en contours fermés. Chaque contour a son origine dans un segment choisi par un mécanisme d'attention contrôlé par un système de productions. Le contour est choisi de façon à minimiser la somme pondérée des longueurs des segments impliqués. Les facteurs de pondération permettent de tenir compte d'informations fournies par l'organisation perceptuelle, ainsi que par les caractéristiques des groupes de pixels séparant les segments impliqués. C'est ainsi que si un segment est impliqué avec le segment d'origine dans une relation perceptuellement significative comme le parallélisme ou la symétrie, un facteur de pondération plus petit que l'unité lui sera appliqué, de façon à augmenter ses chances de faire partie du contour le plus court. Des facteurs de pondération semblables sont appliqués aux segments reliés au segment d'origine par une ou des régions de pixels uniformes. Après la fermeture de chaque contour, le système à base de connaissances choisit le segment d'origine du prochain contour.

1. Introduction

La segmentation d'images consiste à partitionner une image numérisée en un ensemble de régions mutuellement exclusives qui couvrent l'ensemble de l'image. Son objectif est de décomposer l'image en parties qui ont du sens dans le cadre d'une application particulière, consistant par exemple en la séparation d'un objet du reste de l'image. Les connaissances requises pour ce faire ont trait à l'apparence des objets, et aussi aux stratégies d'analyse permettant de les identifier dans une image. L'utilisation de systèmes experts comme aide à la segmentation n'est pas nouvelle: Levine et Nazif [1], par exemple, l'ont déjà explorée il y a plusieurs années. Une revue des travaux effectués depuis apparaît en [2]. La nouveauté de notre méthode consiste à incorporer au système expert des connaissances portant sur l'organisation perceptuelle. Inspirée de la vision humaine, l'organisation perceptuelle [3] se

sert de la détection de caractéristiques significatives pour fixer l'attention visuelle et interpréter le contenu de l'image. Une caractéristique est dite significative si elle a une faible probabilité de se produire accidentellement, et qu'elle est préservée lors de la projection de trois dimensions à deux dimensions. De telles caractéristiques incluent le parallélisme, la symétrie, la colinéarité, et les jonctions d'arête.

Les caractéristiques généralement acceptées d'une bonne segmentation d'image sont les suivantes:

- les régions sont uniformes en brillance, couleur et texture
- des régions voisines ont des valeurs différentes des attributs pour lesquels elles sont homogènes
- les frontières entre les régions correspondent à des lieux de fort gradient de brillance
- les frontières entre les régions sont lisses
- les intérieurs des régions sont simples et dépourvus de petits trous.

Il existe plusieurs façons d'obtenir des segmentations correspondant à ces critères. Elles sont basées soit sur la croissance de régions, soit sur l'extraction des frontières et la fermeture de contours. La référence [4] en donne une bonne revue.

À ces critères, l'organisation perceptuelle ajoute les contraintes suivantes:

- les paires de segments symétriques, colinéaires, convergents ou parallèles doivent avoir une probabilité accrue d'appartenance à un même contour.
- les objets doivent être partitionnés comme la vision humaine le ferait, c'est-à-dire de façon simple et compacte. Par exemple, les contours extraits de l'image d'une tasse devraient être ceux du corps de la tasse, de son ouverture et de l'anse. Le contour correspondant à la silhouette de la tasse ne devrait pas être extrait. Les techniques conventionnelles de suivi de contour ont du mal à respecter cette dernière contrainte, car elle implique souvent la préférence de contours contenant des vides à celle de contours pleins.

Il n'est pas facile d'incorporer ces contraintes à des méthodes de segmentation conventionnelles: nous proposons ici une façon de le faire. Notre méthode unifiera de plus les techniques basées sur la croissance de régions et celles basées sur l'extraction de contours. Elle pallie aussi une faiblesse de la

segmentation basée sur les contours, souvent mise en faute par les artefacts des techniques d'extraction d'arêtes, qui ont tendance à laisser des vides aux points de jonction.

2. Description de la méthode

La Figure 1(a) représente une image typique des contours d'un cube tels qu'on pourrait les obtenir à l'aide d'un détecteur d'arêtes conventionnel comme la méthode de Canny. On suppose dans la figure que les pixels des arêtes ont été liés en segments approximatés par des droites (nous utilisons pour ce faire une procédure décrite en [7], qui permet en plus d'approximer des courbes par des arcs de cercle). Comme c'est typiquement le cas en pratique, la procédure a laissé dans la Figure 1 des vides dans les arêtes séparant les facettes du cube, ainsi qu'au point de jonction des arêtes internes. Le seul contour fermé continu de l'image est celui qui relie les arêtes extérieures du cube. Idéalement, cependant, cette image devrait être segmentée en trois régions représentant chacune une facette du cube.

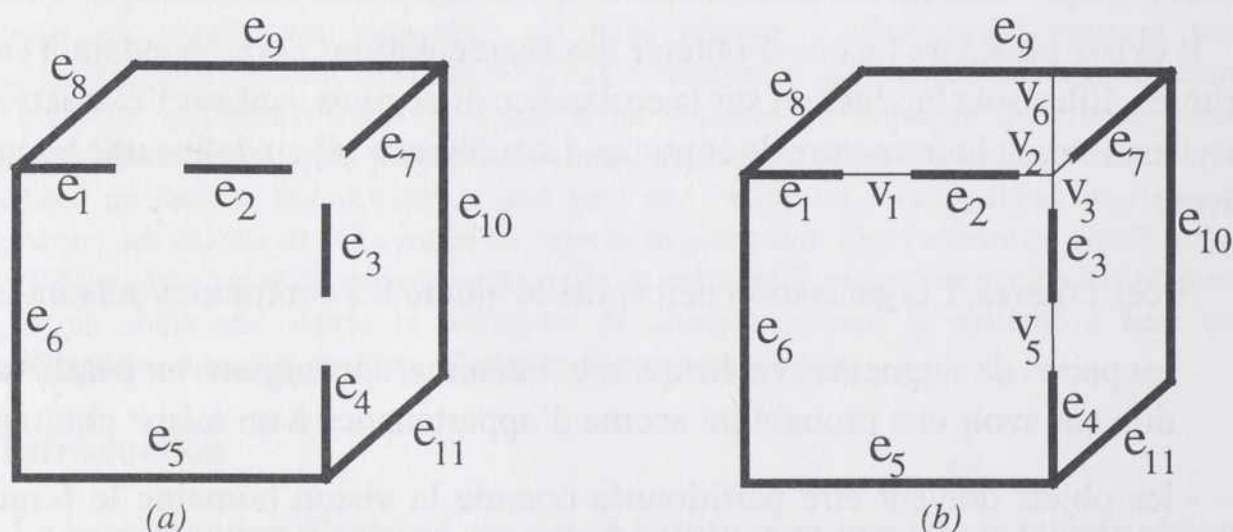


Figure 1 : illustration de la méthode sur un exemple simple. Les vides dans les contours sont comblés par des segments dits virtuels, apparaissant en traits fins en (b). Certains de ces segments, comme v_1 à v_5 , seront choisis judicieusement, alors que d'autres, comme v_6 , représenteront de fausses relations perceptuelles (dans le cas de v_6 , une jonction en T entre les segments e_3 et e_9).

Pour ce faire, les vides entre les segments d'arête e_1 et e_2 , e_3 et e_4 , ainsi que ceux correspondant à la jonction de e_2 , e_3 , et e_7 doivent être remplis. Nous décrivons en [6] une telle procédure, qui consiste en une modification de certains mécanismes proposés par Lowe [3]. Il sera alors possible d'insérer dans l'image des segments potentiels ou virtuels pour représenter la possibilité

de fermer ces vides. Cet ajout ne réglerait cependant pas le problème de l'extraction des contours les plus plausibles: en particulier, une méthode de fouille de graphe identifierait encore le contour extérieur du cube comme un circuit fermé. Dans des images plus compliquées, le nombre de contours fermés possibles dans l'image croîtrait exponentiellement avec le nombre de segments d'arête. De plus, le groupement perceptuel pourrait très bien produire de faux segments virtuels: le segment e_3 , par exemple, pourrait être prolongé jusqu'au segment e_9 , ce qui produirait une jonction en T et induirait la possibilité d'encore plus de contours fermés. Le problème de l'extraction des contours à partir de dessins au trait est l'un des plus anciens en vision artificielle [5]. Quand la scène analysée contient une variété de formes limitée, des techniques de propagation de contraintes peuvent limiter les choix possibles. Il n'existe cependant pas de méthode de propagation de contraintes pouvant s'appliquer à des formes quelconques.

De plus, d'autres informations sont normalement disponibles, et il est avantageux de les utiliser dans des images plus complexes. Il est en particulier possible d'effectuer la segmentation de l'image en procédant par croissance de régions (en réunissant les pixels semblables) plutôt que par détection d'arêtes (séparation de pixels différents). Ces deux méthodes peuvent en général s'appliquer sur une même image, et donnent souvent des résultats différents. Notre méthode permet la combinaison des informations provenant respectivement des lignes et des régions.

2.1 Calcul de l'adjacence des segments aux régions

La solution que nous proposons consiste à activer un segment choisi intentionnellement, et à trouver par une méthode de fouille heuristique le contour fermé le plus court comprenant ce segment. Faisant maintenant référence à la Figure 2, supposons que nous voulions extraire la facette du cube adjacente au segment e_6 , après avoir identifié par détection d'arêtes les segments e_1 à e_{11} . Supposons de plus que l'image ait été segmentée par croissance de régions tel qu'illustré dans la figure. (Pour fins de simplification, nous ignorerons ici le fond de l'image. En général, ce fond sera lui aussi segmenté en régions, dont certaines seront adjacentes aux segments formant les contours du cube.) De façon à faire coïncider le plus possible les frontières des régions avec les segments d'arête, on a interdit pendant la croissance de régions la traversée des arêtes. On peut minimiser la croissance des régions à travers les vides par l'adoption de méthodes de croissance de régions dite par

relation de centroïde [4]: ces méthodes incorporent un nouveau pixel à une région seulement si les caractéristiques du pixel sont suffisamment rapprochées des caractéristiques moyennes de la région. De telles méthodes sont moins sujettes à l'extension indéfinie des régions à travers des vides ou des arêtes de faible intensité que les méthodes de croissance dites individuelles, dans lesquelles un nouveau pixel est incorporé à une région si ses caractéristiques se rapprochent de celles de ses voisins déjà incorporés à la région.

Dans la segmentation résultante, les segments adjacents à la même facette visible du cube que e_6 ont au moins une région en commun avec e_6 , alors que les autres segments d'arête n'en ont aucune. Il est possible d'exploiter ce fait

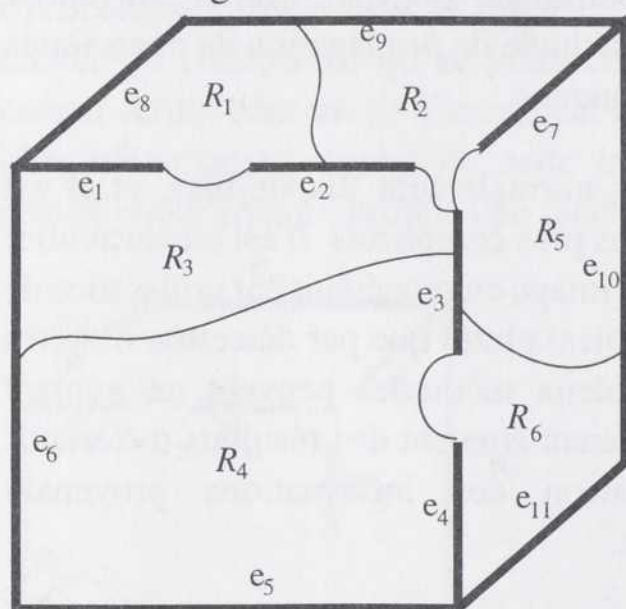


Figure 2 : Une autre source d'information pour la segmentation d'images est l'adjacence de régions communes aux segments.

lors de la fouille pour le contour le plus court en réduisant artificiellement les longueurs des segments bordant la facette, ainsi que les distances correspondant aux vides entre ces segments. On pourra réduire ces distances par un facteur proportionnel au pourcentage de leur longueur pour lesquels ces segments sont adjacents aux mêmes régions que e_6 . Cette réduction augmentera la probabilité que le contour le plus court soit celui formé par les segments e_1 à e_6 .

Notons cependant que cet ajustement ne changera pas la position des extrémités des arêtes ou la forme des objets contenus dans l'image. Il ne s'agit que de modifier temporairement les attributs *longueur* des segments dans la mémoire de travail du système à base de connaissances, ce qui affectera le calcul de la longueur totale d'un contour pour fins d'évaluation dans la fouille.

La technique utilisée devra cependant disposer d'assez de subtilité pour ignorer les fausses informations produites par le phénomène de croissance des régions à travers les vides, qu'il est en général impossible d'éliminer complètement. Elle devra aussi pouvoir tenir compte du fait que les orientations données aux lignes sont arbitraires, ainsi que les notions de droite et de gauche qui leur sont associées. La méthode suivante possède les attributs

désirés. Définissons les matrices d'adjacences gauche et droite G et D comme suit. S'il y a n_l lignes et n_r région, alors D est une matrice de taille $n_l \times n_r$ telle que:

$$d_{ij} = \begin{cases} p & \text{si la région } j \text{ est adjacente à la ligne } i \text{ du côté } \textit{droit} \text{ sur } p \text{ pourcent de sa longueur} \\ 0 & \text{autrement} \end{cases}$$

De même,

$$g_{ij} = \begin{cases} p & \text{si la région } j \text{ est adjacente à la ligne } i \text{ du côté } \textit{gauche} \text{ sur } p \% \text{ de sa longueur} \\ 0 & \text{autrement} \end{cases}$$

De cette façon, si les régions ne se chevauchent pas et couvrent toute l'image, la somme des éléments de chaque ligne ou chaque colonne de D et G donnera 100%. Si nous dénotons maintenant par c_{ik} le *coefficient d'adjacence de régions* des lignes i et k , alors ce coefficient se calculera comme suit:

$$c_{ik} = \frac{1}{400} \left| \sum_{j=1}^{n_r} [(d_{ij} + d_{kj})\alpha_{ijk} + (g_{ij} + g_{kj})\beta_{ijk} - (d_{ij} + g_{kj})\gamma_{ijk} - (d_{kj} + g_{ij})\delta_{ijk}] \right| \quad (1)$$

où les barres verticales signifient la valeur absolue et:

$$\alpha_{ijk} = \begin{cases} 1 & \text{si } d_{ij} \neq 0 \text{ et } d_{kj} \neq 0 \\ 0 & \text{autrement} \end{cases}$$

$$\beta_{ijk} = \begin{cases} 1 & \text{si } g_{ij} \neq 0 \text{ et } g_{kj} \neq 0 \\ 0 & \text{autrement} \end{cases}$$

$$\gamma_{ijk} = \begin{cases} 1 & \text{si } d_{ij} \neq 0 \text{ et } g_{kj} \neq 0 \\ 0 & \text{autrement} \end{cases}$$

$$\delta_{ijk} = \begin{cases} 1 & \text{si } g_{ij} \neq 0 \text{ et } d_{kj} \neq 0 \\ 0 & \text{autrement} \end{cases}$$

Cette technique revient à calculer les coefficients d'adjacence d'une paire de lignes en additionnant les pourcentages de leur longueur pour lesquels les lignes partagent des régions du même côté, et en soustrayant de cette somme les pourcentages pour lesquels elles partagent des régions de côtés différents. De cette façon, si une région est adjacente aux deux côtés d'une ligne, ses contributions au coefficient d'adjacence se réduisent l'une l'autre. Il faut prendre la valeur absolue du résultat parce que les lignes sont orientées de façon arbitraire, et il n'est pas possible *a priori* de prédire si les termes positifs ou négatifs seront les plus significatifs.

2.2 Prise en compte de l'organisation perceptuelle

Nous avons déjà tenu compte des relations perceptuelles comme la

colinéarité, et les jonctions en L et en T, en insérant des segments virtuels entre les extrémités pertinentes des segments impliqués. Les segments virtuels, cependant, ne permettent pas de tenir compte du parallélisme ou de la symétrie. Nous avons donc choisi, comme dans le cas des régions, de réduire la longueur des segments impliqués dans une relation de parallélisme ou de symétrie avec le segment de départ. La longueur sera réduite par un facteur proportionnel à la qualité de la relation en question.

Soit l_1 la longueur du segment le plus long de la paire: ce segment servira de référence. Les quantités utilisées dans la mesure du degré de parallélisme sont définies dans la Figure 3.

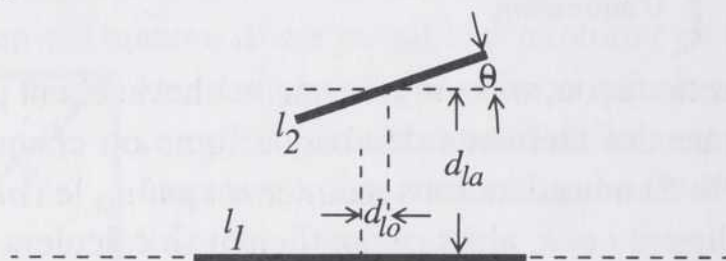


Figure 3 : Quantités utilisées pour la mesure du degré de parallélisme de deux segments.

Nous considérerons que deux segments forment une paire de parallèles perceptuellement significative si elles présentent les caractéristiques suivantes:

- l'angle θ entre les deux lignes est petit.
- la valeur absolue de la distance latérale d_{la} entre les lignes est petite par rapport à la longueur moyenne des lignes: cela revient à dire que la valeur absolue de $d_{la}/(l_1+l_2)$ doit être petite.
- la valeur absolue de $d_{lo}/(l_1+l_2)$ est petite, où d_{lo} est la séparation longitudinale illustrée à la Figure 3.
- la signification perceptuelle diminue si les deux lignes diffèrent largement de longueur. Si nous définissons la variable r comme le quotient des longueurs des segments

$$r = \frac{l_2}{l_1}, \quad (2)$$

nous exigerons que la valeur absolue de son logarithme p soit petite. Le facteur de pénalité totale $\tilde{p}$ correspondant à une paire de parallèles sera donc:

$$\tilde{p} = k_1 \frac{|d_{lo}|}{l_1 + l_2} + k_2 \frac{|d_{la}|}{l_1 + l_2} + k_3 |\theta| + |\rho| \quad (3)$$

où k_1 , k_2 et k_3 sont des poids sans dimension physique servant à ajuster l'importance relative des différents facteurs. Comme $\tilde{p}$ ne met en jeu que des rapports de longueurs et un angle, cette quantité est indépendante de l'échelle, et prend la valeur zéro pour deux segments superposés de même longueur. $\tilde{p}$ peut cependant devenir arbitrairement grand pour des paires de segments quelconques: il est donc contenu dans l'intervalle $[0, \infty]$. Donc la quantité

$$p = \frac{2}{\pi} \text{atan}(\tilde{p}) \quad (4)$$

sera contenue dans l'intervalle $[0, 1]$. La valeur $(1-p)$ pourra servir de facteur multiplicatif pour ajuster la longueur d'un segment de la paire si l'autre segment est le segment de départ. On pourra, en première approximation, étendre cette méthode à des paires de segments comprenant un ou deux arcs de cercles si ces arcs sont remplacés pour fins de calcul par des segments droits de même longueur, dont les points milieu ont les même position et orientation que celles des arcs.

Une relation semblable, impliquant deux paires de segments formant des jonctions en L, permet de détecter certaines symétries. Nous ne pouvons cependant en donner le détail faute d'espace.

2.3 Mécanisme de contrôle

Le système à base de connaissances choisit le segment de départ du prochain contour de la façon suivante. Les segments sont initialement triés par ordre de longueur, et placés dans une liste. Le segment le plus long est choisi comme point de départ du premier contour. Le point de départ du contour suivant est le segment suivant de la liste qui ne fait pas déjà partie d'un contour, et ce jusqu'à épuisement de la liste.

3. Résultats

La Figure 4(a) représente une image monochrome prise dans un de nos laboratoires. Les arêtes extraites de cette image par filtrage de Canny apparaissent à la figure 4(b). La figure 4(c) représente une superposition des arêtes (en blanc) et des régions extraites par croissance de région limitée par les arêtes. Les niveaux de gris sont uniformes dans les régions et correspondent à la brillance moyenne dans chaque région. Il est intéressant de constater que cette dernière image, bien que considérablement simplifiée par

rapport à l'image originale, est beaucoup plus facile à interpréter visuellement que l'image en (b), qui ne contient que les arêtes. Cette observation confirme que les régions remplissant les vides entre les arêtes contiennent une quantité importante d'informations, que notre méthode permet d'utiliser pour interpréter l'image. Cette dernière a une résolution de 256 x 256 pixels, et contient 204 segments d'arêtes et 120 régions. Les figures 4(d) à (f) représentent des contours extraits par la méthode décrite plus haut. On remarquera la présence de vides importants dans tous ces contours: ces vides ont été comblés durant l'analyse par des segments virtuels que nous avons choisi d'omettre des résultats. Les avantages de l'utilisation du partage des régions sont particulièrement apparents dans le cas de la figure 4(e), pour laquelle des méthodes ne reposant que sur le suivi de contours ont tendance à inclure dans un même contour des segments provenant des différents rectangles concentriques formés par l'écran et les contours du moniteur. Notons que même si les trois contours présentés sont rectangulaires, la méthode n'effectue aucune supposition a-priori sur la forme des contours.

4. Conclusion

Nous avons présenté les éléments d'un système à base de connaissances qui effectue la segmentation d'images en objets ou parties d'objet. Les connaissances de ce système portent sur la façon d'exploiter les relations de parallélisme et de symétrie des segments de contour, de même que leur adjacence à des régions communes, pour obtenir des contours correspondant à ceux de l'intuition humaine, et pouvant servir de support à la reconnaissance d'objets.

5. Bibliographie

- [1] M. D. Levine, A.M. Nazif, «Rule-Based Image Segmentation: a Dynamic Control Strategy approach.» *Computer vision, Graphics and Image Processing*, Vol. 32, Oct. 1985, pp. 104-126.
- [2] D. Crevier, R. Lepage, «Knowledge-Based Image Understanding Systems: a Survey.» *Computer Vision and Image Understanding (CVIU)*, sous presse.
- [3] D.G. Lowe, *Perceptual Organization and Visual Recognition*. Kluwer Academic Publishers, Boston, Dordrecht, Lancaster, 1985.
- [4] R.M. Haralick, L.G. Shapiro, *Computer and Robot Vision, Volume I*, Addison-Wesley, 1992, pp. 509-554.
- [5] D. Waltz, Understanding Line Drawings of Scenes with Shadows, in *Psychology of Computer Vision*, P.H. Winston ed., MIT Press, Cambridge, MA, 1975.

- [6] D. Crevier, *On the probabilities of random occurrence of edge configurations in digital images*. Rapport technique, Département de génie électrique, ÉTS 1996.
- [7] D. Crevier, H.J. Yoo, A new method for Straight Line and Circular Arc Extraction, *SPIE Proceedings on Intelligent Robots and Computer Vision XIII: Algorithms and Computer Vision*. Volume 2353, 1994, pp. 154-162.

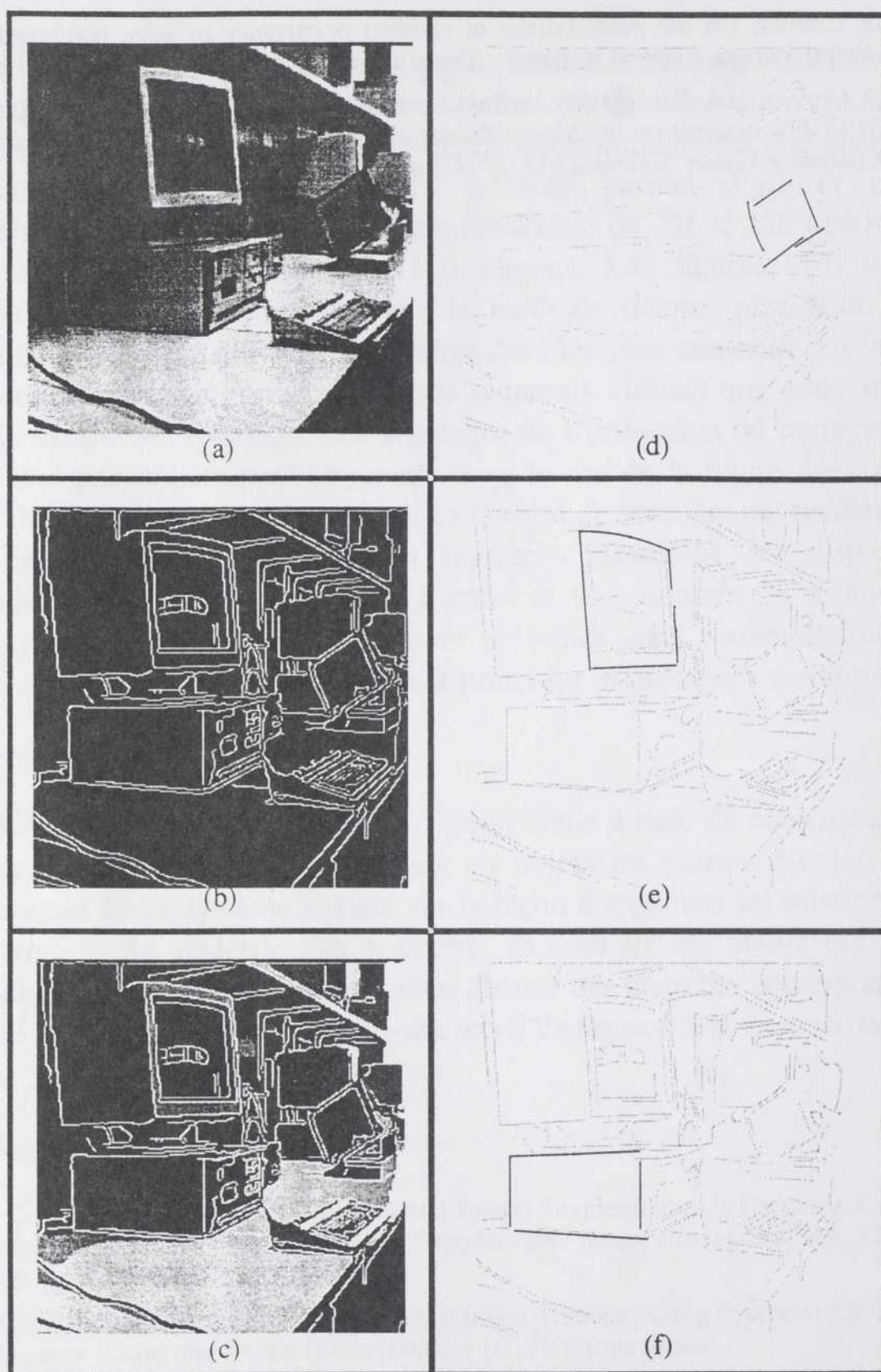


Figure 4 : Image originale (a); arêtes extraites par filtrage de Canny (b); superposition des arêtes et des régions (c); trois contours extraits (d, e, f).

Le traitement optique en visionique

Roland Schaefer et Denis Gingras

Traitement de l'information, Institut national d'optique

Sainte-Foy, Québec, Canada, G1P 4N8

[adresse électronique: gingras@ino.qc.ca]

Résumé

Le traitement optique peut jouer un rôle majeur en visionique. Dans notre article, nous présenterons les différentes structures optiques et les technologies de base qui conviennent aux applications reliées à la visionique en temps réel. Nous étudierons également les problèmes soulevés par le traitement optique, ses limites et ses possibilités ainsi que les défis à venir. Nous résumerons quelques exemples d'utilisations de telles structures dans les systèmes de vision artificielle mis au point à l'Institut national d'optique et dans d'autres laboratoires.

1. Introduction

Les technologies optiques, qui ont connu d'énormes progrès au cours des dernières années, sont omniprésentes dans notre société. L'optique et la photonique se retrouvent à tous les niveaux du traitement de l'information, notamment dans la détection et la mesure, la transmission, la mémorisation, la visualisation. Dans cet article, nous étudierons plus précisément les aspects du traitement de l'information pour lesquels l'optique et la photonique peuvent être utilisées au mieux, en mettant l'accent sur les applications reliées à la visionique.

L'optique et la photonique regroupent deux principaux domaines: le calcul optique numérique et le traitement optique analogique. En traitement numérique, les propriétés de commutation des phénomènes physiques optiques ou optoélectroniques ont permis la fabrication de commutateurs de base servant à la conception de portes et de fonctions booléennes plus complexes, qui produiront, à terme, des éléments de traitement plus rapides et la génération de signaux d'horloge. Dans ce domaine, l'accent est porté sur l'amélioration des caractéristiques, telles la réduction du délai et de l'énergie de commutation. Des commutateurs photoniques (ex. les composants électro-optiques autorégulés SEED [1] ou les portes logiques à fibres optiques SAGNAC [2]), caractérisés par des délais de commutation de l'ordre de la picoseconde et des énergies de l'ordre du femtojoule par bit, ont récemment été conçus. Malgré les progrès considérables réalisés au cours des dernières années, les technologies reliées au traitement numérique

n'ont encore atteint ni la maturité, ni les niveaux de production en série nécessaires pour concurrencer l'électronique en ce qui a trait aux processeurs de haut niveau en visionique.

En traitement analogique, on tire plutôt profit du parallélisme massif de la propagation de la lumière, de son potentiel à réaliser des topologies d'interconnexions denses et sans interférences (aucune diaphonie), ainsi que de ses propriétés fondamentales de Fourier. Par exemple, le spectre complexe bidimensionnel de Fourier d'un objet à traiter, est obtenu au foyer d'une lentille convergente simple. Au cours des 30 dernières années, ces propriétés inhérentes ont été exploitées afin de réaliser diverses transformations, dont la transformée de Fourier, ainsi que de nombreuses opérations de filtrage et de traitement morphologique. Ce type de traitement analogique s'avère très utile au cours des premières étapes de la vision artificielle, pendant lesquelles des opérations fondamentales de transformation et de nettoyage sont appliquées sur les images. Ces opérations visent principalement au traitement et au rehaussement des images, à la reconnaissance des formes, au filtrage adapté, à l'évaluation et à la détection des mouvements ainsi qu'à la détection des contours.

Dans cet article, nous décrirons les structures fondamentales du traitement analogique et cernerons certains avantages qu'offre le traitement optique dans la réalisation des primitives de la vision artificielle. Nous évoquerons de nombreux exemples d'utilisation de processeurs optiques dans les systèmes de vision artificielle et nous ajouterons quelques commentaires quant à l'avenir prévisible de ce domaine.

2. Structures optiques pour le traitement analogique en visionique

Plusieurs technologies et structures diverses reliées au traitement optique analogique ont été présentées au fil des ans. Le traitement du signal acousto-optique, par exemple, relève d'une technologie qui a atteint son plein développement et qui a connu un certain succès dans les applications militaires notamment le traitement des signaux radar. Le lecteur intéressé par les processeurs acousto-optiques unidimensionnels pourra se référer au document [3] de notre bibliographie. Quant aux applications en visionique traitées dans cet article, nous ne présenterons que les principales structures bi-dimensionnelles. Toutes ces dernières sont construites sur le modèle de la transformée de Fourier produite par les effets de diffraction de la lumière.

2.1 La transformée optique de Fourier

La transformée de Fourier est probablement la fonction la plus utilisée dans les systèmes de traitement des signaux et de vision. Il s'agit d'une fonction lourde en calculs qui peut être réalisée rapidement et efficacement à l'aide d'un système optique [4]. La figure 1 représente la structure la plus couramment utilisée pour cette opération. Le faisceau laser est élargi et collimé; il traverse le plan d'une image où il est modulé dans l'espace. Le milieu de modulation, appelé modulateur spatial de lumière (MSL), peut être passif (comme une diapositive) ou actif (comme un dispositif à cristaux liquides). Le signal d'entrée bidimensionnel $g(x, y)$ y est codé; un tel codage se fait spécifiquement en mode d'amplitude, bien qu'une modulation de phase pour le signal d'entrée soit également utilisée par moments [5]. Dans le plan focal arrière de la lentille, l'amplitude complexe du signal optique est:

$$G(u, v) = A \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \cdot e^{-j \frac{2\pi}{\lambda f} [ux + vy]} dx dy, \quad (1)$$

où λ représente la longueur d'onde de la lumière cohérente, f , la longueur focale de la lentille et A , une constante complexe. Une simple modification dans les coordonnées donne la transformation de Fourier (TF) en fonction des coordonnées spatiales de sortie x' et y' de sorte que:

$$u = \frac{x'}{\lambda f}, \quad v = \frac{y'}{\lambda f}. \quad (2)$$

En pratique, la cadence d'un tel système est limitée par la cadence de rafraîchissement du MSL actif dans le plan d'entrée et par la rapidité de lecture de la matrice de détection dans le plan de Fourier. Il est à noter que la TF est invariante au décalage dans le plan d'entrée, ce qui constitue une qualité importante pour la détection de cibles multiples.

2.2 Le corrélateur à plan fréquentiel

Le système TF décrit ci-dessus peut être monté en cascade: c'est-à-dire que l'image résultante (la TF) est transmise à un second étage, ajusté de manière à en effectuer la transformée de Fourier inverse. Cette idée a été

proposée pour la première fois par Vander Lugt en 1964 [6]. Un corrélateur à plan fréquentiel est alors créé en plaçant un filtre dans le plan de Fourier (plan TF) d'un tel système, comme l'illustre la figure 2. Lorsque la fonction du filtre utilisée répond à la fonction complexe conjuguée ($H^*(u,v)$) de la TF de la fonction de filtrage $h(x,y)$, le signal détecté à la sortie du second étage est exprimé par:

$$\begin{aligned} c(x,y) &= FT^{-1} [G(u,v) \cdot H^*(u,v)], \\ &= g(x,y) \otimes h(x,y), \end{aligned} \quad (3)$$

ce qui correspond mathématiquement à l'opération de corrélation.

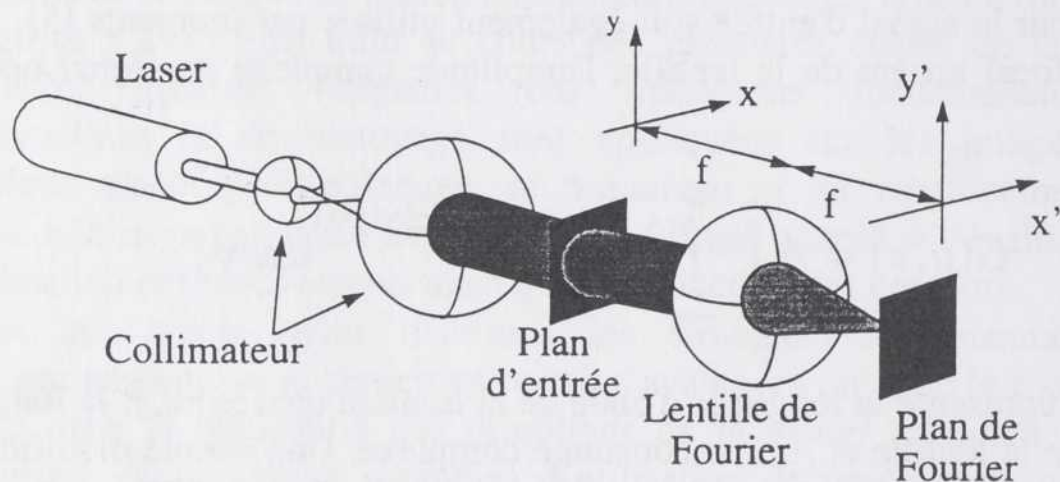


Figure 1. Système optique de transformation de Fourier.

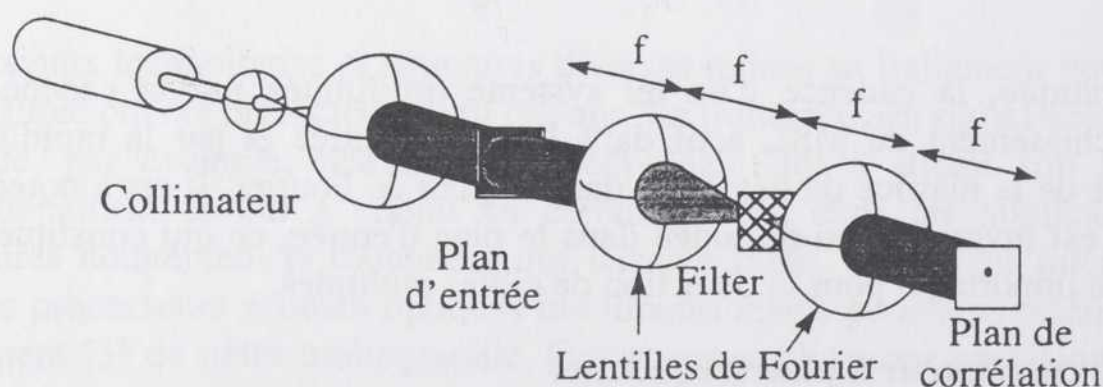


Figure 2. Corrélateur à plan fréquentiel

Il faut noter que le codage du filtre $H^*(u,v)$ doit spécifiquement se faire à la fois en modulation d'amplitude et en modulation de phase afin de représenter complètement la fonction TF du filtre $h(x,y)$. De nombreux algorithmes de conception de filtres ont été développés pour produire des filtres de phase analogiques [7], des filtres de phase binaires [8], des filtres d'amplitude et de phase ternaires [9], etc. Beaucoup d'efforts ont été investis afin de créer des filtres invariants aux distortions géométriques (invariants à l'échelle et aux rotations) [10] et de coder des fonctions multiples de référence sur un même filtre (appelé alors filtre composite) [11] permettant ainsi l'utilisation d'un seul filtre au lieu de plusieurs, appliqués en séquence. Formellement, le filtre le plus général nécessite une modulation complexe. Les travaux actuels visent à réaliser cela, tout en intégrant les contraintes de modulation des modulateurs existants [12].

2.3 Le corrélateur à transformée jointe

Le corrélateur à transformée jointe [13,14], illustré à la figure 3, a été proposé vers la fin des années 60. Son architecture comporte deux voies: l'une d'enregistrement, l'autre de restitution. Au plan d'entrée de la première voie, on présente simultanément la référence (g) et la scène (h), séparées l'une de l'autre par une distance $2b$. La figure d'interférences enregistrée au plan du détecteur de cette voie est alors décrite par:

$$I(u,v) = |H(u,v)|^2 + |G(u,v)|^2 + H(u,v)G^*(u,v)e^{j\frac{4}{\lambda f_1}ub} + H^*(u,v)G(u,v)e^{-j\frac{4}{\lambda f_1}ub}, \quad (4)$$

fonction qui décrit l'intensité de la TF du plan d'entrée ($g + h$).

Cette figure d'interférences est ensuite transférée au modulateur spatial d'entrée de la seconde voie du système. On dira que le rôle de cette voie est « d'effectuer la TF inverse », car la figure restituée au plan du détecteur correspond mathématiquement à l'intensité de la TF inverse de l'équation 4. Il en résulte deux termes constants de chevauchement (dus aux deux premiers termes dans l'équation (4) et des termes de corrélation croisée, spatialement dissociés d'une distance proportionnelle à $2b$).

Le regain d'intérêt suscité par ce type de corrélateur provient de l'avènement des dispositifs à cristaux liquides (et des matériaux photoréfractifs), qui permettent un mode opératoire en temps réel. De plus, avec ce type de structure, il n'est pas nécessaire d'afficher de signal complexe (phase et amplitude) aux modulateurs, car les fonctions à afficher

sont réelles. Par contre, ce système n'est pas particulièrement performant pour le codage des filtres invariants aux distorsions ou des filtres mixtes. Dans ce cas, une séquence de filtres devra être utilisée, ce qui implique un ralentissement important du traitement. En général, ce type de corrélateur est beaucoup moins utilisé dans l'industrie que le corrélateur à plan fréquentiel [15].

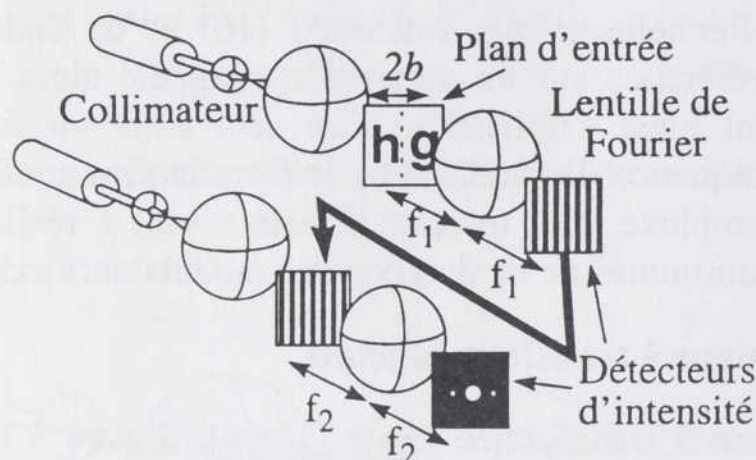


Figure 3. Corrélateur à transformée jointe

2.4 Performances du traitement optique analogique

Une question revient souvent: Pourquoi utiliser l'optique? La réponse est simple: le parallélisme massif et la vitesse élevée de calcul. Comme le système optique fonctionne en parallèle, la durée du traitement d'une TF ou d'une corrélation n'est pas déterminée par le volume des données. Le traitement peut théoriquement s'effectuer durant l'intervalle de temps très bref que prend la lumière pour passer du plan d'entrée au plan de sortie. Cependant, il serait plus juste de préciser que la vitesse du système est limitée par les dispositifs présentement disponibles. Il existe des MSL sur le marché fonctionnant à 10 000 images/s pour des images binaires de 256X256 pixels [16]. (La figure 4 illustre un assemblage possible d'un corrélateur optique d'un volume de 30 X 19 X 12 cm.) Ainsi, en théorie, 10 000 TF ou corrélations de cette dimension peuvent être calculées à la seconde. On dit bien en théorie car actuellement les détecteurs bidimensionnels actuels ne peuvent fonctionner aussi rapidement. Les modulateurs fonctionnant en tons de gris proviennent généralement de systèmes visuels de projection et fonctionnent à la cadence vidéo. Cependant, un dispositif de 512X512 pixels, avec une modulation en niveaux de gris de 7 bits, fonctionnant à près de 100 images/s a été conçu récemment [17]. La capacité absolue

équivalente de traitement des données des circuits numériques n'est tout simplement pas suffisante pour atteindre les performances des systèmes optiques correspondants. Les puces spécialisées de TF rapides peuvent, dans des conditions spécifiques, effectuer des TF rapides de 512×512 à des rythmes d'environ 4 images/s [18]. Quant aux corrélations, des circuits intégrés numériques de convolution peuvent calculer une corrélation de 512×512 à approximativement 120 images/s; mais ceci n'est valable que pour une matrice de filtrage de 3×3 [19]. Il faudrait plusieurs milliers de circuits intégrés électroniques (chacun fonctionnant à une puissance d'environ 1,3 W) pour obtenir une puissance de calcul équivalente à celle d'un corrélateur optique. Toutefois, il est évident que la TF et la corrélation constituent des opérations plutôt spécialisées et qu'un processeur optique est quelque peu restreint aux algorithmes qui peuvent être formulés en termes de ces opérations. Heureusement, en visionique, plusieurs opérations peuvent être réalisées en fonction des TF, convolutions et autres fonctions de filtrage similaires. Les processeurs optiques utilisés pour la mise en application de tels algorithmes surpassent nettement les processeurs numériques disponibles au niveau de la puissance et de la vitesse. Les limites dans la précision des calculs constituent le désavantage des traitements optiques analogiques. Elles résultent des limites de sensibilité et de la plage dynamique des matrices de détection ainsi que des limites de modulation des MSL. Encore une fois, en visionique, la plupart des primitives de vision artificielle requièrent relativement peu de bits par pixel de précision; les exigences liées à la vitesse de calcul étant beaucoup plus importantes.

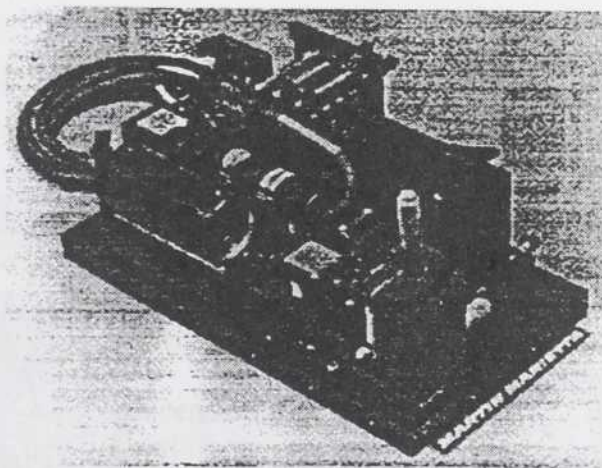


Figure 4. Corrélateur rapide compact [44].

3. Transformation de coordonnée et autres transformées

Selon le type de filtre utilisé dans le corrélateur, on peut réaliser différentes fonctions. De nombreuses recherches ont été entreprises portant notamment sur les fonctions morphologiques [20], les filtres d'ondelettes [21] et les filtres de Gabor [22]; la rotation d'images et le calcul de distributions de Wigner à l'aide de TF fractionnaires ont également été proposées [23]. Des systèmes semblables aux TF à lentille unique ont aussi été utilisés afin d'exécuter diverses transformations de coordonnées. Dans ce cas, un hologramme généré par ordinateur est habituellement placé après le plan d'entrée (la lentille de TF est souvent inutile puisque l'hologramme est codé) tel qu'indiqué à la figure 5. L'hologramme est codé de manière à rediriger la lumière incidente sur le plan d'entrée, vers le plan de sortie, selon un diagramme particulier correspondant à la transformation voulue. Par exemple, les transformations de Hough sont réalisées à l'aide d'un hologramme généré par ordinateur (composant diffractif), équivalent à la superposition de plusieurs lentilles cylindriques (composants réfractifs) [24]. Chaque « lentille cylindrique » fait ressortir un trait du plan de sortie bidimensionnel. Les transformations de coordonnées cartésiennes-polaires sont exécutées de façon similaire: chaque point de l'hologramme dirige la lumière vers un point précis du plan de sortie. De telles transformations de coordonnées présentent un intérêt particulier car les variations de rotation et d'échelle dans le plan d'entrée se traduisent par des décalages dans le plan de sortie. Ainsi, la combinaison d'un système cartésien-polaire et d'un système TF peut servir efficacement à l'extraction des caractéristiques d'objets à invariance d'échelle et de rotation (soit pour toute rotation ou à toute échelle de l'objet).

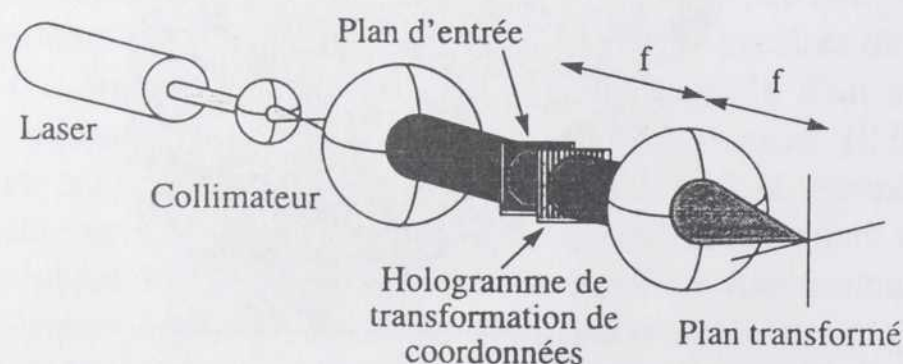


Figure 5. Système de transformation de coordonnées

En effet, lorsque le système cartésien-polaire est utilisé afin de traiter préalablement les données, le décalage dans l'espace cartésien-polaire qui en résulte n'a aucune conséquence sur le système TF qui suit (invariance au décalage). Un système de ce type opère en temps réel.

4. Exemples de traitement optique en visionique

Le traitement optique a engendré une abondance d'articles [25], d'éditions spéciales de revues [26] et de nombreuses conférences [voir par exemple 27, 28, 29] sur les processeurs optiques, leur fonctionnalité et leur applicabilité. Le domaine militaire a fourni la plupart du financement et des travaux de recherche liés à l'avancement du traitement optique. Toutefois, ces recherches se sont généralement limitées à la détection d'objets et à la reconnaissance de formes. Une grande partie des travaux produits dans ce domaine s'applique à la visionique. De bons exemples de systèmes hybrides et optiques fonctionnels sont présentés dans les documents [30, 31] des références bibliographiques. Dans ces deux systèmes, un processeur Datacube sert à détecter les objets et à extraire les contours. Le système conçu par l'INO (voir la figure 6 et [30]) utilise un système optique afin de calculer d'abord la TF des objets détectés et filtrés au préalable et ensuite de déterminer les vecteurs de caractéristiques utilisés par un réseau neuronal pour classifier ces objets; ce système permet la classification et la détection d'objets en temps réel. Il est important de noter que le processeur optique est utilisé dans ce système afin de permettre le traitement de gros débits de données au début du processus de reconnaissance.

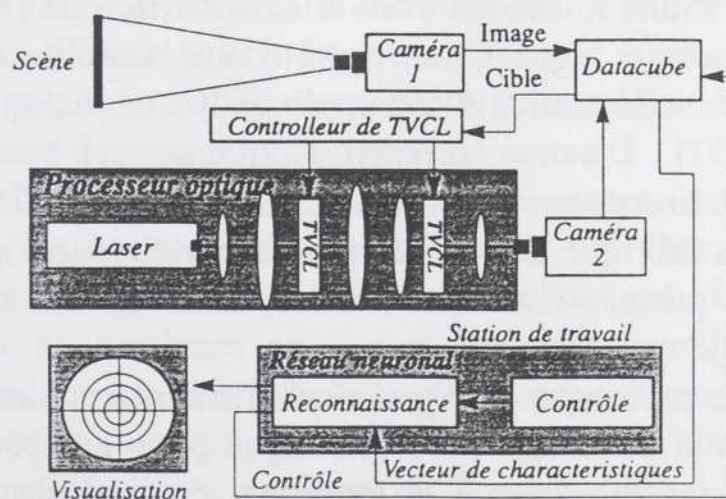


Figure 6. Système hybride de reconnaissance d'objets

Le traitement à haut niveau, responsable de la classification des cibles, vise un vecteur de caractéristiques à dimension réduite et s'effectue sur un processeur numérique universel. Un corrélateur optique a déjà été utilisé pour la classification d'objets par corrélation séquentielle (à vitesse élevée) avec plus de 1000 filtres de référence [31]; le système a réussi 18 identifications d'objets à la minute.

Le domaine aérospatial a également manifesté un grand intérêt dans l'utilisation de processeurs optiques pour la vision artificielle. À la référence [32], on présente un système hybride destiné à l'atterrissage automatisé d'un engin spatial. Cette application nécessite des transformations de coordonnées cartésiennes-polaires pour l'invariance de rotation et d'échelle, des corrélateurs optiques pour la recherche de sites d'atterrissage et des filtres de points de repère prédéterminés qui seraient calculés a priori. Aux références [33] et [34], on décrit des techniques liées au traitement optique et destinées à des applications dans les stations orbitales. Dans ces trois derniers exemples, les processeurs optiques ont été choisis en raison de leurs avantages par rapport aux dispositifs électroniques classiques au niveau de la dimension, du poids et de la consommation de puissance en rapport à des impératifs de traitement donnés. La firme française PSA donne un exemple de l'utilisation sur terre de processeurs optiques destinés à la navigation automatisée dans lequel un corrélateur à transformée jointe sert à la reconnaissance de signaux routiers [35].

Un corrélateur optique intégré a été mis au point par Teledyne Brown pour des applications militaires et de vision artificielle [36]. Ce système a été conçu de façon à assurer une rigidité, une stabilité et une facilité de production. Sa fonctionnalité et son applicabilité en visionique sont décrites à la référence [37]. D'autres entreprises ont consacré beaucoup de temps à la production de processeurs optiques. C'est notamment le cas de Grumman Aerospace qui a fabriqué des prototypes de corrélateurs optiques destinés à l'évaluation de pièces tridimensionnelles dans un système d'assemblage robotisé [38, 39].

De nombreux prototypes de systèmes de transformation de coordonnées ont également été présentés [40] et conçus pour l'inspection de produits [41]. Ces systèmes produisent en parallèle des tranches multiples d'une transformation de Hough à l'aide d'un hologramme créé par ordinateur. Dans les applications particulières étudiées, les caractéristiques précises du produit sont identifiées à des fins de rejet ou d'acceptation du produit.

5. Quel sera l'avenir des processeurs optiques?

On critique de plus en plus le fait qu'après plus de 20 ans d'existence, les corrélateurs optiques ne se soient pas vraiment avérés populaires. Il y a du vrai dans cet énoncé; la plupart des études effectuées dans ce domaine ont été longtemps théoriques (filtres invariants aux distorsions, filtres composites, etc.). La principale raison provient du manque de composants ayant une résolution et une vitesse suffisamment élevées pour être utilisés dans les systèmes optiques. On a compté (à quelques exceptions près) sur l'industrie de la visualisation pour la conception de MSL, prélevant les écrans de projection pour obtenir des composants. Ainsi, les composants étaient spécifiquement limités à des cadences d'images vidéo. Au cours des dernières années, plusieurs entreprises sont apparues et ont mis au point des MSL à haute résolution (dimension de pixels points < 20 mm) et à cadence élevée (> 500 Hz). Des dispositifs peu répandus comme les modulateurs à bactériorhodopsine construits à partir d'un matériel protéinique [42] à résolution submicronique et les modulateurs à puits quantiques multiples [43] à débits d'images de l'ordre du MHz, conçus en prototypes, sont présentement mis au point. Les difficultés surviendront au niveau de la lecture des fonctions de corrélation de format 512×512 produites par des systèmes à des fréquences supérieures à environ 1 kHz. Les DCC (détecteurs à coupleurs de charges) fonctionnent actuellement jusqu'à 500 images/s dans le cas de modèles spécialisés; cela deviendra un véritable problème très bientôt. Plusieurs chercheurs essaient de rendre les capteurs et les détecteurs «intelligents» afin de résoudre ces problèmes d'acquisition et de post-traitement.

Il y a également toute la partie de la mémorisation optique de données qui n'a pas été abordée dans cet article et qui, sans aucun doute, sera aussi utile dans l'avenir; les prochains systèmes devront nécessairement mémoriser et récupérer rapidement des «pages» de données.

Après avoir été pendant des années destinés presque exclusivement au domaine militaire, les corrélateurs prêts à installer font leur apparition dans le monde civil; la conception optique et optomécanique de même que la mise sous boîtier de systèmes solides et portatifs constituent des défis auxquels on commence seulement à s'attaquer. De nombreuses entreprises et organismes, dont l'INO, travaillent à la réalisation et à la commercialisation de corrélateurs destinés à des applications industrielles et utilitaires. Cela permettra sûrement d'intégrer les corrélateurs à d'autres domaines dont la visionique où le traitement de données constituent un véritable problème.

Au cours des prochaines années, les corrélateurs devraient vraisemblablement sortir des laboratoires et faire leur entrée dans le champ de la visionique industrielle.

6. Conclusion

Nous avons tenté, dans cet article, de présenter certains aspects du traitement optique appropriés à l'industrie de la visionique. À plus ou moins long terme, le traitement optique analogique aura probablement un impact dans ce domaine sur le plan de systèmes fiables (soit des systèmes ayant dépassé le stade de prototypes de laboratoire). Le traitement optique analogique, par son parallélisme, convient aux opérations bidimensionnelles requises en visionique; quant aux applications en temps réel nécessitant une concordance des modèles, les systèmes optiques peuvent présentement surpasser les systèmes numériques actuels au niveau du volume et de la consommation d'énergie relativement au traitement de gros débits de données. Les systèmes optiques ne sont cependant que des processeurs spécialisés; leur souplesse de programmation est limitée (généralement reliée aux opérations linéaires de Fourier). Mais il existe un large éventail d'opérations en visionique qui peuvent être réalisées à l'intérieur de cette contrainte.

L'avenir du traitement optique s'annonce donc brillant. Avec l'apparition de nouveaux composants et le développement de systèmes destinés au marché industriel, l'idée d'intégrer des processeurs optiques dans des systèmes industriels commence à se concrétiser et pourra se réaliser d'ici quelques années.

7. Références bibliographiques

- [1] D.A.B. Miller, «Quantum wells for optical information processing.» *Opt. Eng.*, vol. 26, no 5, (1987), p. 368-372.
- [2] A. Huang et al., «Signac fiber logic gates and their possible applications.», *Appl. Opt.*, vol. 33, no 26, (1994), p. 6254-6259.
- [3] P.K. Das et C.M. DeCusatis, *Acousto-optic signal processing: fundamentals and applications*. Artech House, Boston, (1991).
- [4] J.W. Goodman, *Introduction to Fourier optics*. McGraw-Hill, New York, (1968).
- [5] S. Maze et Ph. Refregier, «Opticall correlation: influence of the coding of the input plane.», *Appl. Opt.* vol. 33, no 29, (1994), p. 6788-6796.

- [6] A. Vander Lugt, «Signal detection by complex spatial filtering». IEEE Trans. Infor. Theory, vol. IT-10, (1964), p. 139.
- [7] J. Horner et P. Gianino, «Phase-only matched filtering.» Appl. Opt., vol. 23, (1984), p. 812-816.
- [8] D. Psaltis, E. Paek et S. Venkatesh, «Optical image correlation with a binary spatial light modulator.» Opt. Eng. vol. 23, (1984), p. 698-704.
- [9] B.V.K. Vijaya Kumar et Z. Bahri, «Efficient algorithm for designing a ternary valued filter yielding maximum signal to noise ration.» Appl. Opt., vol. 28, no 10, mai (1989), p. 1919-1925.
- [10] Y. Hsu, H. Arsenault et G. April, «Rotation-invariant digital pattern recognition using circular harmonic expansion.» Appl. Opt., vol. 21, (1982), p. 4012-4015.
- [11] B.V.K. Vijaya Kumar, «Tutorial survey of composite filter designs for optical correlators.», Appl. Opt., vol. 31, (1992), p. 4773-4801.
- [12] V. Laude et Ph. Refregier, «Multicriteria characterization of coding domains with optimal Fourier spatial light modulator filters.» Appl. Opt., vo. 33, (1994),, p. 4465-4471.
- [13] C. Weaver et J. Goodman, «A technique for optically convolving two funcitons.» Appl. Opt., vol. 5, (1966), p. 1248-1249.
- [14] B. Javidi, «Nonlinéal joint power spectrum based optical correlation.» Appl. Opt., vol. 28, (1989), p. 523-542.
- [15] H. Rajbenbach et al., «Compact photorefractive correlation for robotic applications.» Appl. Opt., vol. 31, n o 26, (1992), p. 5666-5674.
- [16] Boulder Nonlinear Systems, Boulder, CO, É-U.
- [17] Meadowlark Optics, Longmont, CO, É-U.
- [18] Raytheon TMC2310 par exemple.
- [19] Convolutionneur numérique Harris HSP48908 par exemple.
- [20] R. Schefer et D. Casasent, «Optical implementation of gray scale morphology.» In Nonlinear Image Processing III, vol. 1658, Proc. of Soc. Photo-Opt. Instrum. Eng., février (1992), p. 287-296.
- [21] D. Roberge et Y. Sheng, «Optical composite wavelet-matched filters.» Opt. Eng., vol. 33, (1994), p. 2290-2295.
- [22] D. Casasent, A. Ye, J.-S Smokelin et R. Schaefer, «Optical correlation filter fusion for object detection.», Opt. Eng., vol. 33, no 6, (1994), p. 1757-1766.

- [23] A. Lohmann, «Image rotation, Wigner rotation and the fractional Fourier transform.» JOSA-A, vol. 10, no 10, (1993), p. 2181-2186.
- [24] N. Carender, D. Casasent, F. Coetzee et D. Yu, «Hough transform computer generated holograms: New output format.» in Computer and Optically Generated Holographic Optics, vol. 1555, Proc. of Soc. Photo-opt. Instrum. Eng., (1991), p. 182-193.
- [25] D. Flannery, J. Loomis, M. Milkovich et P. Keller, «Application of binary phase-only correlation to machine vision.» Opt. Eng., vol. 27, no 4, (1988), p. 309-320.
- [26] «Special issue on optical computing systems.» Proc. IEEE, novembre, (1994).
- [27] D. Pape, éditeur. Advances in Optical Information Processing VI, vol. 2240, Proc. of Soc. Photo-Opt. Instrum. Eng., avril, (1994).
- [28] D. Casavent, éditeur. Transition of Optical Processing into Systems 1994, vol. 2236, Proc. of Soc. Photo-Opt. Instrum. Eng., avril, (1994).
- [29] D. Casasent et T.-H. Chao, éditeurs. Optical Pattern Recognition V, vol. 2237, Proc. of Soc. Photo-Opt. Instrum. Eng., avril (1994).
- [30] E. Eustache, D. Gingras et D. Poussart, «Hybrid system for automatic target recognition.» In Infrared Technology XX, vol. 2269, Proc. of Soc. Photo-Opt. Instrum. Eng., juillet (1994), p. 280-291.
- [31] D. Ben-Meir et S. Mills, «Optical correlator-based ATR system implementation and evaluation.» In Optical Pattern Recognition III, vol. 1701, Proc. of Soc. Photo-Opt. Instrum. Eng., (1992), p. 50-61.
- [32] R. Juday et T.-H. Chao, «Hybrid vision for automated spacecraft landing.» In Optical Pattern Recognition, H.-K. Liu, éditeur, vol. 1053, Proc. of Soc. Photo-Opt. Instrum. Eng., (1989), p. 131-139.
- [33] M. Scholl, «Optical procession for semiautonomous terminal navigation and docking.» Appl. Opt., vol. 31, no 26, (1993), p. 5049-5055.
- [34] D. Ennis et D. Jared, «Optical processing space station automation.» Opt. Eng., vol. 25, no 7, (1986), p. 808-820.
- [35] L. Guibert, G. Keryer et M. Attia, «Optical roadside recognition to improve active safety features.» In Optical Pattern Recognition IV, vol. 1959, Proc. of Soc. Photo-Opt. Instrum. Eng., (1993), p. 230-234.
- [36] M. Temmen et C. Hester, «Optical correlator for TOPS.» In Transition of Optical Processors into Systems 1993 D. Casasent, éditeur, vol. 1958, Proc. of Soc. Photo-Opt. Instrum. Eng., avril (1993), p. 19-24.

- [37] J. Brasher et al., «Optical processing architectures for machine vision functions.» In Machine Vision Architectures, Integrations, and Applications, vol. 1615, Proc. of Soc. Photo-Opt. Instrum. Eng., (1991), p. 39-45.
- [38] R. Brandstetter, N. Fonneland et D. Blitz, «Optical correlator simulation and experimental evaluation for three-dimensional parts.» In Optical Pattern Recognition III, vol. 1701, Proc. of Soc. Photo-Opt. Instrum. Eng., (1992), p. 66-76.
- [39] R. Brandstetter, N. Fonneland, R. Zanella et M. Yearwood, «Optical correlator vision system for a manufacturing robot assembly cell. » In Optics, Illumination, and Sensing for Machine Vision V, vol., 1385, Proc. of Soc. Photo-Opt. Instrum. Eng., (1990), p. 173-189.
- [40] L. Yulin, Z. Mingjun et Z. Li, «A hybrid pattern recognition system for robotic vision.» In Optics, Illumination, and Sensing for Machine Vision V, vol. 1385, Proc. of Soc. Photo-Opt. Instrum. Eng., (1990), p. 200-205.
- [41] J. Richards et D. Casasent, «Real-time optical Hough Transform for industrial inspection.» In Intelligent Robots and Computer Vision VIII: Algorithms and Techniques, vol. 1192, Proc. of Soc. Photo-Opt. Instrum. Eng., (1989) p. 2-21.
- [42] Q. Song et al., «Chemically enhanced bacteriorhodopsin thin-film spatial light modulation.» Opt. Lett., vol. 18, no 16, (1993), p. 1373-1375.
- [43] A. Parvoti et al., «High-speed joint-transform optical image correlator using GaAs/AlGaAs semi-insulating multiple quantum wells and diode lasers.» Opt. Lett., vol. 18, no 11, (1993), p. 906-908.
- [44] S. Serati, T.K. Ewing, R.A. Serati, K.M. Johnson et D. M. Simon, «Programmable 128X128 ferroelectric liquid crystal spatial light modulator compact correlator.» In Optical Pattern Recognition IV, vol. 1959, Proc. of Soc. Photo-Opt. Instrum. Eng., avril (1993), p. 55-68.

Approche connexionniste au problème de l'extraction de connaissances terminologiques à partir de textes.

Jean-Guy Meunier and Georges Nault
Université du Québec à Montréal
Laboratoire LANCI

Résumé

L'extraction de connaissances terminologiques sur des bases de données textuelles posent des problèmes. Comment maintenir un processus d'extraction opérationnel malgré la quantité massive de textes à traiter et qui sont en constante modifications. Nous proposons une approche connexionniste au problème. Cette approche permet d'appliquer les capacités classificatoires dynamiques des réseaux de neurones à un corpus textuel. Le domaine particulier d'application choisi pour cette expérience fut l'identification de réseaux lexicaux susceptibles d'interprétation sémantique par un terminologue. Le traitement connexionniste est intégré à des processus linéaires rapides de pré-traitement et post-traitement du texte. La chaîne de traitement ainsi produite allie dynamicité et rapidité pour fournir à l'utilisateur terminologue un appui inestimable dans sa tâche d'extraction des connaissances.

Mots-clés : connexionnisme, connaissances terminologiques, textes, extraction de connaissance, etc.

Approche connexionniste au problème de l'extraction de connaissances terminologiques à partir de textes

Jean-Guy Meunier and Georges Nault

*Laboratoire LANCI
Université du Québec à Montréal*

1. Le problème

De nos jours, un nombre croissant d'institutions accumulent très rapidement des quantités de documents qui ne sont souvent classés ou catégorisés que très sommairement. Rapidement, les tâches de dépistage, d'exploration et de récupération de l'information présente dans ces textes, c'est-à-dire des "connaissances", sont devenues extrêmement ardues, sinon impossibles. Un type particulier de ces connaissances que certains utilisateurs désirent extraire des textes est de nature terminologique. On veut connaître soit le vocabulaire propre du texte soit la signification de certains mots selon le contexte. Cependant, en raison de l'ampleur et de la dynamique des corpus, cette extraction terminologique devient de moins en moins possible dans des temps raisonnables ou faisables avec des ressources restreintes.

Du point de vue méthodologique, cette question de l'extraction des connaissances dans les textes rencontre cependant des difficultés épistémologiques sérieuses. En effet, la situation est relativement inverse au cadre traditionnel de l'IA qui réclame pour le traitement de l'information que le système en appelle à des dépôts de connaissances préformées et utilisées comme gabarit dans le dépistage et la reconnaissance. De plus les systèmes experts qui opèrent dans ce domaine doivent être dotés des mécanismes habituels (moteur d'inférence, maintien de cohérences, tests de plausibilité, etc.) leur permettant d'effectuer des déductions et des tests d'hypothèses avec un haut niveau de confiance et de réussites. Les connaissances comporteraient des représentations d'objets, de propriétés, de relations d'événements et de situations propre à l'objet à traiter en l'occurrence le contenu informationnel du texte. En possession de ce savoir, un système expert pourrait alors réussir à "comprendre" le texte et donc en extraire les connaissances. De nombreuses recherches ont d'ailleurs montré la nécessité de connaissances de multiples niveaux (syntaxiques, psycholinguistiques, lexicales, sémantiques, encyclopédiques, etc.). [19], [11], [17], [27]

Pour un ordinateur, le problème de l'extraction des connaissances dans le cas d'un corpus textuel ne se pose pas en termes des connaissances qu'il doit posséder pour "comprendre" le texte mais en termes de ce qu'il lui faut faire pour les extraire du texte. Il est, en effet, délicat de donner à l'ordinateur, de manière a priori, les connaissances que le texte avait pour fonction de transmettre sauf peut-être pour celles qui sont de nature générales, encyclopédiques ou techniques. Dans le cadre classique de l'intelligence artificielle, il est possible de donner de telles connaissances mais pour autant qu'elles aient été acquises via des enquêtes cognitives (analyse de protocole) chez les experts ou puisées dans le répertoire encyclopédiques du savoir partagé. Mais dans un horizon textuel, la connaissance se trouve dans le texte lui-même et doit en être extraite. Et les techniques qui ont donné des résultats intéressants sur de petits textes bien maîtrisés (scénario de restaurant, etc.) s'avèrent vite problématiques lorsqu'elles sont appliquées à des domaines dont on ignore en partie ou en totalité la teneur. Un texte ne contient normalement que des énoncés originaux qui n'ont pas encore été lus et dont le contenu tant lexical, sémantique et encyclopédique est inconnu du lecteur.

2. Extraction et classification.

La littérature technique relative au traitement de l'information textuelle a montré qu'il était devenu nécessaire d'explorer des outils d'extraction des connaissances dans des textes (data mining). Or, l'extraction de connaissances peut être vue de plusieurs manières. Dans notre perspective, elle n'est pas une "compréhension" du texte ni une paraphrase ni un rappel d'information mais une identification qui classe des segments de textes en regard d'une interprétation possible par un lecteur. Autrement dit, l'extraction des connaissances est définie comme résultant d'une opération de classification. Par ailleurs, cette classification ou extraction doit respecter les propriétés naturelles d'un texte. Il lui faut, en effet, respecter la dynamique et la plasticité d'un texte c'est-à-dire il faut classer une information qui se modifie constamment tant dans sa quantité que dans son contenu.

Pour les chercheurs dans le domaine de LATAO (lecture et analyse de textes assistés par ordinateur), cette problématique n'est pas nouvelle. Dans la recherche antérieure, plusieurs techniques et méthodes ont déjà été proposées pour tenter d'organiser le contenu d'un texte en des configurations interprétables. Outre les approches linguistiques et narratives, l'une des méthodes qui fut souvent explorée fut celle de l'analyse de données et de la reconnaissance de formes. Ces méthodes, moins fines certes que les approches linguistiques et conceptuelles

permettent un premier traitement générale. Elles sont en mesure , par exemple, d'identifier dans un corpus des classes ou des groupes de lexèmes qui entretiennent entre eux des associations dites de cooccurrence et donc de détecter leurs réseaux sémantiques. Parmi les modèles les plus couramment utilisés, on trouve habituellement, l'analyse des cooccurrences, l'analyse corrélationnelle, l'analyse en composante principale, l'analyse en groupe, l'analyse factorielle, l'analyse discriminante, etc.[14]. Malgré le succès obtenu par ces méthodes, on a dû constater que ces méthodes posaient deux problèmes importants.

Premièrement, les modèles classiques ne peuvent traiter que des corpus stables. Toute modification du corpus exige une reprise de l'analyse numérique. Ceci devient un problème majeur dans des situations où le corpus est en constante modification (par exemple les dépôts de l'autoroute électronique). Deuxièmement, les types de résultats qu'elles produisent ne sont pas sans problèmes théoriques. Elles posent des problèmes d'interprétation linguistique importants [5] Ces associations ne sont pas toujours facilement interprétables. Pourtant, malgré leurs limites, ces approches ont été reconnues des plus utiles pour l'extraction des connaissances et plus particulièrement les connaissances terminologiques. D'une part, ces stratégies classificatoires permettent une immense économie de temps dans le parcours exploratoire d'un corpus, et à ce titre, elles sont incontournables lorsqu'on est confronté à de vastes corpus textuels. D'autre part, elles servent d'indices pour détecter rapidement certains liens sémantiques et textuels. Associées à des stratégies linguistiques plus fines et intégrées dans des systèmes hybrides (i.e. , avec analyseurs linguistiques d'appoint) elles livrent une assistance précieuse pour des analyses globales. Elles permettent un premier déblaiement général du texte. Peuvent alors suivre des analyses plus fines. De telles approches hybrides émergentistes commencent à apparaître et à montrer leur efficacité. [39],[40] .

3. Les approches émergentistes.

La recherche récente permet de penser qu'on peut améliorer ces techniques de classification de l'information . En effet, de nouveaux modèles classifieurs dits émergentistes commencent à être explorés pour ce type de tâche. Une des thèses principales soutenue est que le traitement "intelligent" de l'information est avant tout associatif et surtout adaptatif. Parmi ces modèles dits "de computation émergente" on trouve ceux appelés "génétiques", (Holland 73) markoviens (R. Kindermann and L. Snell, 1980; Bouchaffra et Meunier, 1993) et surtout connexionnistes. Parmi ces derniers, on trouve une grande variété de modèles: entre autres,

les modèles matriciels linéaires et non linéaires (Anderson; Kohonen, 1989;), les modèles thermodynamiques (Hinton et Sejnowski, 1986), de même que les modèles basés tantôt sur la compétition, tantôt sur la rétropropagation mais surtout sur des règles complexes d'activation et d'apprentissage (Kohonen, 1989 ; Rumelhart et McClelland, 1986). Les principaux avantages de ces modèles tiennent au fait que leur structure parallèle leur permet de satisfaire un ensemble de contraintes qui peuvent être faibles et même, dans certains cas, contradictoires et de généraliser leur comportement à des situations nouvelles (le filtrage), de détecter des régularités et ce, même en présence de bruit . Outre les propriétés de généralisation et de robustesse, la possibilité pour ces modèles de répondre par un état stable à un ensemble d'inputs variables repose sur une capacité interne de classification de l'information.

Cependant, tous ces modèles opèrent sur des données bien contrôlées et qui toutes doivent être présentes au début et tout au long du traitement. De plus, ils exigent souvent divers paramètres d'ajustement qui relèvent souvent d'une description statistique du domaine. Il s'ensuit que les résultats de classification obtenus sont valides pour autant qu'ils portent sur les données où peu de modification sont possibles. Si, après la période d'apprentissage, pour quelque raison que ce soit, les systèmes sont confrontés à des données qui n'étaient pas prévues dans les données de départ, ils auront tendance à les classer dans les prototypes déjà construits. Donc à produire une surclassification.

Le domaine dans lequel nous opérons, à savoir le texte, présente précisément ce type de problème. Chaque nouvelle page peut possiblement contenir des informations que le système peut ne jamais avoir rencontré et donc il ne peut se permettre de les classer dans ses prototypes qu'il a antérieurement construits. Il faut donc, outre la dynamicité de l'apprentissage, un système qui soit aussi plastique.

4. Le cadre théorique et l'hypothèse

Il n'existe pas, à notre connaissance, un grand nombre de modèles connexionniste pouvant confronter simultanément les problèmes de dynamicité et de plasticité de l'information. La dynamicité touche la capacité du système à traiter de manière adaptative les informations auquel il est confronté alors que la plasticité touche la capacité du système à traiter de l'information pour laquelle il n'avait été paramétrisé. Un des modèles qui avait ces visées est ART ONE (Grossberg et Carpenter, 1988). En effet, dans sa définition originale le modèle ART ONE se veut

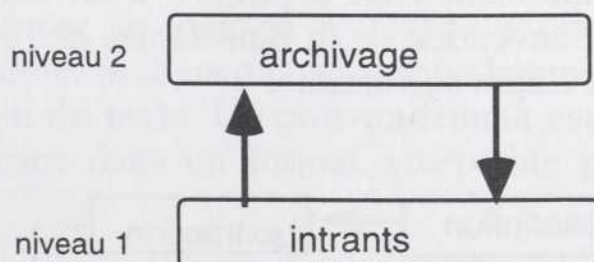
un système classifieur auto-organisationnel, sans supervision pouvant opérer sur des stimuli non contrôlés et bruités.

"An Art system is capable of self stabilizing the self organization of their recognition codes in response to an arbitrarily complex environment of input pattern in a way that parsimoniously reconciles the requirements of plasticity, stability and complexity." Carpenter et Grossberg 1988 : 254)

Ce modèle a évolué à travers les années. Il a livré de nombreuses variantes de la règle de transmission et de la règle d'apprentissage, Il a introduit des facteurs de multiplication dans l'activation et des facteurs de dégradation dans l'encodage de l'information. L'une de ses prétentions importantes est qu'il est en mesure de traiter de manière adaptative des stimuli qui sont changeants (plasticité) c'est-à-dire qui ne font pas partie d'un corpus contrôlé d'avance. L'objectif ultime de ce modèle est de créer une grande stabilisation dans la représentation des patterns de stimuli. Au début ART-ONE ne pouvait traiter que des informations de nature binaires. Plus tard, ART 2, le modèle accepte de informations dont les valeurs ne sont plus discrètes ou binaires. Enfin dans ART 3 le modèle a consolidé ses stratégies et offre un traitement plus fiable.

"An Art system can adaptively switch between stable and plastic modes. It is capable of plasticity in order to learn about significant new events, yet it can also remain stable in response to irrelevant events, in order to make this distinction, an ART system is sensitive to novelty. It is capable, without a teacher, of distinguishing between familiar and unfamiliar events, as well as between expected and unexpected events." Grossberg 1988: 254

L'idée centrale du modèle ART est celle d'un système de l'interaction entre deux niveaux qui entrent en résonance mutuelle.



Le système reçoit en un premier niveau N1 des stimuli qui sont envoyés mais aussi modifiés (selon une distribution et un poids particulier) au deuxième niveau N2 qui est un niveau d'archivage.

Arrive donc au deuxième niveau un pattern différent de ce qui était à l'intrant. Ce pattern est alors conservé et servira de gabarit ou de prototype contre lequel les intrants seront éventuellement comparés. A ce moment, ce prototype sera retourné (selon les mêmes règles de sommation) au stimuli intrant.

En fait, le pattern en N2 servira d'hypothèse vis-à-vis lequel un intrant sera comparé. i.e. ce qui entre est-il vraiment différent ou semblable au pattern archivé? S'il est différent, un autre pattern sera essayé jusqu'à ce qu'une correspondance soit acceptable (selon des paramètres) et s'il y a correspondance acceptable l'intrant sera alors classé avec le prototype. Mais s'il n'est pas acceptable, le nouveau pattern sera considéré comme une prototype en émergence et il servira éventuellement de nouveau gabarit aux autres intrants que le système rencontrera.

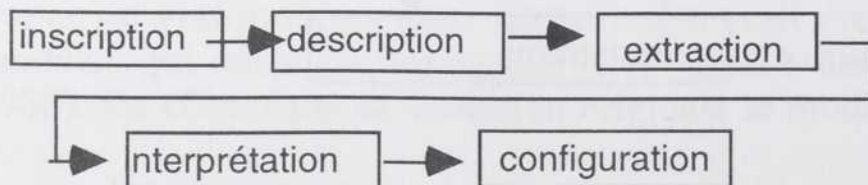
La résonance est définie comme la correspondance entre le pattern prototypal et le pattern intrant. Et au fur et à mesure que l'apprentissage se poursuit il y a consolidation de cette résonance. L'adaptativité survient par la modifications constante des interconnexions entre les niveaux.

Pour réaliser cette interaction le système doit être contrôlé par divers paramètres qui assurent la solidité du traitement.

5. Le traitement connexionniste du lexique d'un texte.

Le modèle Art a été utilisé dans une application d'extraction de connaissances terminologiques. La recherche servait à déterminer s'il était possible de structurer l'information lexicale dans un texte de manière à déterminer l'environnement sémantique d'un lexème particulier. Par exemple, quel est pour le mot CODE son ou ses usages sémantiques spécifiques? Autrement dit, les connaissances extraites consistent en des réseaux lexicaux susceptibles d'interprétation par un terminologue.

Notre chaîne de traitement était appliquée à un texte de quelques 900 pages. A savoir la Convention de la Baie James de l'Hydro Québec. Cette chaîne présentait les étapes suivantes:



Elle commence par une gestion du document, suit alors une description morphologique (lemmatisation) et matricielle du corpus. Vient ensuite une extraction classificatoire par réseaux de neurones ART ONE et, enfin, une interprétation des catégories par des tests inter-juges à des fins de validation. Une étape ultérieure à l'expérimentation s'ajoutera éventuellement pour opérer une mise (configuration) en forme hypertextuelle du plein texte à partir de la catégorisation effectuée par les réseaux de neurones. L'objectif global de la chaîne vise à développer un outil fonctionnel d'aide à l'analyse terminologique.

5.1 L'expérimentation

Dans la première étape de sa gestion, le texte est reçu et traité par un logiciel spécialisé dans l'analyse de contenu (SATO). A cette étape un filtrage sur le lexique du texte est appliqué. Via divers critères de discrimination SATO retire du texte certains mots accessoires (mots fonctionnels statistiquement non significatifs etc.) ou qui ne sont pas porteurs de sens d'un point de vue strictement sémantique et dont la présence pourrait nuire au processus de catégorisation soit parce qu'ils alourdiraient indûment la représentation matricielle soit parce que leur présence nuirait au processus interprétatif qui suit la tâche de catégorisation. Vient ensuite une description morphologique minimale de type lemmatisation. Cette opération consiste à remplacer chaque mot par son équivalent canonique. (vg. aimerions-. AIMER) Ce processus se justifie en ce que les déclinaisons propres à la grammaire ou à la syntaxe d'une langue n'affecte en rien le contenu sémantique profond des termes. De la même façon, remplacer un mot décliné (soit dans sa forme verbale, adverbiale, adjective, pronominale ou autres) par exemple, sa forme nominale n'a aucun impact significatif sur le contenu sémantique principal de ce dernier. Ces dimensions morphologique touchent surtout des aspects genres, aspects, temps etc.

Par après, une transformation est effectuée pour obtenir une représentation matricielle du texte. Cette transformation est opérée moitié par SATO et moitié par un logiciel développé explicitement à cette fin. SATO produit un fichier indiquant pour chaque lemme choisi sa fréquence dans chaque segment du texte. Un post-traitement est alors effectué pour construire une matrice dans un format acceptable pour les réseaux de neurones.

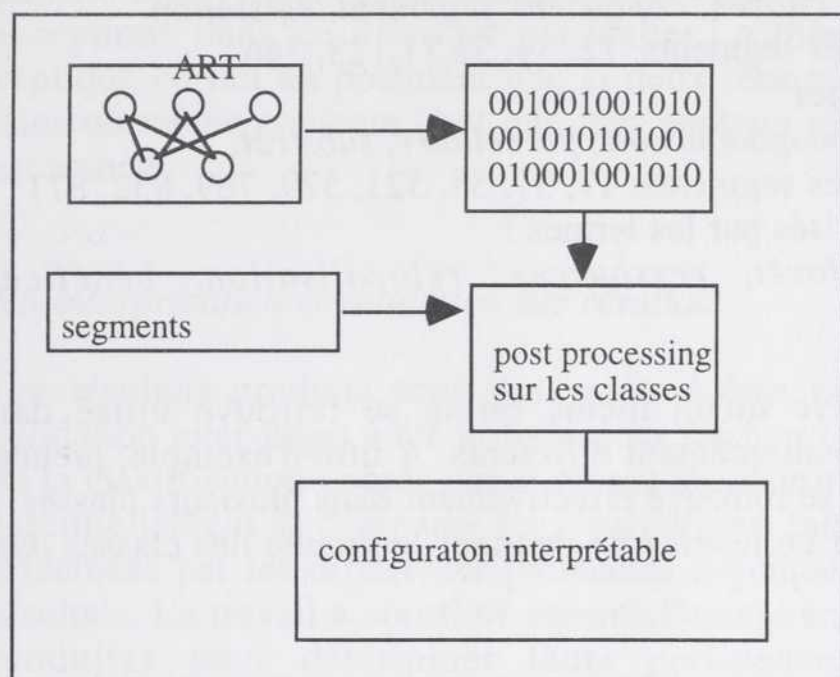
Selon le réseau utilisé (ART-ONE ou FUZZY-ART), la matrice générée pouvait être constituée exclusivement de données binaires (ART-

ONE) ou de données non-binaires (FUZZY-ART). Dans le cas du réseau ART-ONE, les données subissent alors une réduction, puisque la fréquence d'apparition des lemmes est alors remplacée par une simple indication de présence ou d'absence.

La matrice générée représente la distribution des lemmes dans le texte. Pour chaque segment du texte (qu'il s'agisse de pages, de paragraphes, de chapitres, ou autres), la matrice contient une ligne dont chaque élément indique, pour un mot donné, la fréquence de son apparition dans le segment). Le choix du type de réseau de neurones utilisé repose sur plusieurs facteurs, principalement la nécessité d'opérer une classification dynamique, non-supervisée sur des intrants de longueurs fixes. Les réseaux de type ART se prêtent particulièrement à ces contraintes. Le second type de réseau semble donner des résultats supérieurs en ce que le processus de catégorisation produit moins de classes contenant une seul élément caractéristique (un seul lemme), ce qui du point de vue du terminologue n'est d'aucune utilité.

La matrice ainsi produite sert alors d'intrant au réseau de neurones 1 : les réseaux de neurones ART ONE utilisés pour notre expérimentation ont été développé sur une plate-forme de programmation matricielle disponible sur MATLAB. Cette implantation informatique particulière avait été réalisée dans le cadre d'expérimentations sur le traitement massif de signaux. Des modifications mineures d'un certain nombre de caractéristiques ont permis à ce programme de générer des résultats interprétables du point de vue linguistique.

Le réseau neuronal, commence par générer une matrice de résultats qui représentent la classification trouvée.



Chaque ligne (ou vecteur) de cette matrice est constituée d'éléments binaires ordonnés. Cette ligne indique pour chaque terme du lexique original si il fait, oui ou non, partie du prototype de la classe. Ainsi est créé un "prototype" pour chacune des classes identifiées.

On dira alors que la classe no. X est "caractérisée" par la présence d'un certain nombre de termes. Autrement dit, chaque classe identifie quels sont les termes qui se retrouvent dans les segments de textes qui présentent, selon le réseau de neurone une certaine similarité du point de vue lexical. Les résultats du réseau de neurones se présentent donc (avant interprétation) sous la forme d'une séquence de *classes de segments*. De cette classe de segments on extrait alors les termes qui les caractérisent c'est-à-dire qui constituent le lexique de ces segments.

5.2 Exemple

Dans notre analyse, nous obtenions dans

La classe # 17 contient les segments 14,29,34,95,181

Tous ces segments sont caractérisées par les termes :

barrage, eau, réserve, digue, indiens, entente, rivière

Il existera évidemment plusieurs classes de ce type.

Par exemple:

la classe # 56 contient les segments 23, 34,55,76, 11 120

caractérisés par les termes :

attribution, citoyen, compétent, jugement, législation

la classe # 57 contient les segments: 12, 39, 52, 71, 123, 346

caractérisés par les termes :

comté, Chibougaouccôte, parcellaire, substrat.

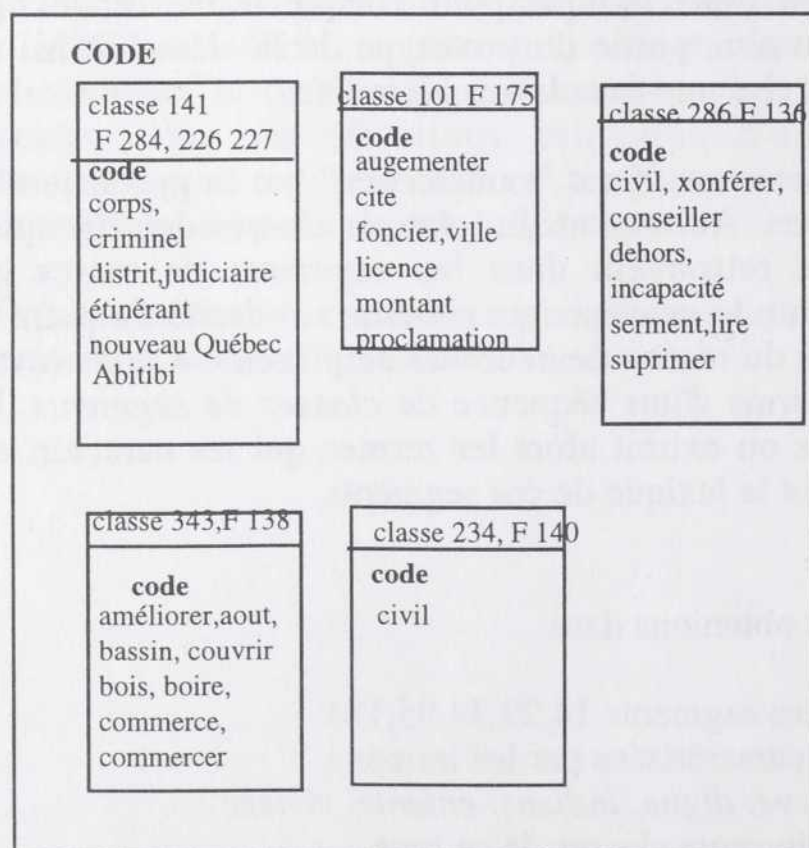
la classe # 58 contient les segments: 17, 31, 58, 521, 579, 769, 832, 871

qui étaient tous caractérisés par les termes :

mineral, forêt, ressources, relocalisation, bénéfices, expropriation

Dans un texte, il arrive qu'un même terme se retrouve utilisé dans plusieurs contextes sémantiquement différents. A titre d'exemple, prenons le mot CODE. Celui-ci se retrouve effectivement dans plusieurs classes. Il était alors possible pour ce lexème de dessiner le graphe des classes dans lesquels il apparaît.

TABLEAU POUR LE MOT CODE.



On voit alors surgir la différence des réseaux sémantiques de ce terme, qui tantôt est utilisé dans le sens de code civil, dans le sens de code de

comportement, code criminel, etc. Pour un terminologue, ce graphe est parlant. Il sert d'indice aux réseaux sémantique de ce termes et donc de ses acceptions dans ce discours particulier. La thèse associative classique explique ce fait en postulant que si deux termes se retrouvent ensemble dans un même contexte c'est que leur contenu sémantique ou conceptuel est associé.

5.3 Interprétation et validation des résultats

Les résultats produits sont alors prêts à être validés. Une méthode de validation inter-juges a été utilisée pour évaluer la qualité et la pertinence de la classification opérée par le réseau de neurones. La méthode consiste essentiellement à comparer les évaluations faites et les appréciations effectuées par les différentes personnes impliquées dans l'évaluation des résultats. Le travail a constitué essentiellement en une analyse des classes produites pour déterminer leurs pertinences d'un point de vue terminologique.

De façon générale, l'analyse a montré que plus de 80 % des classes constituées seraient pertinentes et utilisables dans le processus d'établissement de bases terminologiques.

La forme actuelle des résultats ne se prêtent pas à l'interprétation directe par le terminologue. Avant de pouvoir servir concrètement, les données devront passer par une interface spéciale qui devrait dans un premier temps permettre l'approche catégorielle des lemmes et dans un second temps permettre la mise en forme hypertextuelle du plein texte à partir de la catégorisation effectuée par le réseau de neurones.

L'outil ainsi développé devrait permettre au terminologue d'accéder le plein texte suivant une stratégie de recherche guidée par le travail de catégorisation ou, advenant le cas, de construire les sens des lemmes par une recherche incrémentielle sur ces derniers.

6. Conclusion

L'objectif de notre recherche était de permettre l'extraction de connaissances terminologiques à partir du texte plein. Cette extraction devait, de plus, pouvoir se faire sur un corpus en évolution constante (plasticité) et les catégorisations effectuées se devaient de rester pertinentes et utilisables. Le processus de catégorisation quel qu'il soit se devait

d'opérer sans supervision aucune (adaptivité) et sans faire appel à des connaissances pré-formées ou pré-digérées, celles-ci n'étant simplement pas disponible dans le cas qui nous concerne.

Une méthode connexionniste comme solution au problème de l'extraction terminologique sur des textes entiers a été expérimentée avec des résultats très encourageants. L'approche choisie montre un intérêt certain et des avantages indéniables. Par exemple, le gain de temps estimé par rapport au travail manuel requis pour effectuer un travail terminologique équivalent est considérable. De plus, la précision et la richesse des suggestions faites par le système ne sont en aucune mesure comparable avec ce qu'on obtient avec les méthodes actuelles.

L'approche choisie s'inscrit parfaitement dans l'optique d'une solution opérationnelle aux problèmes réels qui minent l'industrie de la langue et toute industrie qui implique la manipulation et la classification de masses importantes de documents.

7. Recherches et développements futurs

Plusieurs variantes de l'expérimentation sont possibles et envisagées. On peut caractériser les classes autrement que par l'intersection entre les segments qui la constitue. Pourrait-on par exemple tirer d'autres informations pertinentes de l'union des classes de segments classés? On pourrait sûrement obtenir des résultats encore plus probants en tenant compte de la non-indépendance entre les variables (ou unifs) pris en considération. Un pré-processing sémantique (non-approfondie pour conserver l'avantage de temps) pourrait amener des améliorations considérables. Plusieurs variantes des filtres de pré-traitement utilisés sont à l'étude.

Il va de soi qu'on pourrait aussi filtrer à la sortie les classes ne représentant pas une richesse d'information suffisante (selon un critère donné, ex. : pas assez d'unifs dans cette classe).

Plusieurs problèmes restent à résoudre (grandeurs fixes des intrants, dégradation du temps d'apprentissage avec le nombre des intrants, interprétation, etc.).

Un module d'interprétation des résultats (avec interface graphique) est en chantier.

8. Références

- [1] Anderson, J. A., & Hinton, G. E. (1981). *Parallel Models of Associative Memory*. Hillsdale: Erlbaum
- [2] Barto, A. G. (1990). "Neural Network Learning for Control: An overview.". In T. Miller, R. S. Sutton, & W. P.J. (Ed.), *Neural Network for control*. Cambridge: MIT Press.
- [3] Burr, D. J. (1987). "Experiments with a connectionist text reader". *IEEE First International Conference on Neural Networks*, San Diego, 717-24
- [4] Carpenter, G. , & Grossberg, G. (1991). "An Adaptive resonance Algorithm for Rapid Category Learning and Recognition". *Neural Networks* 4, 493-504.
- [5] Church, K. W., & Hanks, P. (1990). "Word association norms, mutual information and lexicography". *Computational Linguistics* 16 (1), 22-29.
- [6] Cheeseman, P. , Self, M. , Kelly, J. , Stutz, J., Taylor, W. , & Freeman, D. (1988). "Bayesian Classification". *Proceedings of AAAI 88*, Minneapolis, 607 -611
- [7] Delany, & P. Landow (Ed.), *The Digital Word: Text Based Computing in the Humanities*. Cambridge, Mass: MIT Press.
- [8] Delisle, S. (1994). *Text Processing without a priori domain knowledge: semi automatic linguistic analysis for incremental knowledge acquisition*. PH Thesis, Ottawa University. :
- [9] Frey, S. , Reyle, U. , & Rohrer, C. (1983). "Automatic Construction of a Knowledge Base by Analysing Texts in Natural Language". *Proc of IJCAI*, 83 727-729,
- [10] Grefenstette, G. (1992). "Sextant: Exploring Unexplored Contexts for Semantic Exeraction from Syntactic Analysis". *Proc of the 30th Annual Meeting fo the ACL* 324- 326,
- [11] Grefenstette, G. (1992). "Use of syntactic Context to Produce Term Association Lists for Text Retrieval". *Proc of SIGIR 92 ACM*, Copenhagen, June 21-24,
- [12] Grossberg, S. , & Carpenter, S. (1987). "Self Organization of Stable Category Recognition Codes for Analog Input Patterns". *Applied Optics* 26, 4919- 4930.
- [13] Grossberg, S. (. (1988). *Neural Network and Natural Intelligence*. Cambridge: MIT Press.
- [14] Hinton, G. E. (1987). "Learning and Relearning in Boltzman Machines". In D. E. & M. Rumelhart J. (Ed.), *Parallel Distributed Processing :Explorations in the Micro Sctucture of Cognition*. vol. 1, (pp. 282-318). Cambridge: MIT Press.
- [15] Holland, J. (1975). *Adaptation in Natural and Artificial Systems*. Ann Arbor, Mi.: University of Michigan Press.
- [16] Jacobs, P. , & Zernik, U. (1988). "Acquiring Lexical Knowledge from Text A case Study". *Proceedings of AAAI*, 88 (St Paul. Min.),
- [17] Jansen, S. , Olesen, J. , Prebensen, H. , & Tharne, T. (1992). *Computanional approaches to text Understanding*. in Copenhagen: Museum Tuscalanum Press,
- [18] Kahonen, T. (1982). "Clustering, taxonomy and topological Maps of Patterns". *IEEE Sixth International Conference on Pattern Recognition*, 114-122
- [19] Lebart, L. , & Salem, A. (1988). *Analyse statistique des données textuelles*. Paris: Dunod.
- [20] Lin, X. , Soergel, D. , & Marchionini, G. (1991). "A Self Organizing Semantic Map for Information Retrieval". *SIGIR 91*, Chicago, Illinois,
- [21] Meunier J. G. , F. Daoust. , Rolland, S. . (1976,). "Sato: A system for Automatic Content Analysis of Text,". *Computer and the Humanities* vol. XX. , p 45. 64
- [22] Moulin B, & Rousseau, D. (1990). "Un outil pour l'acquisition des connaissances a partir de textes prescriptifs". *ICO*, Québec 3 (2), 108-120.

- [23] Recoczei, S. , & E. P. O, P. (1988). "Creating the Domain of Discourse: Ontology and Inventory". In J. & B. G. Boose (Ed.), *Knowledge Acquisition Tools for Experts and Novices*. Academic Press:
- [24] Regoczei, S. , & Hirst, G. (1989). *On extracting knowledge from Text. Modelling the Architecture of Language Users*. (TR CSRI 225). Computer Systems Research Institute University of Toronto.
- [25] Rumelhart, D. E. , & Mc Clelland, J. (1987). *Parallel Distributed Processing :Explorations in the Micro Structure of Cognition*, 2 vol. Cambridge: MIT Press.
- [26] Salton, G. (1988). "On the Use of Spreading Activation". *Communications of the ACM* vol 31 (2),
- [27] Salton, G. , Allan, J. , & Buckley, C. (1994). "Automatic Structuring and Retrieval of Large Text File". *Communications of the ACM* 37 (2), 97-107.
- [28] Snell, K. R. e. L. (1980). *Markov Random Fields and their applications*, *Contemporary mathematics*, AMS, Vol. 1.
- [29] Sutton, & W. P.J. (Ed.), *Neural Network for control*. Cambridge: MIT Press.
- [30] Shaw, M. L. G. , & B. R. , G. (1988). "Knowledge Initiation and Transfer Tools for Expert ad Novices". In J. B. & B. Gaines (Ed.), *Knowledge Acquisition Tools for Expert Systems*. Academic Press.
- [31] Tapiero, I. (1993). *Traitement cognitif du texte narratif et expositif et connexionnisme: expérimentations et simulations*. PhD thesis in Université de Paris VIII,
- [32] Thrane, T. (1992). "Dynamic Text Comprehension". In J. O. S. Jansen H Prebensen,T. Thrane (Ed.),*Discourse analysis* Copenhaguen: Museum Tuscalanum Press.
- [33] Veronis, J. , Ide, N. M. , & Harie, S. (1990). "Utilisation de grands réseaux de neurones comme modèles de représentations sémantiques". *Neuronimes*,
- [34] Virbel, J. (1993). "Reading and Managing Texts on the Bibliothèque de France Station". In P. Delany, & P. Landow (Ed.), *The Digital Word: Text Based Computing in the Humanities*. Cambridge, Mass: MIT Press.
- [35] Williams, M. (1990). " Connectionist Models and Information Retrieval". 25, 209-259.
- [36] Young, T. , & Calvert, T. (1987). *Classification, Estimation, and Pattern Recognition*. Amsterdam: Elsvier.
- [37] Zarri, G. P. (1990). "Représentation des connaissances pour effectuer des traitements inférentiels complexes sur des documents en langage naturel. ". In Office de la langue française (Ed). *Les industries de la langue*. Perspectives 1990. Gouvernement du Québec.
- [39] Balpe,J.P. Lelu,A. Papy.F. Saleh, I *Techniques avancées pour l'hypertexte*. Hermes, 1996.
- [40] Lelu, A, *Modèles neuronaux pour l'anlyse de données documentaires et textuelles*.Doctrorat, Paris VI, 1993.

note

Plusieurs chercheurs explorent actuellement l'utilisation du modèle neuronal dans le domaine du texte soit a des fins de rappel d'information (Salton et al. 1994;Veronis J. 1990, Williams, M. (1990), soit a des fins d'analyse sémantique (Lin et al. 1991) ou l'analyse discursive,(Tapiero 1993, Kosko 1991Burr,1987. Balpe et al. 1996, Lelu 1993). Notre propre recherche porte sur l'acquisition des connaissances terminologiques sur des corpus en constante mutation.

La reconnaissance de la parole en français

Pierre Dumouchel¹, Rivarol Vergin¹, Douglas O'Shaughnessy² et Jean Rouat³

¹Ecole de technologie supérieure, département de génie électrique, Montréal (Québec).

²INRS-Télécommunications, Ile-des-Soeurs (Québec).

³UQAC, département de sciences appliquées, Chicoutimi (Québec).

e-mail: dumouchel@ele.etsmtl.ca

Résumé

La reconnaissance automatique de la parole retient aujourd'hui l'attention de beaucoup de centres de recherche. Cet intérêt pour ce champ particulier de recherche s'explique d'abord, pensons-nous, par les applications visées dans les domaines les plus divers. Cet article donne une description sommaire du système de reconnaissance de l'INRS-Télécommunications. Des solutions à certains problèmes spécifiques au français, comme le phénomène de liaison qui est moins problématique en anglais, sont présentées. La performance du système, en terme de taux de reconnaissance, est évaluée à partir de romans audios lus par un locuteur.

1. Introduction

Dans la théorie du langage formel, un langage est défini par un ensemble de séquences finies de la forme: $x_1 \dots x_n$, où chaque symbole, x_n , provient d'un ensemble fini V appelé le vocabulaire du langage.

Une classe importante de langages, connue sous le nom de *langages réguliers*, sont ceux caractérisés par un automate fini [1]. Un automate fini est défini par un graphe comprenant: (a) un ensemble fini d'états S ; (b) deux sous-ensembles disjoints de S , les éléments de l'un sont appelés les états initiaux, et les éléments de l'autre sont appelés les états finaux; (c) un ensemble fini de transitions, chacune de la forme (s, x, s') où s et s' sont des états, s n'étant pas un état final et x un symbole appartenant à l'ensemble V qui forme le vocabulaire.

Nous disons qu'une transition de la forme (s, x, s') a s comme état de départ et s' comme état d'arrivée. Un chemin est une séquence de transitions $\tau_1 \dots \tau_N$, où l'état d'arrivée de τ_{n-1} est le même que l'état de départ de τ_n . Un chemin est dit complet s'il commence à un état initial et se termine à un état final [2]. Chaque chemin complet définit une séquence de symboles $\mathbf{x}$; le

langage associé à l'automate est alors défini par l'ensemble des séquences de symboles obtenues de cette façon.

La notion d'automate fini a été étendue de manière à ce qu'on puisse associer à une séquence de symboles donnée $\mathbf{x}$ une probabilité $P(\mathbf{x})$. Le modèle stochastique qui en résulte, connu sous le nom de chaîne de Markov cachée (HMM), est utilisé dans la plupart des systèmes de reconnaissance de la parole dont celui de l'INRS-Télécommunications [8], parce qu'il permet de tenir compte des variations temporelles et spectrales observées sur le signal à l'entrée du système.

Dans les sections qui suivent, nous allons décrire chacun des principaux modules qui composent le système de reconnaissance automatique de la parole, en faisant ressortir à chaque fois les particularités propres au fait français.

2. Système de reconnaissance

2.1 Vue d'ensemble

La majorité des systèmes de reconnaissance, y compris celui de l'INRS-Télécommunications, représente la parole par un modèle statistique. Ce modèle est caractérisé par une probabilité jointe d'une séquence de mots et du signal de parole. La séquence de mots est issue d'un dictionnaire qui comprend aussi le mot "silence", et le signal de parole est représenté par un ensemble de vecteurs de paramètres. Ces vecteurs de paramètres, calculés tous les 10 ms, constituent les éléments fournis à l'entrée du système de reconnaissance. Notre modèle comporte trois parties:

1. Un modèle de langage dont le rôle est d'évaluer une distribution de probabilité *a priori* des séquences de mots du dictionnaire.
2. Une composante phonologique qui permet d'avoir une ou plusieurs transcriptions phonétiques de chaque mot du dictionnaire.
3. Un modèle acoustico-phonétique qui permet de faire le lien entre le signal de parole à l'entrée et les transcriptions phonétiques des mots du dictionnaire.

Le modèle de langage et le modèle acoustico-phonétique permettent de

calculer la probabilité conjointe d'une séquence donnée de mots et du signal de parole. La séquence de mots choisie par le reconnaiseur est celle pour laquelle cette probabilité conjointe est maximale.

2.2 Modèle de langage

Un modèle de langage est un modèle discret de chaîne de Markov cachée dont les symboles sont les mots du dictionnaire [4]. Il comporte une collection d'état $\{\sigma\}$.

La probabilité jointe d'une séquence de mots, $w_1 \dots w_N$, et d'une séquence d'états, $\sigma_0 \dots \sigma_N$, est donnée par:

$$P(w_1 \dots w_N) = \prod_{n=1}^N P(w_n, \sigma_n | \sigma_{n-1}) \quad (1)$$

La probabilité d'une séquence donnée de mots w est définie par $P(w, \Sigma)$ où Σ représente l'ensemble des séquences d'états possibles pouvant produire la séquence de mots w .

2.2.1 Modèle de langage déterministe

Un modèle de langage déterministe est un modèle où une seule séquence possible d'états peut être associée à une séquence de mots. Donc la chaîne de Markov n'est plus cachée. De manière générale pour construire un modèle déterministe, on utilise une relation d'équivalence sur les séquences de mots ayant la propriété suivante:

$$w \equiv w' \Rightarrow ww \equiv w'w \quad (2)$$

où w est un mot et w et w' sont des séquences de mots. Le modèle comporte un état pour chaque classe de séquences de mots équivalents.

Un exemple simple de modèle de langage déterministe est celui des N-grammes [3]. Ce modèle a un état pour chaque (N-1)-grammes possibles. La distribution de probabilité $P(w|\sigma)$ est définie comme suit, si l'état courant est σ avec:

$$\sigma = w_1 \dots w_{N-1} \quad (3)$$

alors l'état suivant, σ' , est totalement défini par le mot courant w parce que:

$$\sigma' = w_2 \dots w_{N-1} w \quad (4)$$

Si on suppose qu'il existe une collection de statistiques sur les n -grammes, avec $n = 1 \dots N$, alors:

$$P(w_n | w_1 \dots w_{n-1}) = \frac{C(w_1 \dots w_n)}{C(w_1 \dots w_{n-1})} \quad (5)$$

Dans cette équation, $C(w_1 \dots w_n)$ représente le nombre de fois que la séquence $w_1 \dots w_n$ a été observée dans le corpus d'entraînement, c'est-à-dire sur l'ensemble des phrases qui servent à entraîner le système, et $C(w_1 \dots w_{n-1})$ est donnée par:

$$C(w_1 \dots w_{n-1}) = \sum_{w \in V} C(w_1 \dots w_{n-1} w) \quad (6)$$

où V est l'ensemble des mots du dictionnaire. La règle [5] d'évaluation de $P(w|\sigma)$ est la suivante: si $P(w|w_{n-1}) > 0$ alors $P(w|\sigma) = P(w_{n-1})$. Sinon, si $P(w|w_{n-2}) > 0$ alors $P(w|\sigma) = \lambda P(w|w_{n-2})$. λ représente la pénalité de passer d'un modèle de n -gramme à un modèle de $(n-1)$ -gramme.

2.3 Modèle acoustico-phonétique

Soit $Y_1 \dots Y_T$ un ensemble de vecteurs de paramètres caractérisant un signal de parole donné et $F_1 \dots F_N$ un ensemble de phonèmes. Supposons que l'ensemble des vecteurs Y puissent être décomposés en N , chacun correspondant à un phonème F_n ($n = 1 \dots N$) et que ces segments soient statistiquement indépendants. Sous cette hypothèse, l'analyse acoustico-phonétique peut se faire en ayant seulement des modèles caractérisant des segments de phonèmes. Posons $N = 2$ et définissons par $[t, t']$ la durée du signal de parole, alors la probabilité que la séquence de phonèmes $F: F_1, F_2$ ait été émise est donnée par:

$$P([t, t'] | F) = \sum_{i=t}^{t'-1} P([t, i] | F_1) P([i+1, t'] | F_2) \quad (7)$$

L'index i varie de t à $t'-1$ parce que les segments des phonèmes F_1 et F_2 doivent être au moins de longueur 1. En itérant cette procédure, on peut calculer la probabilité de n'importe quel signal d'entrée pour n'importe quelle séquence de transcription phonétique. Cette mesure est connue sous le nom de "stochastic segment modelling". Cependant l'approche qui permet d'obtenir de meilleurs résultats jusqu'à aujourd'hui, lorsqu'on utilise des modèles de segments de phonèmes, est celle basée sur le formalisme des chaînes de Markov cachées (HMM). Cette approche a deux avantages majeurs: elle est facile à implanter et elle permet de tenir compte des variations spectrales et temporelles des segments de phonèmes.

2.3.1 Les modèles HMMs

Un modèle HMM ou chaîne de Markov cachée peut être vue comme un automate probabiliste si l'on considère que les transitions d'un état à un autre ne se font pas à des instants précis, mais sont plutôt régies par une loi probabiliste. Dans le cadre de la reconnaissance automatique de la parole [7], le modèle HMM sera spécifié par:

1. Un ensemble de distributions gaussiennes qu'on dénote par C . C peut être considéré comme un "livre de codes".
2. Un ensemble d'états; chaque état s a un mot de code c_s qui lui est associé.
3. Un ensemble de transitions d'un état à l'autre et une distribution de probabilité rattachée à ces transitions.

Recherche avant

La représentation d'un ensemble d'observations $Y_1, Y_2, \dots$ par des modèles HMMs se fait de la façon suivante. Un état de départ s_1 est d'abord choisi auquel est associé un mot de code c_1 appartenant à C . La probabilité d'observer Y_1 étant donné l'état s_1 est alors définie en considérant la distribution de probabilité rattachée à c_1 . La généralisation de la procédure est immédiate, considérant que les observations $Y_1 \dots Y_t$ ont été générées et que

l'état courant est l'état s_t . Une transition τ est alors choisie parmi l'ensemble des transitions dont l'origine est s_t . L'état qui sera occupé au temps $t+1$ est tout simplement l'état d'arrivée de la transition τ . Si cet état est un état final, alors la séquence d'observation est terminée; sinon une nouvelle observation Y_{t+1} sera générée en considérant la distribution de probabilité c_{t+1} associée à l'état s_{t+1} .

Ainsi, une séquence d'observation $Y_1 \dots Y_T$ est générée en considérant une séquence de mot de codes $c_1 \dots c_{T+1}$ qui a été produite par une séquence de transitions d'états $s_1 \rightarrow s_{T+1}$. Pour $t = 2 \dots T$, définissons par $\alpha_t(s) \big|_{s=s_t}$ la mesure de vraisemblance des observations $Y_1 \dots Y_t$. $\alpha_t(s)$ est donnée par la formule récurrente suivante:

$$\alpha_t(s) = P(Y_t | c) \max_{\tau} \alpha_{t-1}(s') P(\tau | s') \quad (8)$$

Dans cette équation, c est le mot de code associé à l'état s , τ appartient à l'ensemble des transitions qui aboutissent à l'état s , s' est l'état d'origine de τ et $P(\tau | s')$ est la probabilité de passer à l'état s au temps t étant donné qu'on était dans l'état s' au temps $t-1$.

Lorsque pour évaluer la probabilité d'observer une séquence donnée $Y_1 \dots Y_T$, on commence par le dernier élément Y_T la procédure s'appelle alors *recherche arrière* [6]; $\alpha_t(s)$ est dans ce cas remplacé par $\beta_t(s)$. La combinaison de ces deux procédures (recherche avant et recherche arrière) permet en général d'évaluer d'une façon simple les paramètres du modèle HMM, c'est-à-dire les probabilités de transition entre les états ainsi que les probabilités d'émission d'un symbole donné Y_T étant donné l'état considéré s_t .

2.3.2 Principe de reconnaissance

Une hypothèse de reconnaissance est un doublet comprenant une séquence de mots w et une transcription phonétique F de w . Le problème est alors de

trouver la séquence de mots w ainsi que la transcription phonétique F pour lesquelles la quantité

$$P(w)P(Y_1 \dots Y_T | F) \quad (9)$$

est maximale; $Y_1 \dots Y_T$ sont les vecteurs d'observations qui caractérisent le signal de parole à l'entrée, $P(w)$ est la probabilité *a priori* de la séquence de mots w évaluée à partir du modèle de langage et $P(Y_1 \dots Y_T | F)$ est la probabilité d'observer la séquence $Y_1 \dots Y_T$ étant donné la séquence de phonèmes F ; cette dernière est évaluée à partir du modèle acoustico-phonétique. Pour résoudre ce problème, certains systèmes de reconnaissance, y compris celui de l'INRS-Télécommunications [9], utilisent une procédure à deux passes. Au cours de la première passe des modèles de langages et des modèles acoustico-phonétiques grossiers sont utilisés et toutes les hypothèses de reconnaissance ayant un score élevé sont retenues. L'ensemble de ces hypothèses forme un graphe qui sera parcouru au cours de la deuxième passe avec des modèles de langages et des modèles acoustico-phonétiques plus fins afin de définir laquelle de ces hypothèses est la meilleure. Le système accepte ainsi à l'entrée un ensemble de vecteurs Y représentant le signal de parole et produit à la sortie un ensemble de mots w correspondant à la meilleure hypothèse retenue au cours de la seconde passe.

2.4 Transcription phonétique

Une étape importante dans l'élaboration d'un système de reconnaissance est la construction des modèles HMMs. Ces modèles peuvent aussi bien servir à représenter des mots que des phonèmes. Cependant, en reconnaissance de parole continue, on préfère utiliser les HMMs afin de modéliser des phonèmes plutôt que des mots. Dans le cas du français, par exemple, on considère que 40 phonèmes permettent de représenter adéquatement l'ensemble des sons qui constituent la langue, alors que le nombre de mots différents dans le corpus de test et d'apprentissage peut être supérieur à 50 000 bien souvent.

Cette utilisation des HMMs pour modéliser les phonèmes au lieu des mots ajoute une contrainte supplémentaire au système: chaque mot doit avoir une transcription phonétique. Par transcription phonétique, on entend l'ensemble des sons distincts qui forme un mot donné. La prononciation d'un mot peut varier d'une personne à l'autre, donc chaque mot peut avoir une ou plusieurs transcriptions phonétiques. L'ensemble de ces transcriptions phonétiques des mots du dictionnaire forme ce qu'on appelle un lexique.

Un des problèmes posé par le français est le phénomène de liaison. La prononciation de certains mots change en fonction du contexte. Prenons le mot «ils» par exemple, si ce mot est suivi du mot «font» comme dans «ils font», le «s» ne sera pas prononcé; cependant si le mot «ils» est suivi du mot «ont» comme dans «ils ont», alors le «s» sera prononcé comme un «z».

Pour résoudre le problème de liaison, nous avons choisi d'ajouter à la transcription phonétique classique des mots du dictionnaire des transcriptions qui tiennent compte du phénomène de liaison. Dans le cas du mot «ils» par exemple, on a deux transcriptions phonétiques: «ils» et «ilz». Cette approche nous permet d'utiliser le même système de reconnaissance qu'en anglais avec un lexique différent.

3. Tests et résultats

Nous avons testé la performance du système de reconnaissance en français sur un roman audio. Il s'agit d'un livre: «Et tournons la page ...», écrit par Solange Chaput-Rolland, et qui est lu par une femme. La durée totale de la lecture a été de 146,51 minutes. Le livre comporte 5 901 mots différents. 90% du texte a été utilisé pour entraîner le système. Les tests de reconnaissance ont été effectués sur les 10% du texte restant. Le tableau ci-dessous montre les taux de reconnaissance obtenus.

Tableau 1 : Taux de reconnaissance obtenus avec certains fichiers

Fichiers	Nombre de mots	Taux de reconnaissance
sol010	188	37,23%
sol026	366	57,92%
sol077	268	40,67%
sol094	378	52,65%
sol102	342	61,11%

Deux éléments importants retiennent notre attention dans ce tableau: d'abord le faible taux de reconnaissance obtenu, et aussi un écart significatif entre les taux de reconnaissance lorsqu'on passe d'un fichier à un autre.

Le meilleur taux de reconnaissance obtenu, lorsque l'on fait de la reconnaissance du français avec notre système, est de 61,11% alors que le

même système donne un taux de reconnaissance moyen en anglais de 90%. Il y a là un écart considérable qu'il convient de combler.

Outre le problème de liaison que nous avons mentionné précédemment, il existe en français plusieurs combinaisons de mots ou de groupes de mots qui ont exactement la même transcription phonétique, comme «chasseur» et «chasse heure», ou «eau» et «haut», et qui sont des homophones. Le système de reconnaissance arrive difficilement, compte tenu du contexte, à choisir le mot ou le groupe de mots qui convient le mieux.

4. Conclusion

Le problème de reconnaissance automatique de la parole peut être défini de façon simple: il s'agit d'associer à un ensemble de vecteurs donné Y caractérisant le signal à l'entrée, un ensemble de mots w . Pour un même mot ou une même phrase, la vitesse d'élocution ainsi que le mode de prononciation change d'une personne à l'autre. Les HMMs utilisés pour modéliser la parole permettent de tenir compte des variations temporelles et spectrales du signal à l'entrée.

Au cours du processus de reconnaissance, le système propose plusieurs séquences possibles de mots compte tenu des observations Y , il convient donc dans un second temps de choisir la séquence la plus vraisemblable, d'où l'intérêt du modèle de langage qui spécifie la probabilité que N mots se suivent. Cependant, dans le cas de la reconnaissance du français, nos expériences nous laissent croire que le modèle de langage seul ne nous permettra pas de résoudre tous les problèmes, en particulier ceux posés par les homophones comme "eau" et "haut". Un élément nouveau, pensons-nous, doit être ajouté au système afin d'accroître sa robustesse. Nos recherches se poursuivent en ce sens et nous espérons bientôt trouver une solution à ce problème.

5. Références

- [1] J. E. Hopcroft et J. D. Leman, *Introduction to Automata Theory, Languages and Computation*, Addison-Wesley, 1979.
- [2] M. Oerder et H. Ney, "Word Graphs: An Efficient Interface Between Continuous Speech Recognition and Language Understanding", *Proc. ICASSP-93*, 1993, pp. 119-122.
- [3] P. Placeway, R. Schwartz, P. Fung et L. Nyuyen, "The Estimation of Powerful

- Language Models from Small and Large Corpora", *Proc. ICASSP-93*, 1993, pp. 33-36.
- [4] F. Jelinek, B. Merialdo, S. Roukos et M. Strauss, "A Dynamic Language Model for Speech Recognition", *Proc. DARPA Workshop*, 1991, pp. 293-295.
- [5] M. H. Talon, P. Dumouchel et D. O'Shaughnessy, "Modélisation du langage pour la reconnaissance de la parole: une approche stochastique non-déterministe à contexte de longueur variable", *Association Canadienne Française pour l'Avancement de la Science*, Mai 1996.
- [6] S. Austin, R. Schwartz et P. Placeway, "The forward backward search algorithm", *Proc. ICASSP-91*, 1991, pp. 697-700.
- [7] D. B. Paul, "The lincoln tied-mixture HMM continuous speech recognition", *Proc. ICASSP-91*, 1991, pp. 656-659.
- [8] P. Kenny, R. Hollan, G. Boulianne, H. Garudadri, M. Lennig et D. O'Shaughnessy, "An A* Algorithm for very large vocabulary continuous speech recognition", *DARPA Workshop on Speech and Natural Language*, 1992.
- [9] P. Kenny, R. Hollan, G. Boulianne, H. Garudadri, Y.-M. Cheng, M. Lennig et D. O'Shaughnessy, "Experiments in Continuous Speech Recognition with a 60,000 Word Vocabulary", *International Conference on Spoken Language Processing*, Banff, Canada, 1992, pp. 225-228.

Vision artificielle

17. J. J. Gibson, *The Perception of the Visual World*, Houghton Mifflin, Boston, 1950.
18. W. H. Taylor, *The Perception of the Visual World*, Houghton Mifflin, Boston, 1950.
19. J. J. Gibson, *The Perception of the Visual World*, Houghton Mifflin, Boston, 1950.
20. J. J. Gibson, *The Perception of the Visual World*, Houghton Mifflin, Boston, 1950.
21. J. J. Gibson, *The Perception of the Visual World*, Houghton Mifflin, Boston, 1950.
22. J. J. Gibson, *The Perception of the Visual World*, Houghton Mifflin, Boston, 1950.
23. J. J. Gibson, *The Perception of the Visual World*, Houghton Mifflin, Boston, 1950.
24. J. J. Gibson, *The Perception of the Visual World*, Houghton Mifflin, Boston, 1950.

Vision artistique

L'effet du désapprentissage dans les modèles autoassociatifs de la catégorisation

Robert Proulx, Nicolas Ledoux et Jean Bégin

Université du Québec à Montréal, Laboratoire d'études en intelligence naturelle et artificielle, Montréal (Québec), CANADA. e-mail: proulx.robert@uqam.ca.

Résumé

L'implantation de systèmes efficaces de vision artificielle dans la complexité du monde naturel repose essentiellement sur l'élaboration de modèles adéquats de la catégorisation de l'information. À ce titre, les modèles connexionnistes et les systèmes à base de réseaux de neurones formels possèdent de nombreux avantages sur les modèles classiques de type computationnel. De plus, parmi l'ensemble des modèles proposés, le modèle EIDOS [1] apparaît particulièrement intéressant en ce qui a trait à ses applications en reconnaissance de forme. En effet, grâce à sa structure auto-associative et à l'utilisation d'une règle hebbienne et anti-hebbienne pour l'apprentissage, celui-ci débouche sur une modélisation naturelle de la catégorisation, tout en assurant la stabilisation automatique du processus d'apprentissage. Cependant, comme le montre certaines études récentes [1], sa sélectivité dépend de l'ajustement précis de l'un de ses paramètres libres, ce qui limite à la fois sa généralité et sa plausibilité biologique.

Dans cet exposé, nous étudions l'effet de l'introduction dans le modèle d'une procédure de désapprentissage semblable à celle proposé par Crick et Mitchison [2]. Plus spécifiquement, nous vérifions l'hypothèse que le désapprentissage permet l'élimination adéquate du bruit, sans avoir recours à des ajustements arbitraires. Les résultats des simulations montrent que sous l'application de cette procédure, le système converge vers des états finaux optimaux, et que celle-ci demeure robuste sous les variations de ses paramètres libres. Il est donc conclu qu'en plus d'accroître la plausibilité biologique du modèle EIDOS, le désapprentissage constitue une solution intéressante au problème de la stabilisation autonome de l'apprentissage dans les réseaux neuronaux non supervisés.

1. Introduction

S'il existe un problème auquel l'intelligence artificielle s'est heurtée depuis sa fondation, c'est bien celui de la reconnaissance de forme. En effet, contrairement aux systèmes experts et aux applications en résolution

de problèmes, les systèmes de reconnaissance de forme se doivent quant à eux d'être en mesure d'opérer directement dans la complexité du monde réel et de s'adapter à un environnement sans cesse variable et mal défini. Il en résulte donc que les stratégies habituelles, basées sur des algorithmes essentiellement symboliques et sériels se trouvent rapidement confrontées à l'explosion combinatoire, et deviennent, par conséquent, difficilement applicables à des domaines autres que des micro-mondes ou des univers conceptuels restreints.

C'est en fait principalement pour pallier à ce type de problème, qu'au cours de la dernière décennie, plusieurs individus se sont tournés vers une approche massivement parallèle, basée sur les opérations de réseaux neuronaux associatifs et distribués [3] [4]. Le principal avantage de ce type d'approche est qu'en plus de fournir un modèle plausible du traitement de l'information, celle-ci permet de déboucher sur des systèmes adaptatifs auto-organisés, capables de généraliser et de converger rapidement vers des états stables optimaux, et ce, même lorsque le système doit opérer en présence d'un haut niveau de bruit. De plus, le fait que ces propriétés émergent par le biais d'un processus d'apprentissage les rend parfaitement adaptés aux situations mal définies, où la structure des catégories est inconnue à l'avance et où il est, par conséquent, impossible de référer à des caractéristiques distinctives pré-établies.

Cependant, même si ce domaine a connu récemment un essor considérable, peu de modèles se sont penchés directement sur le problème de la catégorisation autonome. En effet, les modèles basés sur des règles d'apprentissage supervisées tel les modèles à rétropropagation [5] se trouvent difficilement applicables, puisque ceux-ci supposent l'existence dans le système d'une information déjà pré-catégorisée. Il en est de même des modèles compétitifs [6] et de certains modèles ACP (Analyse en Composantes Principales) [7], qui dépendent quant à eux d'indices non calculables localement comme, par exemple, la norme des vecteurs ou le nombre de catégories. Il apparaît donc que seul les modèles auto-associatifs basés sur les opérateurs de projection [8] semblent en mesure d'attaquer directement le problème de la catégorisation. À ce titre le modèle EIDOS développé dans notre laboratoire [1] constitue un candidat intéressant.

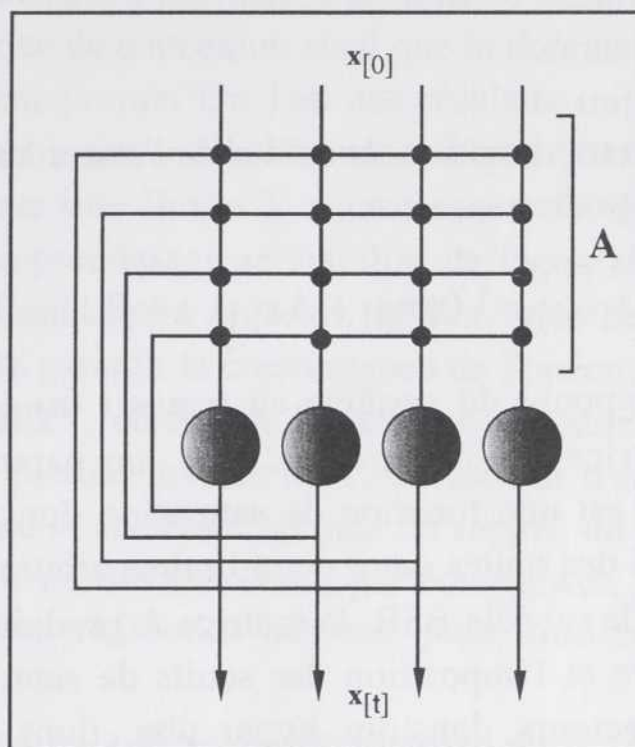


Figure 1. Illustration schématique du modèle EIDOS

2. Le modèle EIDOS

Comme pour tous les modèles connexionnistes, le modèle EIDOS se trouve entièrement spécifié par son architecture, ainsi que par sa règle de transmission et sa règle d'apprentissage.

2.1 Architecture

Issu des travaux sur le modèle BSB [9], le modèle EIDOS repose sur une architecture auto-associative récurrente, où chacune des unités du réseau neuronal établit des connexions avec l'ensemble des autres unités. Cette situation se trouve illustrée dans la figure 1. Lorsqu'un stimulus x se présente à l'entrée du réseau au temps $t=0$, celui-ci est alors constamment recyclé à travers les connexions de la matrice A , jusqu'à ce que le système se retrouve dans un état stable et invariant. L'état final ainsi obtenu constitue la *réponse* du réseau. De plus, puisqu'il n'existe qu'un nombre fini d'attracteurs par rapport à un nombre infini d'états initiaux possibles, le système effectue donc une *catégorisation* de son environnement. Cependant, les propriétés émergentes de cette opération dépendent de la forme précise de sa règle de transmission et sa règle d'apprentissage.

2.2 Transmission

Au niveau de la transmission, la valeur de l'état $\mathbf{x}$ au temps $t+1$ peut se représenter par l'équation

$$\mathbf{x}_{[t+1]} = L(\psi \mathbf{x}_{[t]} + \mathbf{A} \mathbf{x}_{[t]}), t = 0, 1, \dots, T \quad (1)$$

où $\mathbf{x}_{[t]}$ est la réponse du système au temps t ($\mathbf{x}_{[0]}$ étant le stimulus initial), $\mathbf{A}$, la matrice des connexions, et ψ , un paramètre de maintien ($0 \leq \psi \leq 1$). Ici, $L[\bullet]$ est une fonction de saturation dont le rôle consiste à maintenir l'activité des unités entre deux limites arbitraires et égales à $\pm \gamma$. Tout comme dans le modèle BSB, la matrice $\mathbf{A}$ produit un effet global de rétroaction positive et l'imposition des seuils de saturation a pour effet d'enfermer les vecteurs dans un hypercube, dont les coins stables constituent les réponses finales du système.

2.3 Apprentissage

Au niveau de l'apprentissage, la transformation que subit la matrice $\mathbf{A}$ après chaque essai engendré par la présentation d'un stimulus $\mathbf{x}$ quelconque est définie égale à

$$\mathbf{A}_{i+1} = \zeta \mathbf{A}_i + [\alpha \mathbf{x}_{[p]} \mathbf{x}_{[p]}^T + \beta \mathbf{x}_{[n]} \mathbf{x}_{[n]}^T] \quad (2)$$

où $\mathbf{A}_i$ représente la valeur de la matrice $\mathbf{A}$ au i ème essai, ζ , un paramètre général d'oubli ($0 \leq \zeta \leq 1$), $\mathbf{x}_{[p]}$ et $\mathbf{x}_{[n]}$, l'état du réseau après p et n itérations de la fonction (1) ci-dessus, et α et β , les poids respectifs des changements de type hebbien et anti-hebbien. Pour que l'algorithme fonctionne il faut et il suffit que les paramètres α , β , p et n soient choisis de façon à ce que la magnitude de $\alpha > 0$ soit supérieure à celle de $\beta < 0$, et que p soit inférieur à n [10].

2.4 Propriétés du modèle

Ce qui distingue le plus le modèle EIDOS des modèles hebbiens conventionnels, c'est surtout l'emploi conjoint de l'adaptation hebbienne et anti-hebbienne dans une même règle d'apprentissage. En effet, l'un des problèmes majeurs que doivent surmonter les modèles basés sur une règle

de Hebb simple consiste à prévenir la croissance exponentielle des valeurs propres de la matrice de connexion ainsi que la domination du système par son premier vecteur propre. Or, l'un des résultats importants concernant la double règle hebbienne/anti-hebbienne est qu'il peut être démontré [10] que celle-ci impose une limite à la croissance de chacune des valeurs propres, et que l'apprentissage se stabilise de façon automatique lorsque toutes les valeurs ont atteint ce seuil. Mieux, sous certaines conditions, l'emploi de la règle garantit la convergence de l'opérateur de transmission vers la forme $\Lambda \mathbf{X} \mathbf{X}^+$, où $\mathbf{X}$ est la matrice colonne des prototypes à identifier, $\mathbf{X}^+$, sa pseudo-inverse et Λ , un facteur d'échelle quelconque. Ceci démontre donc le caractère optimal du rappel, du moins, au sens des moindres carrés. Il ne devient donc, par conséquent, plus nécessaire de s'en remettre à des méthodes *ad hoc* pour limiter l'adaptation des connexions.

Toutefois, même si la règle d'apprentissage du modèle EIDOS semble lui attribuer des propriétés optimales, un problème se pose lorsque le système doit opérer en présence de bruit à l'apprentissage (ce qui constitue généralement la norme). En effet, dans ce cas précis chaque objet est unique, et même s'il existe un taux de croissance différentiel des valeurs propres associées au signal (prototypes), par rapport à celles associées au bruit, le fait que leur magnitude soit bornée par la règle entraîne que, tôt ou tard, toutes les valeurs propres s'égaleront et le système perdra sa sélectivité. Jusqu'à maintenant, Bégin et Proulx [1] ont proposé l'utilisation d'un paramètre d'oubli pour contourner ce problème. Cependant, comme celui-ci applique à chaque essai une réduction uniforme à chacune des valeurs propres, il faut donc l'ajuster de façon précise à partir des connaissances *a priori* des particularités de chacune des situations. Un tel état de fait limite donc la généralité et la plausibilité biologique du modèle, et il devient nécessaire de se tourner vers des méthodes moins contraignantes pour permettre l'élimination efficace du bruit.

2.5 La filtration du bruit et le désapprentissage

L'une des façons de permettre à un réseau neuronal d'éliminer les valeurs propres associées à la présence de bruit consiste tout simplement à insérer entre les essais d'apprentissage, des essais de désapprentissage,

c'est-à-dire des essais ne comportant que du bruit, et où les signes des paramètres de la règle d'apprentissage se trouvent inversés. De cette façon, à long terme, seules les valeurs propres associées au signal ne devraient émerger et atteindre leur seuil de saturation. Une telle procédure a déjà été proposée antérieurement par Crick et Mitchison [3] et celle-ci a été associée au phénomène biologique du rêve, dont le rôle au niveau de la consolidation des apprentissages a souvent été mentionné.

Dans les paragraphes qui suivent, nous vérifions, à l'aide de simulations d'une tâche en reconnaissance de caractères optiques, l'hypothèse que le désapprentissage permet d'éliminer adéquatement le bruit, et que la procédure est relativement insensible aux variations de ses paramètres libres.

3. Méthode de simulation

Les expériences de simulation procèdent toutes à partir d'une méthode commune et similaire à celle proposée par Bégin et Proulx [1].

3.1 Stimuli

Le répertoire des *prototypes* est composé des neuf caractères alphabétiques: "A, E, H, I, N, O, R, S, T", lesquels sont d'abord représentés visuellement dans une grille 5 x 7, pour être ensuite encodés sous forme de vecteurs à 35 dimensions, puis normalisés. L'encodage d'un vecteur s'effectue par l'attribution à chacune de ses composantes de la valeur numérique -1 ou +1, selon l'état blanc ou noir du pixel correspondant. Un *stimulus* pour un essai quelconque est obtenu par la sélection au hasard de l'un des prototypes, suivi de sa contamination par du bruit. La contamination par le bruit s'effectue par l'ajout d'un vecteur de nombres aléatoires distribués selon une loi normale $N(0, \sigma)$, où la valeur de σ dépend du pourcentage de bruit désiré. Le pourcentage de bruit est défini comme étant le rapport, en pourcentage, entre la norme du bruit et la norme du signal, et celui-ci est fixé à 20% pour toutes les expériences.

3.2 Procédure d'apprentissage de désapprentissage et de rappel

Pour réaliser un essai d'apprentissage, la procédure consiste d'abord à engendrer un nouveau stimulus selon la méthode exposée ci-dessus, puis à

l'acheminer dans le réseau neuronal pour n itérations, au cours desquelles celui-ci applique rigoureusement les transformations décrites par les équations (1) et (2).

Dans le cas d'un essai de désapprentissage, la même méthode est employée, sauf que le stimulus ne se compose que d'un bruit de norme comparable à celle des stimuli, et que les paramètres α et β de la règle d'apprentissage sont de signes inversés et de magnitudes réduites de 50%.

Un test de rappel consiste simplement à générer un nouveau stimulus, puis à lui appliquer successivement la règle de transmission, jusqu'à ce que le système atteigne un état invariant. Un succès est défini comme une réponse qui correspond exactement au prototype utilisé.

Finalement, pour l'ensemble des simulations les paramètres libres du modèle sont fixés aux valeurs typiques suivantes: $\psi=0.95$, $\gamma=2.0$, $\zeta=1.0$, $\alpha=0.0050$, $\beta=-0.0025$, $p=5$ et $n=10$. Il convient de noter que la valeur du paramètre d'oubli, ζ , est délibérément fixée à 1.0 pour éliminer totalement son effet et isoler les effets du désapprentissage.

4. Résultats

L'exposé des résultats se divise en trois parties. La première rapporte les résultats de simulations visant à étudier l'effet pur du désapprentissage, de même que sa stabilité en fonction du nombre d'essais. La seconde, quant à elle, étudie l'effet d'une variation de la fréquence, c'est-à-dire le nombre absolu d'essais de désapprentissage par rapport au nombre d'essais d'apprentissage, tandis que la troisième étudie l'effet du patron de désapprentissage, c'est-à-dire le cycle d'apparition à fréquence fixe de chacun des deux types d'essais.

Dans tous les cas, la variable dépendante étudiée est le spectre des valeurs propres de la matrice de connexion. La performance au rappel n'est mesurée que pour s'assurer de la validité du processus de catégorisation. Dans la situation idéale, le fait qu'il n'y ait que neuf prototypes entraîne que la matrice de connexion devrait se développer de façon à n'engendrer que 9 valeurs propres positives et 26 valeurs propres résiduelles inférieures ou égales à zéro.

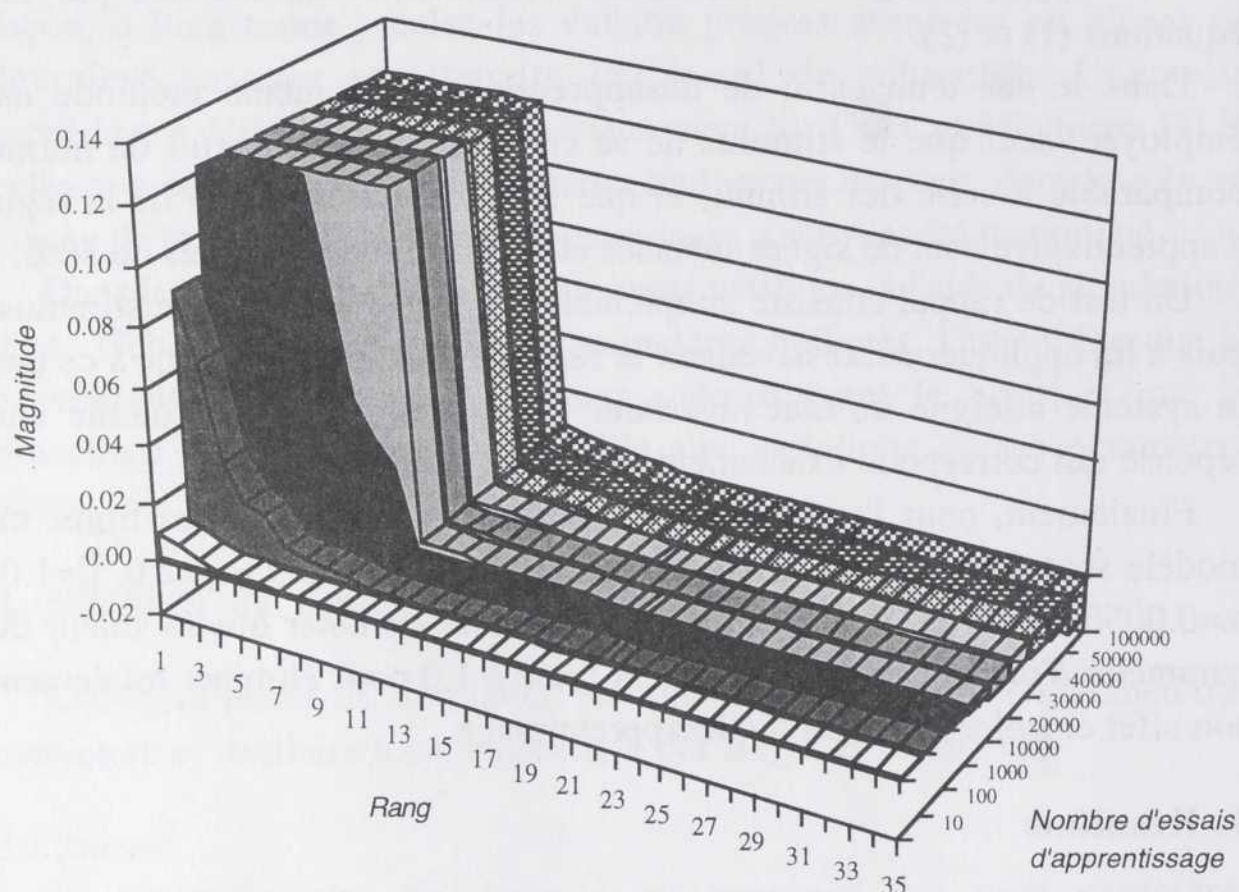


Figure 2. Progression du spectre des valeurs propres en fonction du nombre d'essais d'apprentissage

4.1 Effet du désapprentissage

La figure 2 ci-dessus rapporte les spectres des valeurs propres obtenus suite à une variation de 10 à 100 000 essais d'apprentissage. Dans cette expérience, la fréquence, de même que le patron des essais de désapprentissage est maintenu à un pour un. À l'examen de cette figure, il est facile de constater l'efficacité du processus. En effet, le spectre se stabilise complètement avec 9 valeurs propres positives après 10 000 essais, et la poursuite des essais jusqu'à 100 000 n'altèrent à peu près pas sa valeur. De plus les tests de rappel effectués établissent la performance de catégorisation à 100%. Il donc conclu que contrairement au paramètre d'oubli, le désapprentissage permet la stabilisation autonome du spectre des valeurs propres. Cependant, il reste à démontrer que le processus demeure stable sous la variation de ses paramètres libres tels la fréquence et le patron de succession des essais.

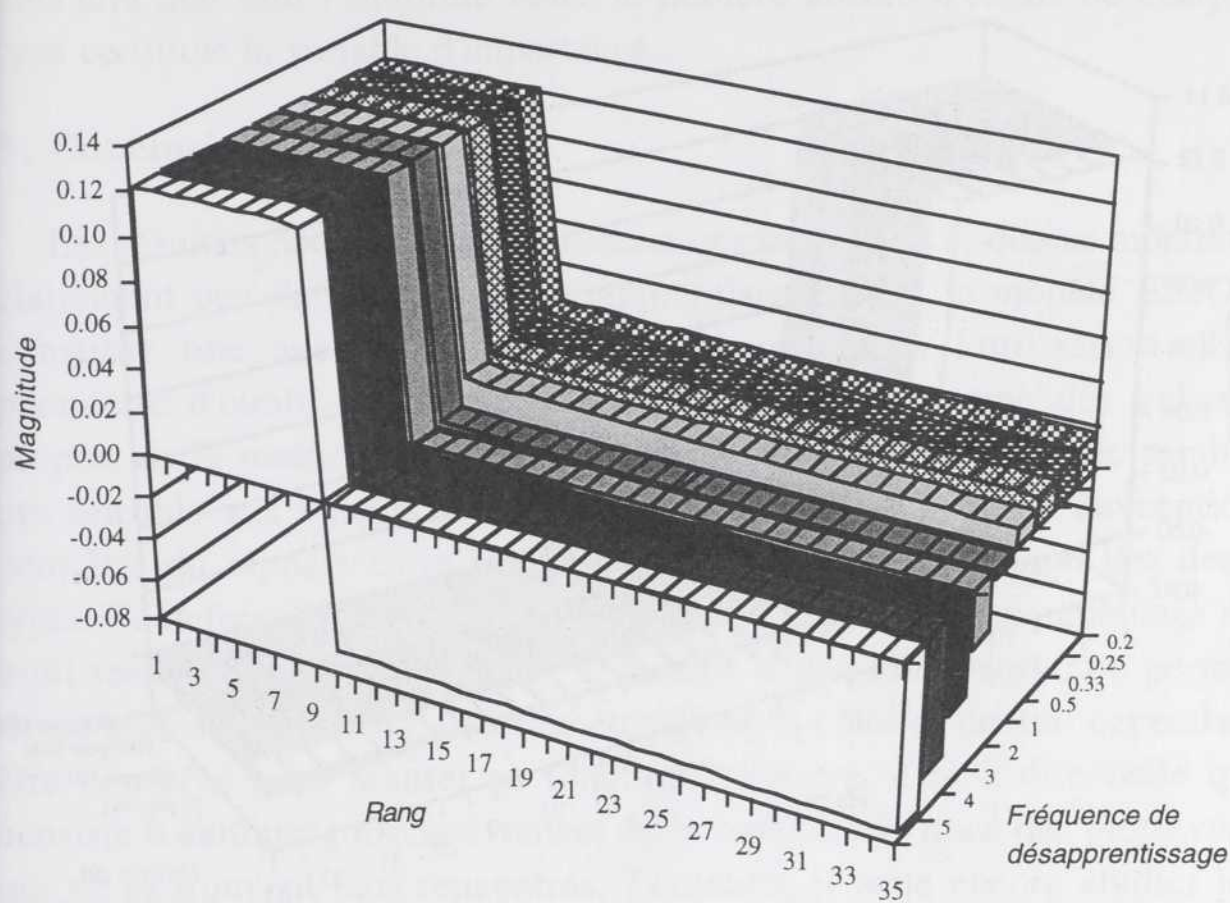


Figure 3. Progression du spectre des valeurs propres en fonction de la fréquence du désapprentissage

4.2 Effet de la fréquence

La figure 3 ci-dessus illustre les spectres des valeurs propres obtenus suite à une variation de la fréquence des essais de désapprentissage allant de 5 à 0.2, c'est-à-dire, de cinq essais de désapprentissage par essai d'apprentissage, à un essai de désapprentissage par cinq essais d'apprentissage. Le nombre total d'essais d'apprentissage est maintenu fixe à 10 000. À l'examen de la figure, on remarque que le processus semble toujours opérationnel sous la variation de la fréquence. Ceci se trouve d'ailleurs confirmé par les tests de rappel qui rapporte encore une performance de 100%. Toutefois, on constate une inversion du signe des valeurs propres résiduelles lorsque le seuil de 1.0 se trouve franchi. De plus, à mesure que diminue le nombre d'essais de désapprentissage par rapport au nombre d'essais d'apprentissage, le système semble de moins en moins capable d'écraser les valeurs propres résiduelles. La présence de valeurs propres négatives ne devraient toutefois pas poser de problèmes,

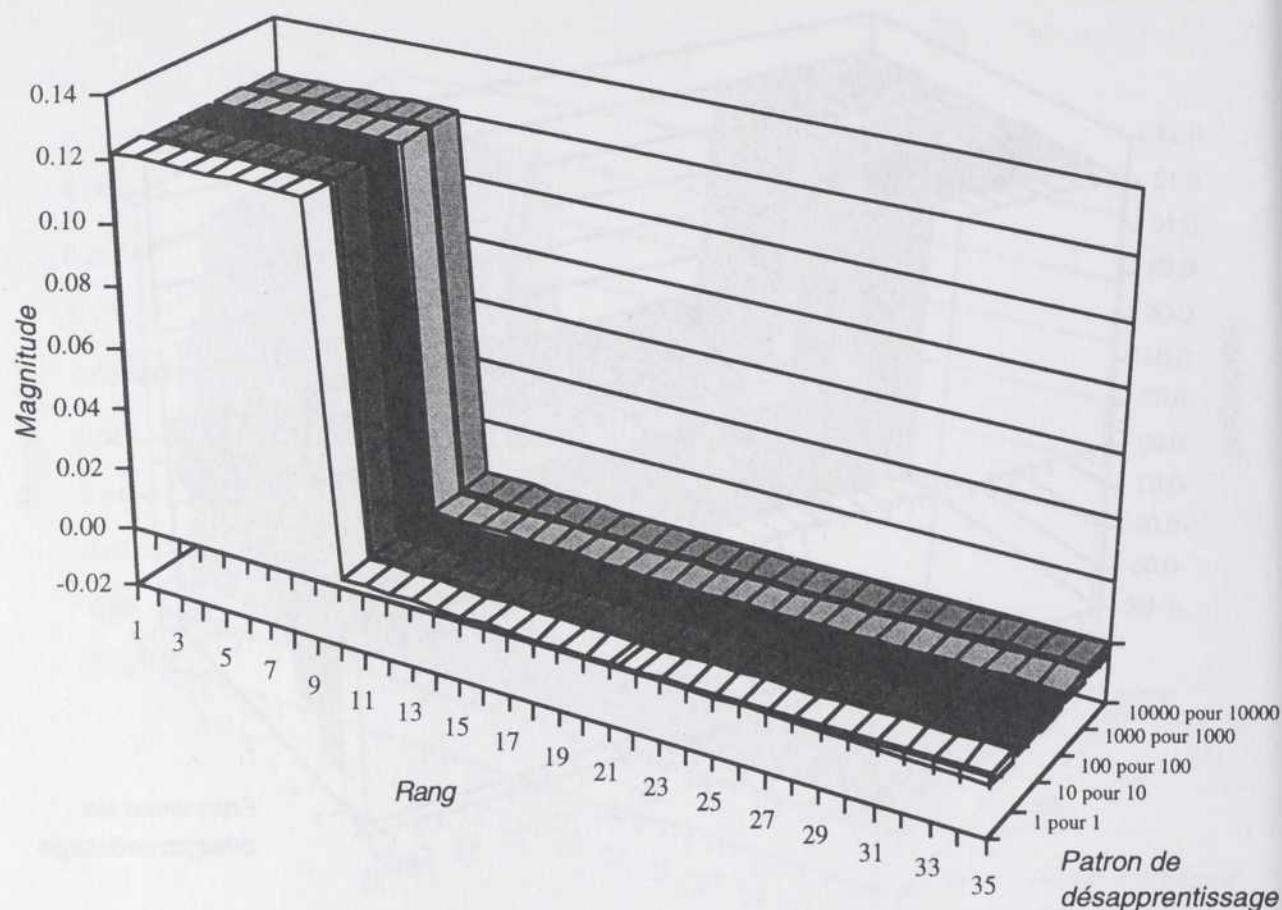


Figure 4. Progression du spectre des valeurs propres en fonction du patron de désapprentissage

dans la mesure où celles-ci demeurent supérieures à -1.0. Par contre, il reste à vérifier si la croissance des valeurs propres positives se maintient à un niveau suffisamment bas à des fréquences jugées normales. De plus, la valeur même du seuil critique de 1.0 est arbitraire, puisque celle-ci dépend probablement du niveau de bruit utilisé. Il faut donc conclure que d'autres expériences devraient se pencher de façon plus précise sur le rapport entre le seuil critique et le pourcentage de bruit.

4.3 Effet du patron de désapprentissage

Puisque l'effet du patron de désapprentissage se trouve partiellement confondu dans les variations de fréquence, il convient donc d'isoler cette variable et d'étudier ses effets à fréquence fixe. La figure 4 ci-dessus illustre les variations du spectre des valeurs propres suite à des variations du patron allant de 1 pour 1, à 10 000 pour 10 000. Les résultats sont assez étonnants puisqu'il démontrent qu'en aucun cas la cyclicité entre l'apprentissage et le désapprentissage ne semble intervenir. Il faut donc en

conclure que seul l'équilibre entre le nombre absolu d'essais de chaque type constitue la variable d'importance.

5. Conclusion

Les résultats obtenus dans les simulations présentées ci-dessus montrent clairement que l'utilisation du désapprentissage dans le modèle EIDOS constitue une avenue intéressante, par rapport à l'utilisation d'un paramètre d'oubli, pour assurer la stabilisation autonome des valeurs propres de la matrice de connexion. De plus, cette procédure ne semble pas sensible aux valeurs précises de ses paramètres libres, à l'exception peut-être du rapport entre le nombre total d'essais de chacun des deux types. Toutefois, l'interaction précise entre ce rapport et le pourcentage de bruit reste à explorer. Enfin, si l'utilisation d'une telle procédure permet de ne plus avoir recours au paramètre d'oubli, celui-ci devrait cependant être conservé pour assurer sa fonction première, c'est-à-dire celle qui consiste à éliminer progressivement de la mémoire la trace des prototypes qui ne se trouvent plus rencontrés. Toutefois, il reste encore étudier les interactions possibles entre ce paramètre et ceux du désapprentissage.

5. Références

- [1] J. Bégin, R. Proulx, "Categorization in unsupervised neural networks: The EIDOS model", *IEEE Trans. on Neural Networks*, vol. 7, no 1, 1996, pp. 147-154.
- [2] F. Crick, G. Mitchison, "REM sleep and neural nets", *Journ. of Mind and Behavior*, no 7, 1986, pp. 229-249.
- [3] J. Hertz, A. Krogh, R. G. Palmer, "Introduction to the theory of neural computation", Addison-Wesley, Reading (MA), 1991.
- [4] S. Y. Kung, "Digital neural networks", Prentice Hall (NJ), 1993.
- [5] D. E. Rumelhart, J. L. McClelland (Eds.), "Parallel distributed processing. Vol. 1: Foundations", MIT Press, Cambridge, 1986.
- [6] S. Grossberg, "Competitive learning: From interactive activation to adaptive resonance", *Cognitive Science*, no 11, 1987, pp.23-63.
- [7] E. Oja, "Neural networks, principal components, and subspaces", *Int. Journ. of Neural Syst.*, no 1, 1989, pp.61-68.
- [8] T. Kohonen, "Self organisation and associative memory", Springer Verlag, Berlin, 1989.
- [9] J. A. Anderson, J. W. Silverstein, S. A. Ritz, R. S. Jones, "Distinctive features, categorical perception and probability learning: some applications of a neural model", *Psychological Review*, no 84, 1977, pp. 413-451.
- [10] R. Proulx, J. Bégin, "The use of a dual hebbian and anti-hebbian learning rule for categorization in neural networks", 1995, (manuscrit soumis pour publication).

Revue des principaux modèles de réseaux neuroniques en vision artificielle préattentive

Richard Lepage

*Ecole de technologie supérieure, Département de génie électrique,
Laboratoire d'imagerie, de vision et d'intelligence artificielle,
Montréal (Québec), CANADA.
e-mail: lepage@ele.etsmtl.ca*

Résumé

Une revue des principaux modèles de réseaux neuroniques artificiels appliqués à des systèmes de vision numérique est présentée. Un modèle computationnel du système de vision est d'abord exposé, identifiant les niveaux de représentation de l'information visuelle requis et les divers algorithmes permettant de passer d'un niveau de représentation au suivant. Des réalisations neuroniques de quelques algorithmes sont explorés.

1. Introduction

La vision artificielle est considérée, dans le cadre de cet article, comme un processus de traitement de l'information. Un tel processus, selon Marr [1], doit être étudié à trois niveaux afin d'en comprendre entièrement le fonctionnement. Le niveau le plus général, la *théorie computationnelle*, identifie le but poursuivi par le processus et la théorie pour y parvenir. On distingue deux catégories de système de vision artificielle selon le but poursuivi: une première catégorie de système a comme objectif de reconnaître et classifier des formes bidimensionnelles - on pense ici à la reconnaissance d'écriture manuscrite, la classification de dessins au trait et l'inspection automatisé d'images provenant d'une caméra; tandis qu'une seconde catégorie de système a comme objectif la reconnaissance et la localisation d'éléments composant une scène tridimensionnelle observée. Nous nous limitons, dans cette étude, à cette seconde catégorie de machine de vision. Le second niveau d'étude du processus de vision artificielle consiste à identifier et décrire les différents niveaux de représentation de l'information visuelle de même que les différents algorithmes permettant de passer d'un niveau de représentation au suivant. Un *modèle computationnel* permet d'illustrer cette décomposition du processus de vision en niveaux de représentation et en algorithmes utilisés pour réaliser la théorie computationnelle. Enfin, le troisième niveau de compréhension du processus de vision artificielle identifie le *support matériel de réalisation* du processus. Cet article propose une revue

de la littérature sur la réalisation de systèmes de vision artificielle tridimensionnelle ayant comme support une architecture de réseaux neuroniques artificiels.

2. Modèle computationnel du système de vision numérique

Un système de vision extrait de l'information à partir d'une séquence d'images d'illuminance monoculaires ou stéréos. L'information est traitée afin d'en arriver à une compréhension raisonnable du monde environnant et permettre ainsi au mécanisme d'effectuer des tâches comme par exemple de se diriger vers un but identifié, manipuler des objets de l'environnement ou éviter des collisions. Une modélisation algorithmique du système de vision consiste à décomposer tout le processus en plusieurs niveaux de représentations internes qui mettront en relief différents aspects de l'information visuelle analysée. Le passage d'une représentation interne à une autre est réalisé par un ou plusieurs algorithmes [2]. La figure 1 illustre la modélisation algorithmique du système de vision basée sur la représentation originale de Marr [3-5] et modifiée par Lepage et Poussart [6] afin d'explicitier la nature multirésolution des premiers niveaux de représentation visuelle, remplacer l'algorithme d'interpolation pour générer l'esquisse 2,5D par un de fusion d'algorithmes, et enfin de permettre à l'algorithme d'analyse de l'ombrage d'alimenter directement celui de fusion. Cette modélisation est caractérisée par un flux d'information qui procède du bas vers le haut. Le croquis, par exemple, est généré à partir de l'image d'illuminance et n'est pas influencé par les niveaux supérieurs de représentation de données. Ce sens unique de bas vers le haut utilisé pour construire les différents niveaux de représentation est contesté par de plus en plus de chercheurs [7] qui observent des interactions haut $\rightarrow$ bas de plus en plus fréquemment.

Le premier niveau de représentation interne est constitué des **images d'illuminance** gauche et droite qui sont produites par les capteurs correspondants du système de vision. Le niveau suivant de représentation, le **croquis**, est une description des variations brusques d'illuminance, appelées arêtes, détectées dans l'image d'entrée. Les arêtes correspondent à des caractéristiques physiques de la scène observée tels un changement de l'orientation de la surface, un changement d'illumination (ombrage), une occlusion de surfaces ou un marquage de la surface (e.g. texture). Le croquis permet une nette diminution de la densité d'information à traiter afin de générer les niveaux ultérieurs de représentation. Lindeberg [8, 9] propose

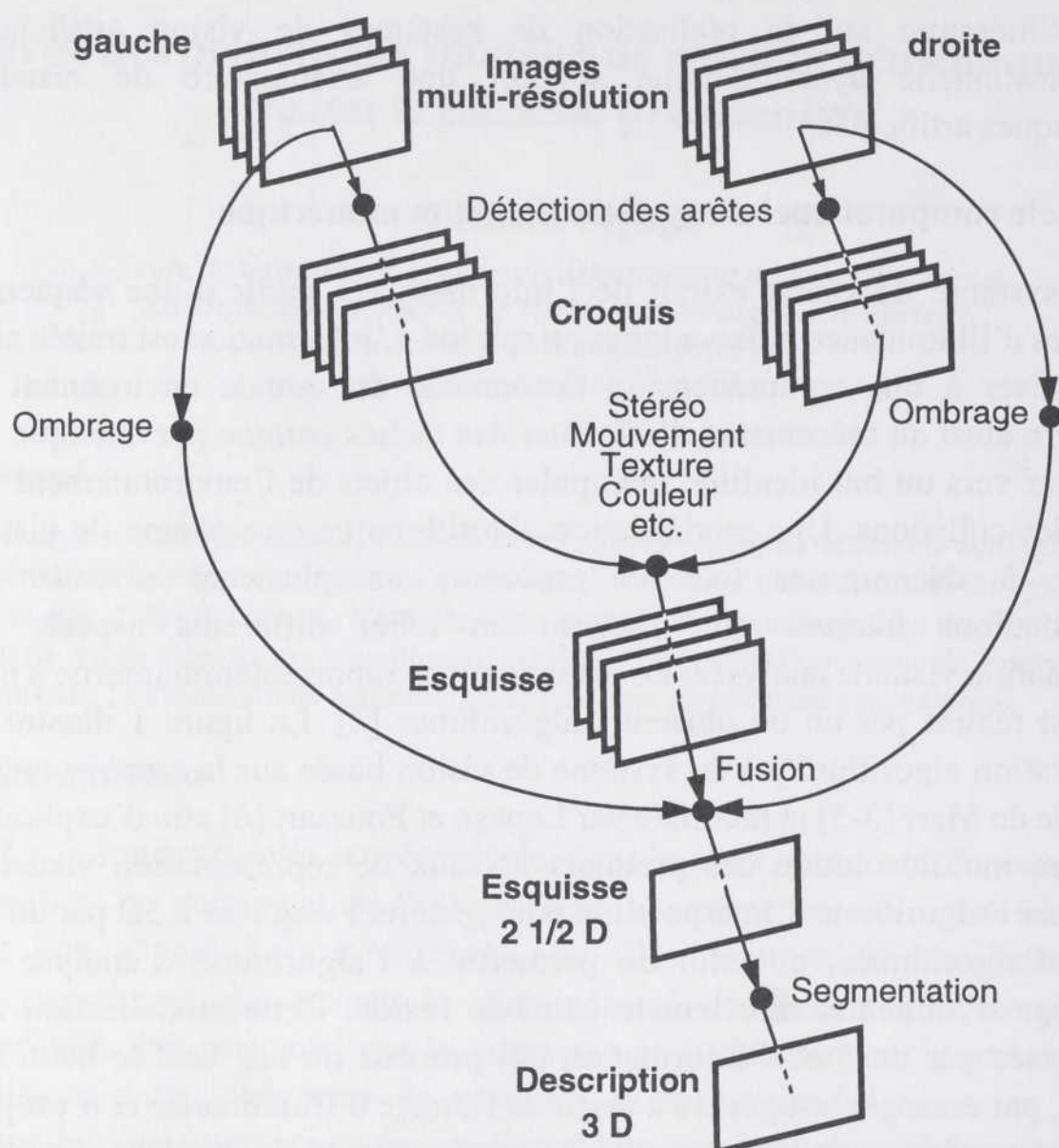


Figure 1 Modélisation algorithmique du système de vision. L'information présente dans l'image est codifiée à plusieurs niveaux de représentation interne. Des algorithmes permettent de passer d'une représentation à une autre.

d'utiliser des primitives encore plus simples pour construire le croquis. Ces primitives sont les taches ou contours fermés qui sont obtenus à partir de l'image d'illuminance. Une description multirésolution avec comme primitives les taches détectées dans l'image d'illuminance a aussi été formulée. La représentation proposée par Lindeberg est plus simple mais plus difficile à réaliser avec des réseaux neuroniques artificiels qui s'accommodent bien d'opérations locales et plus difficilement d'opérations plus régionales comme le demande le processus de détection des taches. L'algorithme pour passer de la représentation par niveaux de gris (image d'illuminance) au croquis des changements abruptes d'illuminance appartient à la famille des **détecteurs d'arêtes**.

L'**esquisse** est une représentation clairsemée qui consiste en une estimation grossière de la profondeur des points caractéristiques de la scène situés sur les surfaces visibles des objets. Plusieurs algorithmes permettent de construire l'esquisse, dont les plus utilisés sont la stéréoscopie, l'ombrage ou la couleur, le mouvement, et la texture.

L'**esquisse 2,5D** est une représentation caméra-centrique de la profondeur de toutes les surfaces visibles de la scène. Une profondeur est donc assignée à chaque coordonnée de la scène et cette information est codée dans un référentiel centré sur l'un des capteurs. Un algorithme d'**interpolation de surface** permet de raffiner l'esquisse et ainsi de générer l'esquisse 2,5D. Un algorithme de **fusion** peut être utilisé lorsque plusieurs algorithmes complémentaires sont utilisés pour la génération de l'esquisse 2,5D. L'existence même de l'esquisse 2,5D en tant que forme de représentation interne de données est contestée par l'école de Grossberg [10, 11] à Boston. Cette école préconise plutôt une forte interaction entre les différents algorithmes et des liens de contre-réaction entre les différents niveaux de représentation. Il en résulte que la description 3D est directement générée à partir des images d'illuminance et des croquis.

La **description 3D** constitue le dernier niveau du système de vision. Ce niveau décrit symboliquement la forme des objets visibles et leur organisation spatiale dans un système de coordonnées objet-centrique. Un algorithme de segmentation permet de générer la description 3D des éléments de la scène observée en regroupant l'information présente dans l'esquisse 2,5D.

3. Extraction des caractéristiques de l'image

La première étape du processus de vision consiste à extraire les arêtes détectées sur les images d'illuminance qui correspondent à des caractéristiques physiques de la scène tridimensionnelle observée, mais également à des artefacts générés par exemple par le bruit de numérisation ou le mécanisme de détection et qu'il faut soustraire du croquis. La génération du croquis nécessite donc à la fois un algorithme efficace de détection des arêtes présentes dans l'image d'illuminance et d'atténuation des artefacts.

3.1 Détection directionnelle des arêtes

La détection d'arêtes par une mesure du maximum de la dérivée première de l'image calculée dans une direction spécifiée est une méthode de détection

qui a fait l'objet de nombreuses recherches. Cette méthode offre l'avantage d'être peu bruyante dû à l'utilisation de la dérivée première et offre la possibilité d'obtenir des arêtes minces en choisissant soigneusement la direction selon laquelle la dérivée est calculée. Ces méthodes sont issues principalement des travaux de Canny [12]. Ce dernier a trouvé que l'opérateur optimal G_n pour trouver une arête bruitée est la dérivée première directionnelle (selon la direction $\hat{n}$ du gradient) d'une fonction gaussienne G_σ de résolution σ . Une arête est alors définie comme le maximum local (dans la direction du gradient) de l'opérateur G_n convolué avec l'image I .

Une réalisation neuronique d'un opérateur de type Canny fait intervenir plusieurs couches de neurones, chacune à une orientation spécifique. L'arête en chaque pixel est choisie selon l'orientation dominante détectée et selon l'orientation des arêtes dans le voisinage [13-17]. Ces travaux s'inspirent largement de ce qui est connu du fonctionnement du cortex visuel chez les mammifères et se prêtent donc particulièrement bien à une réalisation par réseaux neuroniques. La figure 2 illustre le fonctionnement de ce système de

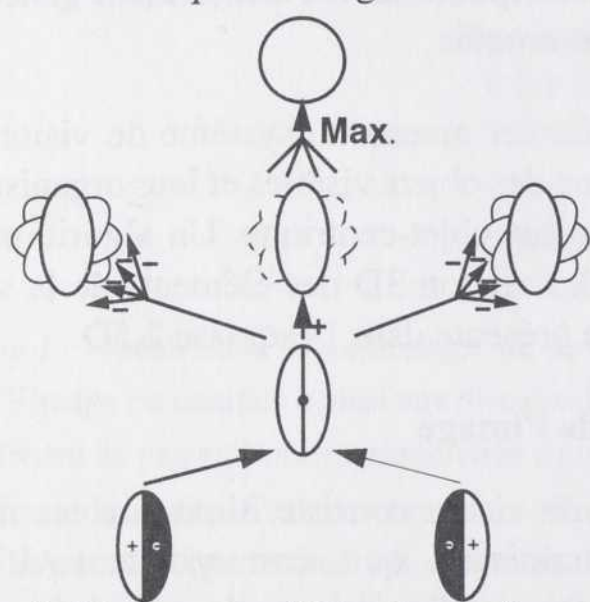


Figure 2 Version simplifiée du système BCS ("Boundary Contour System") développé par Grossberg et son équipe. La première rangée est constituée de cellules simples qui détectent les arêtes dans une direction et une orientation données. La seconde couche est constituée de cellules complexes qui ne sont sensibles qu'à l'orientation de l'arête. Les troisième et quatrième couches constituent les deux couches de cellules hypercomplexes.

détection d'arêtes. Le premier étage est constitué de **cellules simples** qui détectent une arête dans une direction et un sens donnés et à une résolution spatiale spécifique qui dépend de l'étendue du champ récepteur. Les arêtes sont détectées selon 4 orientations: horizontale, 45° , verticale (tel qu'illustré sur la figure), et 135° . Carpenter *et al* [14] utilisent deux grosseurs de champ récepteur, et donc deux échelles spatiales d'analyse. La spécification du champ récepteur détermine l'opération effectuée par la cellule. Une opération simple de différentiation entre le côté

droit et gauche de la cellule est calculée dans le type de détecteur proposé dans [14], mais d'autres champs récepteurs peuvent être utilisés, tel une cellule de Gabor [18] qui modélise assez bien les cellules simples biologiques ou une cellule de Grossberg-Todorovic [19] qui est construite à partir d'une combinaison pondérée de filtres orientés simulant des Différences de Gaussiennes. Le second étage du détecteur est composé de **cellules complexes** qui produisent une détection insensible à la direction du contraste. Pour ce faire, chaque cellule complexe est connectée aux deux cellules simples de même orientation et de même résolution spatiale mais de direction de contraste opposée. Les **cellules hypercomplexes** forment les deux derniers étages du détecteur. Ce sont deux étages de compétition qui servent à supprimer le bruit près des arêtes alors que le bruit loin des arêtes est atténué grâce à l'effet moyennneur des cellules simples et complexes. Les cellules hypercomplexes au voisinage d'une cellule complexe y sont connectées par des liens de poids positif (liens d'excitation) si les cellules sont dans l'axe d'orientation de la cellule complexe et en plus, sont elles-mêmes orientées dans la même direction que la cellule complexe. Toutes les autres cellules hypercomplexes dans le voisinage y sont connectées par des liens de poids négatifs (liens d'inhibition). Les dimensions du voisinage sont choisies en fonction des échelles spatiales utilisées. Cet étage de compétition renforce les arêtes connectées, et donc les contours, et inhibe les arêtes isolées qui sont habituellement des arêtes bruitées. Le second étage de compétition est simplement un réseau de type "le gagnant prend tout" qui sélectionne l'activité maximale à chaque position de pixel.

3.2 Détection isotrope du gradient

La détection d'arêtes par une mesure du maximum de la dérivée première isotrope consiste à détecter la présence d'une arête par une seule mesure indépendante de la direction de l'arête. L'amplitude du gradient local est généralement utilisée comme paramètre mesuré. Cette méthode permet une réalisation simple du circuit de détection mais n'offre pas la robustesse de la méthode du maximum de la dérivée première directionnelle.

La méthode de mesure du gradient utilisée seule n'offre pas une localisation précise de l'arête et est plus susceptible de détecter des artefacts parce que c'est une méthode de détection ponctuelle qui ne tient pas compte des arêtes détectées dans le voisinage. Lepage et Poussart [20] utilisent cette méthode de détection surtout à cause de la simplicité du réseau neuronique

résultant. L'amincissement des arêtes de même que la diminution du nombre d'artefacts sont obtenus grâce à une représentation multirésolution de l'image d'illuminance et à une interaction entre les arêtes détectées aux différentes échelles spatiales. D'autres auteurs proposent des méthodes qui vont dans le même sens: Fleck [21] mesure la dérivée par une différence finie et améliore la détection en combinant des mesures de différence finie de différentes largeurs; et McMichael [22] pour sa part, mesure le gradient de plusieurs images d'une scène éclairée de différentes façons et combine ces mesures, ce qui atténue l'effet de l'ombrage et renforce la détection des arêtes de l'image causées par des caractéristiques de la scène.

3.3 Passage par zéro de la dérivée seconde

La détection isotrope d'arêtes par une mesure du passage par zéro de la dérivée seconde de l'image est une méthode de détection qui fait l'objet d'une recherche diversifiée et qui suscite encore aujourd'hui énormément d'intérêt. Cet engouement pour cette méthode provient du fait qu'il existe un opérateur linéaire simple du second ordre qui détecte une variation brusque de l'illuminance indépendamment de la direction de la variation d'illuminance. Cet opérateur est le Laplacien et opère sur un voisinage restreint du pixel d'intérêt. L'intérêt pour cette méthode provient aussi du fait que l'opérateur de Laplacien peut facilement être combiné avec un filtre passe-bas gaussien et ainsi rendre la détection d'arêtes sensible à l'échelle spatiale. Un dernier avantage apprécié de cette méthode est qu'elle permet, en principe, une localisation intra-pixel précise de l'arête puisque c'est un passage par zéro de l'image convoluée qui indique la présence d'une arête et que ce passage peut être interpolé à partir des valeurs positive et négative mesurées.

La méthode de détection du passage par zéro de la dérivée seconde isotrope est issue principalement des travaux de Marr et Hildreth [23, 24]. Le détecteur de Marr-Hildreth, comme il est habituellement appelé, s'inspire des connaissances physiologiques du système visuel des mammifères: on croit que les contrastes de l'image rétinienne à certaines échelles spatiales sont d'abord accentués par un opérateur neuronique semblable à celui proposé par Marr et Hildreth et que les passages par zéro de l'image résultante de même que leur orientation sont ensuite détectés pour former le croquis.

3.4 Détection des arêtes par apprentissage supervisé

Les arêtes correspondent à des variations brusques d'illuminance. De telles variations sont détectées avec un opérateur de différentiation du premier ou du

second ordre, opérateur qui sera le plus souvent combiné avec un filtre passe-bas afin de diminuer les artefacts générés par l'opération de différentiation. La traduction neuronique de ces opérateurs résulte en des connexions qui relient la couche image à la couche du croquis avec des poids fixes. Des recherches récentes explorent différentes formes d'apprentissage afin de trouver d'autres valeurs de poids des connexions qui parviennent à la couche du croquis. Moura et Martins [25] utilisent un réseau à 3 étages. Les deux premiers étages sont interreliés par des liens à poids fixes qui réalisent un filtre passe-bas gaussien pour le premier étage et un masque de Sobel pour le second afin de détecter l'amplitude et la direction du gradient. Le dernier étage en est un de satisfaction de contraintes pour éliminer le bruit et construit des liens excitateurs entre des neurones voisins orientés perpendiculairement au gradient et des liens inhibiteurs entre des neurones voisins orientés parallèlement au gradient. Bhathia *et al* [26] proposent, à la même époque, un réseau à un seul étage, l'étage du croquis, relié à la couche de l'image d'illuminance. Chaque emplacement du croquis est occupé par quatre neurones, qui indiquent l'illuminance moyenne de même que la force et la direction de l'éventuelle arête détectée. Chaque groupe de 4 neurones du croquis sont interreliés par des liens inhibiteurs pour augmenter la compétition et chaque groupe est relié à un champ récepteur de 16×16 pixels en dessous, sur l'image d'illuminance. Deux séries de connexions composent le champ récepteur, une première à poids fixes réalise un filtre gaussien et l'autre série doit être entraînée de façon non-supervisée à l'aide d'une règle hebbienne modifiée. L'image utilisée pour la phase d'entraînement est constituée de cercles concentriques pour avoir un bon échantillonnage d'arêtes de force et d'orientations variées.

Un peu plus tard, Etemad et Chellappa [27] ont proposé un réseau neuronique approximateur de fonctions et entraîné par rétro-propagation du gradient d'erreur pour résoudre le problème de détection d'arêtes. Le réseau proposé comporte une couche d'entrée de 3×3 qui balaie l'image d'illuminance, d'une couche cachée et d'une couche de sortie également de 3×3 qui représente l'arête détectée débarrassée du bruit. Un ingénieux système de vote entre les couches de sortie voisines permet de diminuer le bruit et les artefacts. Mkaouar et Lepage [28] proposent le même type de réseau à rétro-propagation du gradient d'erreur, mais cette fois en utilisant une couche d'entrée de 5×5 , 2 couches cachées, et un seul neurone à la sortie pour indiquer la présence ou l'absence d'une arête au centre de la couche d'entrée. Les

résultats sont encourageants et des travaux se poursuivent pour interpréter les configurations de poids obtenues et améliorer les performances et diminuer le temps nécessaire pour l'apprentissage.

3.5 Détection multirésolution

Les caractéristiques physiques d'une scène se retrouvent à plusieurs échelles spatiales d'observation. Les artefacts, quant à eux, sont générés à haute résolution et correspondent à de petites variations rapides dans l'image, causés par le bruit de numérisation et par l'opération de différentiation. L'analyse multirésolution consiste à débiter l'analyse à basse résolution, à un niveau où l'on risque peu d'être en présence d'artefacts, et à ensuite poursuivre les arêtes détectées vers les images à plus haute résolution afin de mieux localiser les arêtes détectées à plus basse résolution et ainsi atténuer la présence des artefacts.

Witkin [29] est considéré un pionnier et a jeté les bases de l'analyse multirésolution et étudié le comportement des passages par zéro d'un signal temporel le long de l'échelle de résolution. Yuille et Poggio [30] ont étendu les résultats de Witkin pour les images bidimensionnelles. Bergholm [31] propose par la suite un algorithme pour la poursuite des arêtes dans l'espace continu de résolution. Burt [32], et plus récemment Jolion et Rosenfeld [33] proposent la pyramide comme une forme de représentation multirésolution très compacte. Lindeberg et Perona [34, 35] proposent diverses méthodes pour l'analyse multirésolution. L'analyse multirésolution repose sur la prémisse que des caractéristiques ne sont pas générées à mesure que la résolution diminue, et l'on démontre que seul le filtre passe-bas gaussien peut être utilisé pour générer les différents niveaux de résolution [36-38].

Les réalisations neuroniques de détecteurs d'arêtes multirésolution ne font pas l'objet de recherche répandue probablement à cause de la complexité des modes de représentation de l'information visuelle multirésolution. Lepage et Crevier [20] proposent une extraction multirésolution des arêtes basée sur une représentation pyramidale et utilisant un neurone de vigilance à chaque niveau de la pyramide pour contrôler le renforcement des arêtes dans une direction basse $\rightarrow$ haute résolution. Mkaouar et Lepage [39] proposent une extraction multirésolution des arêtes en utilisant une pyramide pour accélérer les simulations et un filtre gaussien récursif pour la génération des niveaux de représentation multirésolution. Ce modèle leur sert de base pour générer la base d'apprentissage dans [28].

4. L'esquisse

L'esquisse est une représentation clairsemée de la profondeur mesurée à certains points caractéristiques de la scène. L'esquisse 2.5D est obtenue en complétant les mesures de profondeur pour couvrir toutes les surfaces visibles. Plusieurs algorithmes permettent de générer l'esquisse: la stéréoscopie, l'ombrage, la couleur, la texture, le mouvement et la position des sources d'éclairage sont les principaux. Nous nous attardons, dans cette étude, aux deux premiers, soient la stéréo et l'analyse de l'ombrage.

4.1 Stéréoscopie

La stéréoscopie consiste à mesurer la profondeur de points caractéristiques d'une scène en calculant la disparité des mêmes points caractéristiques sur deux ou trois images de la scène observée de points de vue légèrement différents. La profondeur s'obtient alors par triangulation. La nature des points caractéristiques choisis détermine deux grandes catégories de stéréoscopie: la stéréo par *appariement de caractéristiques* utilise les arêtes ou des segments connectés pour calculer la disparité en ces points discrets tandis que la stéréo par *corrélacion de surface* identifie sur les deux images des surfaces similaires et calcule leur disparité. De bonnes revues de littérature qui couvrent la recherche dans les années 70 sont présentées dans [3, 40] tandis que [41] présente une mise à jour complète pour les années 80.

La théorie de base de la stéréoscopie par appariement de caractéristiques remonte à Marr et Poggio [42]. Une réalisation neuronique hiérarchique a été proposée par [6] qui utilisent une représentation multirésolution afin d'atténuer les faux appariements et par [43] qui utilise également une représentation multirésolution et une machine de Boltzmann pour l'apprentissage des contraintes d'appariement par recuit simulé. Auparavant, Marr et Poggio [4] avaient proposé un réseau neuronique d'appariement des caractéristiques par relaxation de contraintes. Les contraintes de continuité et d'unicité étaient pré-programmées dans le réseau. Drumheller et Poggio [44] proposent par la suite une réalisation du réseau de relaxation sur une machine parallèle, la *Connection Machine*. Bokil et Khotanzad [45] proposent une extension du modèle par relaxation en permettant des images d'entrée à niveaux de gris et en apprenant les contraintes d'appariement, au lieu de les fixer à l'avance. Nasrabadi et Choo [46] proposent pour leur part l'utilisation d'un réseau de Hopfield comme réseau neuronique de relaxation.

4.2 Ombrage

Lorsque nous regardons une photographie, la structure tridimensionnelle des éléments de la scène photographiée ressort assez distinctement, nous permettant par exemple de «voir» la forme dans l'espace d'une sphère. L'analyse de l'ombrage sur une image donne des indices sur la structure tridimensionnelle de la scène projetée sur l'image. Le recouvrement de la forme 3D à partir de l'ombrage, c'est-à-dire à partir des variations faibles du niveau de gris sur l'image, a débuté avec les travaux de Horn et Pentland au milieu des années 80. Nous retrouvons deux approches pour résoudre le problème de calcul des variations de profondeur à partir des variations d'illuminance sur l'image: une approche globale, qui cherche des solutions le long de parcours caractéristiques [47], et l'approche locale [48] qui analyse les variations locales d'illuminance sur l'image.

L'analyse locale de l'ombrage présuppose que la direction de l'éclairage et l'albédo de la surface de l'objet sont constants et connus au préalable. Breton *et al* [49] trouvent simultanément une solution, par une analyse locale, aux problèmes de direction de sources lumineuses, d'orientation des surfaces extérieures et de variation d'albédo au moyen de réseaux fibrés qui identifient, pour chaque pixel de l'image, toutes les orientations possibles de surface. Un réseau neuronique résout, par satisfaction de contraintes, la meilleure combinaison de ces orientations possibles.

Hayakawa *et al* [50] proposent un réseau neuronique à deux couches pour reconstituer la profondeur à partir d'une analyse locale de l'ombrage. Les connexions directes approximent l'optique inverse tandis que les connexions inverses calculent l'équation de formation des images.

Lehky et Sejnowski [51, 52] utilisent un réseau à rétropropagation du gradient d'erreurs. L'apprentissage se déroule en présentant au réseau des images de surfaces géométriques simples dont les courbatures de surface sont calculées et fournies comme sorties de supervision. Une fois l'apprentissage terminé, l'analyse des champs récepteurs obtenus montre des similarités avec les champs récepteurs du cortex visuel associés aux détecteurs de coins et de barres, ce qui semble montrer que l'étude seule des champs récepteurs n'est pas suffisante: il faut tenir compte du «champ projectif», c'est-à-dire l'ensemble des neurones connectés en amont.

Lepage et Poussart [6] proposent d'utiliser les valeurs d'illuminance et de son intégrale pour approximer les formes tridimensionnelles. Pour résoudre

les ambiguïtés, les auteurs proposent de combiner les résultats de la stéréoscopie, qui fournissent les données de profondeur aux positions d'arêtes, à ceux de l'ombrage dans un réseau neuronique de relaxation avec comme conditions initiales les valeurs de profondeur obtenues de la stéréo

5. Description et classification 3D

Le niveau de description 3D décrit symboliquement la forme des objets visibles et leur organisation spatiale dans un système de coordonnées centré sur l'objet observé. La représentation des objets est hiérarchique et modulaire avec des primitives de volume et de surface. Cette forme de représentation est souvent réalisée en comparant le résultat de la segmentation de l'esquisse 2,5D à des modèles emmagasinés en mémoire. Cette forme de représentation se veut idéalement indépendante du point de vue et permet ainsi de reconnaître les objets composant la scène observée, ce qui constitue en somme le but ultime du système de vision. Les recherches sont nombreuses et très actives dans ce domaine. La segmentation consiste à identifier les différentes composantes des objets, de leur assigner une description indépendante du point de vue et de coder cette information afin d'effectuer la comparaison avec une banque de modèles ou de primitives comme les *geons* [53]. Hummel et Biederman [54, 55] proposent un réseau neuronique à liens dynamiques pour pouvoir représenter les *geons* constituant la scène observée et leur interrelation spatiales au moyen de connexions à l'activation périodique pour décomposer les relations spatiales en séquences temporelles.

La description 3D obtenue de l'algorithme de segmentation doit être comparée avec des modèles emmagasinés afin d'identifier les objets composant la scène observée. En ce qui concerne la réalisation par réseaux neuroniques, le modèle de réseau neuronique le plus utilisé est le modèle résonant de la famille ART [17, 56]. Il s'agit d'un réseau à deux couches principales connectées dans les deux directions. L'apprentissage est non-supervisé. La comparaison avec les modèles appris s'effectue par résonance, les modèles les plus similaires étant essayé en premier. Un nouveau modèle est emmagasiné en cas d'échec de la comparaison, permettant ainsi au réseau neuronique de s'adapter à l'environnement.

6. Conclusion

Nous avons présenté une revue de la littérature des principaux modèles de

réseaux neuroniques qui peuvent être utilisés dans un système de vision artificielle. La revue permet de constater que la recherche est très active dans ce domaine prometteur. Nous nous sommes limités aux systèmes de vision tridimensionnels, avec une emphase pour les bas niveaux de représentation et leurs algorithmes associés. Une revue de littérature pour les systèmes de reconnaissance de forme (écriture, chiffre, etc.) est en cours de rédaction.

7. Références

- [1] D. Marr, *Vision - A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: Freeman, 1982.
- [2] M. S. Landy and J. A. Movshon, Eds., *Computational Models of Visual Processing*. Cambridge, Massachusetts: The MIT Press, 1991.
- [3] M. Brady, "Computational Approaches to Image Understanding," *ACM Computing Surveys*, vol. 14, no. 1, 1982, pp. 3-71.
- [4] D. Marr, "Early Processing of Visual Information," *Philosophical Transactions of the Royal Society of London B*, vol. 245, 1976, pp. 483-519.
- [5] A. L. Yuille and S. Ullman, "Computational Theories of Low-Level Vision," in *Visual Cognition and Action*, vol. 2, D. N. Osherson, S. M. Kosslyn, and J. M. Hollerbach, Eds. Cambridge, Massachusetts: The MIT Press, 1990, pp. 5-39.
- [6] R. Lepage and D. Poussart, "A multiresolution neural network for the fusion of stereo and shape from shading in machine vision," in *Proc. SPIE Intelligent Robots and Computer Vision XI: Biological, Neural Net, and 3-D Methods*, Boston, Massachusetts, 18-20 November, 1992, pp. 238-249.
- [7] N. Weisstein and E. Wong, "Figure-Ground Organization Affects the Early Visual Processing of Information," in *Vision, Brain and Cooperative Computation*, M. A. Arbib and A. R. Hanson, Eds. Cambridge, Mass.: The MIT Press, 1987, pp. 209-230.
- [8] T. Lindeberg and J.-O. Eklundh, "Scale-space primal sketch: construction and experiments," *Image and Vision Computing*, vol. 10, no. 1, 1992, pp. 3-18.
- [9] T. Lindeberg, *Scale-Space Theory in Computer Vision*. Boston: Kluwer Academic Publishers, 1994.
- [10] S. Grossberg, "The quantized geometry of visual space: The coherent computation of depth, form, and lightness," *The Behavioral and Brain Sciences*, vol. 6, 1983, pp. 625-692.
- [11] S. Grossberg and E. Mingolla, "Neural Dynamics of Surface Perception: Boundary Webs, Illuminants, and Shape-from-Shading," *Computer Vision, Graphics, and Image Processing*, vol. 37, 1987, pp. 116-165.
- [12] J. Canny, "A Computational Approach to Edge Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, 1986, pp. 679-698.
- [13] T. Grossberg, E. Mingolla, and D. Todorovic, "A Neural Network Architecture for Preattentive Vision," *IEEE Transactions on Biomedical Engineering*, vol. 36, no. 1, 1989, pp. 65-84.
- [14] G. A. Carpenter, S. Grossberg, and C. Mehanian, "Invariant Recognition of Cluttered Scenes by a Self-Organizing ART Architecture: CORT-X Boundary Segmentation," *Neural Networks*, vol. 2, 1989, pp. 169-181.
- [15] P. Gaussier and J.-P. Cocquerez, "Neural networks for complex scene recognition: simulation of a visual system with several cortical areas," in *Proc. IJCNN '92 Baltimore - International Joint Conference on Neural Networks*, Baltimore, Maryland, June 7-11, 1992, pp. III-233 - III-259.

- [16] S. Grossberg and E. Mingolla, "Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations," *Perception and Psychophysics*, vol. 38, no. 2, 1985, pp. 141-171.
- [17] S. Grossberg and L. Wyse, "Invariant Recognition of Cluttered Scenes by a Self-Organizing ART Architecture: Figure-Ground Separation," in *Proc. IEEE INNS International Joint Conference on Neural Networks - Seattle*, Seattle, Washington, July 8-12, 1991, pp. I-633 - I-638.
- [18] J. G. Daugman, "Complete Discrete 2-D Gabor Transforms by Neural Networks for Image Analysis and Compression," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 36, no. 7, 1988, pp. 1169-1179.
- [19] S. Grossberg and D. Todorovic, "Neural dynamics of 1-D and 2-D brightness perception: A unified model of classical and recent phenomena," *Perception and Psychophysics*, vol. 43, 1988, pp. 241-277.
- [20] R. Lepage and D. Crevier, "Multiresolution neural network for the extraction of the primal sketch," in *Proc. SPIE Intelligent Robots and Computer Vision XIII: Algorithms and Computer Vision*, Boston, Massachusetts, 31 October - 2 November, 1994, pp. 375-384.
- [21] M. M. Fleck, "Some Defects in Finite-Difference Edge Finders," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 3, 1992, pp. 337-345.
- [22] D. McMichael, "Fusing Multiple Images and Extracting Features for Visual Inspection," Control Systems Centre, UMIST, Manchester, UK, Control Systems Centre Report 762, 1992.
- [23] E. C. Hildreth, "The Detection of Intensity Changes by Computer and Biological Vision Systems," *Computer Vision, Graphics, and Image Processing*, vol. 22, 1983, pp. 1-27.
- [24] D. Marr and E. Hildreth, "Theory of edge detection," *Proceedings of the Royal Society of London, Series B*, vol. 207, 1980, pp. 187-217.
- [25] L. Moura and F. C. M. Martins, "Edge Detection Through Cooperation and Competition," in *Proc. International Joint Conference on Neural Networks*, Singapore, 1991, pp. 2588-2593.
- [26] P. Bhatia, V. Srinivasan, and S. H. Ong, "Single-layer edge detector with competitive unsupervised learning," in *Proc. International Joint Conference on Neural Networks*, Singapore, 1991, pp. 634-639.
- [27] K. Etemad and R. Chellappa, "A Neural Network Based Edge Detector," in *Proc. IEEE International Conference on Neural Networks*, San Francisco, CA, March 28 - April 1, 1993, pp. 132-137.
- [28] M. Mkaouar and R. Lepage, "Le Perceptron multi-couches pour la détection des arêtes d'une image," in *Proc. Vision Interface '96*, Toronto (Ontario) Canada, 21-24 mai, 1996, pp. 41-48.
- [29] A. P. Witkin, "Scale space filtering: a new approach to multi-scale description," in *Image Understanding 1984*, S. Ullman and W. Richards, Eds. Norwood, NJ: Ablex Publishing Corporation, 1984.
- [30] A. L. Yuille and T. A. Poggio, "Scaling Theorems for Zero Crossings," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 1, 1986, pp. 15-25.
- [31] F. Bergholm, "Edge Focusing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-9, no. 6, 1987, pp. 726-741.
- [32] P. J. Burt, "Smart Sensing within a Pyramid Vision Machine," *Proceedings of the IEEE*, vol. 76, no. 8, 1988, pp. 1006-1015.
- [33] J.-M. Jolion and A. Rosenfeld, *A Pyramid Framework for Early Vision*. Boston:

- Kluwer Academic Publishers, 1994.
- [34] T. Lindeberg, "Scale-Space for Discrete Signals," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 3, 1990, pp. 234-254.
 - [35] P. Perona, "Deformable Kernels for Early Vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 5, 1995, pp. 488-499.
 - [36] J. Babaud, A. P. Witkin, M. Baudin, and R. O. Duda, "Uniqueness of the Gaussian Kernel for Scale-Space Filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 1, 1986, pp. 26-33.
 - [37] L. Wu and Z. Xie, "Scaling Theorems for Zero-Crossings," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, 1990, pp. 46-54.
 - [38] V. Anh, J. Y. Shi, and H. T. Tsui, "Scaling Theorems for Zero Crossings of Bandlimited Signals," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 3, 1996, pp. 309-320.
 - [39] M. Mkaouar and R. Lepage, "Extraction des caractéristiques d'une image par analyse à multiple résolutions spatiales," in *Proc. Congrès canadien en génie électrique et informatique*, Montréal, Québec, du 5 au 8 septembre, 1995, pp. 1176-1179.
 - [40] S. T. Barnard and M. A. Fischler, "Computational Stereo," *ACM Computing Surveys*, vol. 14, no. 4, 1982, pp. 553-572.
 - [41] U. R. Dhond and J. K. Aggarwal, "Structure from Stereo - A Review," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 19, no. 6, 1989, pp. 1489-1510.
 - [42] D. Marr and T. Poggio, "A computational theory of human stereo vision," *Proceedings of the Royal Society of London, Series B*, vol. 204, 1979, pp. 301-328.
 - [43] S. T. Barnard, "Stochastic Stereo Matching over Scale," *International Journal of Computer Vision*, vol. 3, 1989, pp. 17-32.
 - [44] M. Drumheller and T. Poggio, "On parallel stereo," in *Proc. IEEE International Conference on Robotics and Automation*, San Francisco, California, April 7-10, 1986, pp. 1439-1448.
 - [45] A. Bokil and A. Khotanzad, "A Constraint Learning Feedback Dynamic Model for Stereopsis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 11, 1995, pp. 1095-1100.
 - [46] N. M. Nasrabadi and C. Y. Choo, "Hopfield Network for Stereo Vision Correspondence," *IEEE Transactions on Neural Networks*, vol. 3, no. 1, 1992, pp. 5-13.
 - [47] B. K. P. Horn, "Shape-from-Shading Bibliography," in *Shape from Shading*, B. K. P. Horn and M. J. Brooks, Eds. Cambridge, Mass.: The MIT Press, 1989, pp. 533-569.
 - [48] A. P. Pentland, "Local Shading Analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, no. 2, 1984, pp. 170-187.
 - [49] P. Breton, L. A. Iverson, M. S. Langer, and S. W. Zucker, "A New Approach to Shape from Shading," in *Neural networks for vision and image processing*, G. A. Carpenter and S. Grossberg, Eds. Cambridge, Massachusetts: The MIT Press, 1992, pp. 111-131.
 - [50] H. Hayakawa, T. Inui, and M. Kawato, "Computational Theory and Neural Network Model of Perceiving Shape from Shading in Monocular Depth Perception," in *Proc. IEEE INNS International Joint Conference on Neural Networks - Seattle*, Seattle, Washington, July 8-12, 1991, pp. I-649 - I-654.
 - [51] S. R. Lehky and T. J. Sejnowski, "Network model of shape-from-shading: neural function arises from both receptive and projective fields," *Nature*, vol. 333, 1988, pp. 452-454.
 - [52] S. R. Lehky and T. J. Sejnowski, "Simplifying Network Models of Binocular Rivalry and Shape-from-Shading," in *Methods in Neuronal Modeling - From Synapses to Networks, Computational Neuroscience*, C. Koch and I. Segev, Eds. Cambridge,

Massachusetts: The MIT Press, 1989, pp. 361-396.

- [53] I. Biederman, "Higher-Level Vision," in *Visual Cognition and Action*, vol. 2, D. N. Osherson, S. M. Kosslyn, and J. M. Hollerbach, Eds. Cambridge, Massachusetts: The MIT Press, 1990, pp. 41-72.
- [54] I. Biederman, J. E. Hummel, P. C. Gerhardstein, and E. E. Cooper, "From image edges to geons to viewpoint invariant object models: a neural net implementation," in *Proc. SPIE Applications of Artificial Intelligence X: Machine Vision and Robotics*, Orlando, Florida, 22-24 april, 1992, pp. 570-578.
- [55] J. E. Hummel and I. Biederman, "Dynamic Binding in a Neural Network for Shape Recognition," *Psychological Review*, vol. 99, no. 3, 1992, pp. 480-517.
- [56] G. A. Carpenter and W. D. Ross, "ART-EMAP: A Neural Network Architecture for Object Recognition by Evidence Accumulation," *IEEE Transactions on Neural Networks*, vol. 6, no. 4, 1995, pp. 805-828.

Reconnaissance des linéaments à partir d'images acquises par satellite. Discours sur la méthode

Desjardins, R., Iris, S. et St-Onge, B. A.

Département de géographie, Université du Québec à Montréal, Montréal, Qc, H3C 3P8

Roy, D.W.

Département des sciences appliquées, Université du Québec à Chicoutimi, Chicoutimi, Qc, G7H 2B1

Résumé

Cette recherche vise à présenter un état de la question sur les méthodes qui permettent la reconnaissance des linéaments de type géologique et géomorphologique et leur organisation à l'échelle régionale. Elles se divisent en deux catégories: analogique et numérique. Nous présentons d'une façon plus détaillée deux approches à caractère numérique: une première basée sur les filtres directionnels et une seconde qui fait appel au concept du réseau de neurones. Enfin, nous présentons le site expérimental où nous les mettrons à contribution et nous indiquons l'orientation de toute la recherche appliquée dans le secteur étudié.

1. Introduction

La notion de linéament fait partie de la littérature géologique moderne depuis le début du siècle. À ce moment, on décrivait les linéaments comme étant des crêtes à des limites de reliefs ou de partage des eaux [1]. Plus récemment, ils ont été définis comme étant des surfaces simples ou complexes dont les éléments sont alignés d'une manière rectiligne ou curviligne [2]. Elles traduisent probablement des phénomènes qui se produisent sous la surface terrestre. Sur des images aériennes ou satellitaires, les linéaments sont mis en évidence par la physiographie particulière du paysage en matière de profil topographique, hydrographique ainsi que différentes limites de la végétation [3]. De plus, la visualisation des linéaments est associée au changement brusque et linéaire des niveaux de gris sur une image. Ceci relève plus d'une description faite à partir d'une photographie ou d'une image. Les linéaments cachent soient des lignes géologiques (cassures ou contacts), des lignes topographiques, hydrographiques et anthropiques (routes, corridors de lignes hydroélectriques, effets d'ombrage à la limite d'une coupe à blanc) [4].

Reconnaître et bien saisir l'organisation des linéaments est un besoin bien réel. En effet, ce genre d'étude est nécessaire dans trois champs d'application particuliers:

- Développement d'indices minéralisés aidant à la détection de nouveaux sites miniers.
- Détection des zones fissurées dans une perspective de reconnaissance du réseau de drainage pour l'eau ou des matières liquides potentiellement polluantes.
- Une meilleure connaissance des risques de tremblements de terre à l'échelle locale.

2. Problématique générale

Un des problèmes dans les méthodes utilisées pour la reconnaissance des linéaments vient du fait que le travail sur le terrain et l'utilisation des photographies aériennes présente des contraintes importantes. En effet, il manque une vision d'ensemble du phénomène à étudier. Souvent, la photo ne présente qu'une partie du territoire à couvrir et les reconnaissances sur le terrain sont limitées par le manque d'accessibilité à certains secteurs, la rareté des affleurements rocheux et les conditions d'ordre climatique parfois difficiles.

3. Objectif de la recherche

L'objectif de cette étude est d'abord de présenter et d'examiner différentes méthodes d'extraction des linéaments contenus dans les images provenant des satellites de télédétection. Nous concentrerons nos efforts sur deux méthodes: le traitement numérique d'images par l'intermédiaire des filtres directionnels et l'approche du problème de reconnaissance par réseaux de neurones.

4. État de la question

4.1 Les méthodes analogiques

L'utilisation de la photographie aérienne, jumelée avec les données acquises sur le terrain constituent les deux items d'une approche dite conventionnelle qui a longtemps constitué la base de travail pour la reconnaissance des linéaments. En effet, dès le milieu des années trente, les compagnies pétrolières ont utilisé d'une façon intensive les données aériennes

à des fins cartographiques. Ainsi, la préparation de cartes systématiques contenant les éléments essentiels en géologie sont réalisables autant pour les régions accessibles que non accessibles (par l'intermédiaire du terrain).

En fait, les possibilités de vision stéréoscopique des photographies aériennes et les effets d'ombrage provoqués par la position du soleil auront fait en sorte qu'il était possible de relever directement sur un calque les linéaments. Plusieurs méthodes visant à l'étude de la fracturation (densité du réseau) ont été développées. Elles sont décrites par différents auteurs [5], [6]. Plus récemment, des chercheurs ont découvert des indicateurs de linéaments sur les photographies aériennes (alignements de végétation naturelle, changements brusques de pente, profil linéaire du réseau hydrographique) [7]. Ils s'ajoutent à ceux qui sont identifiables sur le terrain (brèches, dykes, anomalies stratigraphiques, ruptures de pente). Ces types d'approche ne sont pas nécessairement à l'abri des interprétations subjectives.

Une autre approche à caractère analogique plus récente implique une difficulté certaine d'interprétation "objective". En effet, certains travaux sont à cet égard très éloquents. Bien que les images étaient d'origine numérique et que différents traitements avaient été réalisés afin d'accentuer le contraste visuel pour pouvoir extraire plus facilement l'information linéaire à partir d'un produit papier, les linéaments provenant de trois sources différentes (lire capteurs et modèle numérique d'altitude) étaient mal identifiés. En effet, la comparaison des résultats aura démontré que l'identification varie beaucoup d'un document à l'autre (images satellite, modèle numérique d'altitude)... et d'un interprète à l'autre [8].

4.2 Les méthodes numériques

Cette approche regroupe des méthodes où les analyses mathématiques et statistiques sont effectuées sur les matrices de données. Elles permettent de mettre clairement en évidence l'élément linéaire recherché par rapport à son environnement. En effet, il en résulte que ces types de traitement engendrent une banalisation de la "toile de fond" sur laquelle reposent les linéaments. Afin d'y arriver, différents algorithmes de reconnaissance et de transformation ont été développés et/ou adaptés au traitement numérique d'images.

La représentation des éléments linéaires contenus dans les images satellite peut être à l'occasion fort complexe puisque pour chaque élément, les caractéristiques spectrales et spatiales peuvent varier. Différentes méthodes

ont été développées pour contrer ces difficultés. Des algorithmes de reconnaissance basés sur le rehaussement d'arêtes ont été élaborés [9]. Jusqu'à maintenant, les plus importants développements dans ce domaine ont été réalisés par Wang [10]. Il a mis au point plusieurs algorithmes et les a regroupés dans un système semi-automatisé de détection et d'analyse appelé LINDA (Linear feature Network Detection and Analysis). Ce système contient divers modules dont les deux principaux sont le "linear detection-GDPA" (Gradient Direction Profile Analysis) et le "Edge detection-EFGS" (Edge Following and Graph Searching). Le GDPA fut à l'origine développé pour extraire le réseau de drainage des images Landsat. En ce qui concerne le EFGS, il permet d'extraire l'amplitude et la direction des éléments linéaires.

En 1962, Hough proposa une méthode de détection des éléments linéaires. La "transformée de Hough" consiste à considérer chaque pixel d'arête de l'image filtrée et seuillée comme faisant partie d'une infinité de segments de droite traversant ce point. Les équations des droites, exprimées en coordonnées polaires, sont ensuite regroupées par sous-ensembles angulaires (24 arcs de 25° par exemple). Lorsque plusieurs points colinéaires formant des linéaments incomplets apparaissent sur l'image, certains maxima de l'histogramme des équations de droites ressortent clairement et indiquent la présence d'éléments linéaires. La transformée seuillée, une fois inversée, permet de les tracer sur le plan de l'image.

Par ailleurs, la transformée de FOURIER permet d'exprimer la composition de l'image en termes de fréquences spatiales [11]. Elle génère un diagramme de fréquences cartésiennes qui permet de mettre en évidence les éléments structuraux dominants ainsi que leur orientation, tout en isolant le bruit, ce dernier étant par la suite facile à éliminer.

Les deux prochaines méthodes sont celles sur lesquelles nous nous attarderons. Il s'agit de la détection des linéaments par l'intermédiaire des filtres directionnels et des réseaux de neurones appliqués sur des images numériques provenant des satellites.

5. Les filtres directionnels

La méthode de détection des linéaments basée sur l'utilisation des filtres directionnels (voir figure 1) se réalise en trois étapes:

- Les prétraitements
- La détection et l'extraction des linéaments
- La cartographie de l'information

5.1 Étape 1: Les prétraitements

Pour qu'une ou plusieurs images soient utilisables et que les informations extraites soient comparables ou complémentaires, il faut d'abord s'assurer que la représentation numérique des données soit au même niveau (8 bits ou 16 bits).

Dans la majorité des cas, le faible étalement des valeurs spectrales sur l'histogramme de fréquence fait en sorte que nous ne pouvons pas toujours distinguer clairement les différents objets qui composent l'image d'une façon claire. Nous rehaussons les contrastes à partir de la technique d'étirement linéaire.

Par la suite, il devient impératif d'effectuer une correction géométrique sur les différentes images utilisées. Cette correction permet d'orienter l'image en fonction du nord géographique. Afin d'y parvenir, nous avons opté pour l'approche photogrammétrique d'abord développée pour l'image SPOT[12]. Elle a été adaptée par la suite pour d'autres types d'images telles que celles provenant des satellites Landsat et ERS-1 RAS [13], [14]. Cette approche permet la correction géométrique et la rectification avec un modèle numérique d'altitude (MNA) d'images de sources diverses. Il engendre une représentation mathématique de la réalité physique lors de la formation de l'image. Elle corrige les images en fonction des trois dimensions (coordonnées de la grille UTM et altitude) [15].

Enfin, il faut obtenir une image de base qui présente déjà un contraste entre les éléments linéaires et la toile de fond. Pour ce faire, nous devons d'abord éliminer les bruits ou parasites qui subsistent sur l'image. à l'aide d'un filtre médian. Cette technique de filtrage permet d'atténuer fortement le bruit tout en entraînant une faible perte de résolution spatiale sur les éléments linéaires [16].

5.2 Étape II: Détection et extraction des linéaments

Les filtres directionnels nous permettent donc de mieux percevoir les linéaments car ils provoquent un effet d'ombrage porté sur l'image comme si

le territoire analysé était éclairé par une lumière rasante. Ils offrent comme avantage de rehausser les linéaments qui sont éloignés de la normale de la source d'éclairement que nous avons sur les images [17]. Ce type de rehaussement des linéaments s'effectue en utilisant un opérateur de SOBEL. Afin de détecter le maximum de linéaments en fonction de leur direction, nous devons opter pour deux filtres "diagonaux" et deux filtres "orthogonaux". Ceci est l'équivalent d'une source d'éclairement à 45, 135, 0 et 90° de direction. Cette procédure a l'avantage de détecter aussi bien les linéaments orientés verticalement, en oblique ou horizontalement par rapport au gradient du filtre. De plus, il aide à minimiser le bruit. Sa fenêtre a une dimension variable en fonction des besoins. Seuls les linéaments dont la dimension est supérieure à la moitié de la fenêtre du filtre peuvent être détectés [18]. À titre d'exemple, un filtre directionnel de type SOBEL (ayant une fenêtre de dimension 5 X 5) qui est utilisé sur une image ayant des pixels de 30 m pourra détecter des linéaments supérieurs à 75 m. En deçà de cette dimension, ils se confondent avec le bruit de fond.

Réaliser des filtres dans les quatre directions mentionnées précédemment engendre une redondance du fait que deux filtres voisins (ex: 0° et 45°) vont détecter parfois les mêmes linéaments. Il devient alors inutile, voire même hasardeux de laisser l'information produite telle quelle. Afin d'éviter des erreurs telles que comptabiliser le même linéament deux fois, il faut éliminer cette redondance. Dans cette perspective, l'analyse en composantes principales (basée sur la transformée de Karhuen-Loeve) est l'opération que nous avons retenue. Elle engendre une décorrélation entre les nouvelles bases de données (les composantes). Dans la plupart des cas, les deux premières composantes expliquent plus de 95% de la variance entre les bases d'information d'origine. La perte d'information non redondante est minime (en général de l'ordre de 1 à 2%). Cette opération a aussi comme avantage de banaliser la toile de fond de l'image en y attribuant une valeur uniforme. On a démontré que les premiers et derniers 10% de l'histogramme de fréquence de cette nouvelle image représentent des limites crédibles pour la création d'images binaires [19]. Elle est par la suite superposable à une image d'origine afin de faciliter l'interprétation et d'éliminer les linéaments non géologiques. Une carte est par la suite créée sous forme de document papier.

5.3 Étape III: La cartographie

Une visualisation spatiale des linéaments fait partie des outils nécessaires à une meilleure compréhension de l'organisation structurale d'un territoire. Pour ce faire, nous avons besoin d'une analyse du secteur étudié à partir du document papier qui a été fait à la fin de l'étape précédente. Il s'agit d'une superposition des linéaments sur une image satellite qui présente l'occupation du sol. À titre indicatif, la bande TM03 du satellite Landsat 5 ou XS2 du satellite SPOT constituent d'excellentes images d'origine pour bien identifier cette occupation du sol.

Lorsque la cartographie manuelle des linéaments géologiques est terminée, il devient utile de numériser le résultat à partir d'une table numérisante et de logiciels appropriés. Il reste à effectuer un choix d'échelle pour le produit final. Cela dépend des besoins de l'utilisateur. Ce choix devra être fait dans le but de bien doser le nombre de linéaments (à partir de leur longueur) en fonction la réalité du terrain représentée sur la carte. Une carte surchargée d'information masque souvent le message que l'on souhaite transmettre.

6. Les réseaux de neurones

Les méthodes classiques de classification engendrent généralement une marge d'erreur et nécessitent toujours une assistance humaine lors de la phase d'interprétation. Les réseaux de neurones peuvent aider à minimiser notre intervention en offrant l'avantage de reconnaître rapidement, une fois entraînés correctement, des formes d'une image à l'autre. L'intérêt croissant accordé à cette méthode nous a incité à l'expérimenter pour la détection et l'extraction de linéaments.

Les réseaux de neurones, aussi connus sous le nom de modèles connexionnistes, sont des systèmes informatiques imitant les capacités d'apprentissage et de raisonnement du cerveau humain par l'utilisation d'un grand nombre de neurones artificiels interconnectés entre eux [20]. Ils sont composés d'unités de traitement (noeuds) organisés sous forme de couche: une d'entrée, une de sortie et, entre elles, une ou plusieurs couches cachées. Le nombre de noeuds d'entrées et de noeuds cachés est défini en fonction de la complexité du problème à résoudre. Chacun des noeuds possède deux fonctions: une fonction d'addition et une fonction de seuillage [21]. La figure 2 présente les étapes de cette approche.

6.1 Étape I: Les prétraitements

Comme avec la méthode par filtre directionnel, pour que les résultats puissent être comparés, les données numériques devront être au même niveau (8 ou 16 bits). Par la suite, nous devons effectuer une étape de rehaussement et de correction géométrique afin que les images présentent un contraste visuel de bonne qualité entre les différents objets qui les composent. Finalement, dans le but d'éliminer les pixels isolés qui pourrait nuire à notre traitement, nous utilisons aussi le filtre médian.

6.2 Étape II: Phase de traitement

En premier lieu, nous devons délimiter des sites où les linéaments se distinguent facilement par rapport aux éléments environnants. Une taille variant entre 30 et 100 pixels par site est favorable pour la phase d'apprentissage du réseau [22].

Par la suite, nous devons établir la topologie du réseau en définissant le nombre de noeuds d'entrées, de sorties et de couches cachées. Ce choix peut être établi après un certain nombre d'expérimentations. Ces dernières sont réalisées en modifiant le nombre de noeuds des diverses couches à chacun des essais. Le modèle final correspond à celui où la détection aura été optimale. Pour y arriver nous devons réaliser, avec les différents modèles, un entraînement supervisé par rétropropagation. Cette méthode d'apprentissage permet au réseau de vérifier si la différence entre les résultats obtenus et ceux désirés est supérieure à un certain seuil. Si tel est le cas, les données de sorties sont réintégrées dans le réseau en modifiant les poids des connections entre les couches et ce, jusqu'à ce que la différence de résultats soit inférieure au seuil établi.

6.3 Étape III: Cartographie

Afin de mieux visualiser et comparer les résultats obtenus par le réseau de neurones, nous cartographions les linéaments détectés. Cette étape est réalisée sous format numérique avec sortie papier, à l'aide d'un logiciel de cartographie automatisée.

7. Conclusion

Nous proposons de réaliser l'appréhension des linéaments selon deux approches générales. Celle dite traditionnelle qui met en valeur les données provenant du terrain et des photographies aériennes. Actuellement des efforts

sont déployés afin "d'automatiser" la détection de ces phénomènes linéaires par l'intermédiaire de méthodes basées sur le traitement numérique (mathématique et statistique) des informations.

Dans cette recherche, nous mettons l'emphasis sur deux méthodes à caractère numérique. La première met en valeur l'utilisation de filtres directionnels, jumelés à divers phases menant à l'accentuation de contrastes entre les objets à détecter et leur environnement dans une perspective de cartographie. Nous présentons par la suite la méthode des réseaux de neurones dont le but est de minimiser l'intervention de type qualitatif dans la détection de ces linéaments. Cet état de la question sur les méthodes sera suivi dans les prochains mois par la phase expérimentale qui sera réalisée à partir d'images acquises par satellite, notamment celles du satellite canadien RADARSAT I. Il y aura aussi dans cette étape une validation des résultats obtenus à partir d'images du satellite ERS-1 RAS [23]

8. Bibliographie

- [1] Hobbs, W.H. "Lineaments of the Atlantic border region" *Geology Society of America Bulletin*, vol. 22, 1904, pp. 123-176.
- [2] O'Leary, D.W., Friedman, J.D. et Pohn, H.A. "Lineament, linear and lineation: Some proposed new standards for old terms" *Geological Society of America Bulletin*, vol. 87, 1976, pp. 1463 - 1469.
- [3] Scanvic, J.Y. "Utilisation de la télédétection dans les sciences de la terre" *Bureau de recherches géologiques et minières*, Paris, 1984, pp. 159.
- [4] Roy, D.W. et Boily, J. "Délimitation de domaines structuraux par classification non-supervisée de cartes de linéaments" *Actes du 16^e Symposium canadien de télédétection 8^e Congrès de l'Association québécoise de télédétection, AQT/CRSS Ed.*, 1993, pp. 301 - 305.
- [5] Lattman, L.H. "Technique for mapping geologic fracture traces of lineaments on aerial photographs" *Photogrammetric Engineering*, vol. 24, no 4, 1958, pp. 568 - 576.
- [6] Mollard, J.D. "A study of aerial mosaics in southern Saskatchewan and Manitoba" *Oil in Canada*, 1957, pp. 11.
- [7] Norman, J.W. "Photogeological fracture trace analysis as a subsurface exploration technique" *Transaction of the American Institute of Mining Engineers*, no 86, 1976, pp. 52 - 62.
- [8] Mussakowski, R. et Trowell, N.F. "Comparaison of remote sensing and digital elevation data for geological lineament interpretation, Thunder Bay area, Ontario" *Precambrian*, vol. 14, 1992, pp. 75 - 80.
- [9] Joshi, A.K. "Automatic detection of lineament from landsat data", *Proceeding Igarss. '89/Twelfth Canadian Symposium on Remote Sensing (Vancouver B.C.)*, vol. 1, 1989, pp. 85-88.

- [10] Wang, J "LINDA - A system for automated linear feature detection and analysis" *Canadian Journal of Remote Sensing*, vol. 19, no 1, 1993, pp. 9 - 20.
- [11] Deslandes, S. et Gwyn, Q.H.J. "Évaluation de Spot et de Seasat pour la cartographie des linéaments: comparaison basée sur l'analyse du spectre de FOURIER" *Journal canadien de télédétection*, vol. 17, no 2, 1991, pp. 98 - 112.
- [12] Toutin, Th. "Étude mathématique pour la rectification d'images Spot" *Comptes rendu du XVII^e Congrès international des géomètres*, tenu à Toronto, Canada, du 1^{er} au 11 juin 1986, pp. 379 - 395.
- [13] [Toutin, Th. "Cartographie à partir d'image radar" *Journal canadien de télédétection*, vol. 20, no 2, 1994 pp. 171 - 179.
- [14] Toutin, Th. et Carbonneau, Y. "La création d'ortho-image avec un MNE: description d'un nouveau système" *Journal canadien de télédétection*, vol. 18, no 3, 1992, pp. 136 - 141.
- [15] De Sève, D., Toutin, Th. et Desjardins, R., "Évaluation de deux méthodes de correction géométrique d'images Landsat TM et ERS-1 RAS dans une étude de linéaments géologiques" *International Journal of Remote Sensing*, vol. 17, no 1, 1996, pp. 131 - 142.
- [16] Marion, A. "Introduction aux techniques de traitement d'image" *Eyrolles, Ed.*, Paris, 1987, pp. 278.
- [17] [Drury, S.A. "Remote sensing of geological structure in temperate agricultural terrain." *Geological Magazine*, vol. 123, no 2, 1986, pp. 113-121.
- [18] Colwell, R. (Ed), "Manual of remote sensing" Vol. I et II, *American Society of Photogrammetry and Remote Sensing*, Falls Church, Va, 1983, pp. 2133.
- [19] Moore, G. "Digital processing of Landsat-TM image for lineament occurrence and spatial frequency in sedimentary rocks" *Geological Survey Report DOE/OR21552-1*, EROS Data Centre, 1986, pp. 19.
- [20] [Sui, D.Z., "Recent Applications of Neural Networks for Spatial Data Handling" *Canadian Journal of Remote Sensing*, vol. 20, no. 4, Décembre 1994, pp. 368-380.
- [21] An, P., Chung, C.J.F., "Neural Network approach for geological mapping : technical background and case study", *Canadian Journal of Remote Sensing*, vol. 20, no. 3, Septembre 1994, pp. 293-301.
- [22] Blamire, P.A., "The influence of relative sample size training artificial neural networks", *International Journal of remote sensing*, vol. 17, no. 1, 1996, pp. 223-230.
- [23] De Sève, D., Desjardins, R. et Toutin, Th. "Contribution des données d'ERS-1 dans l'appréhension de l'organisation des linéaments: Le cas de l'astrolème de Charlevoix" *Journal canadien de télédétection*, vol. 20, no 3, 1994 pp. 233 - 244.

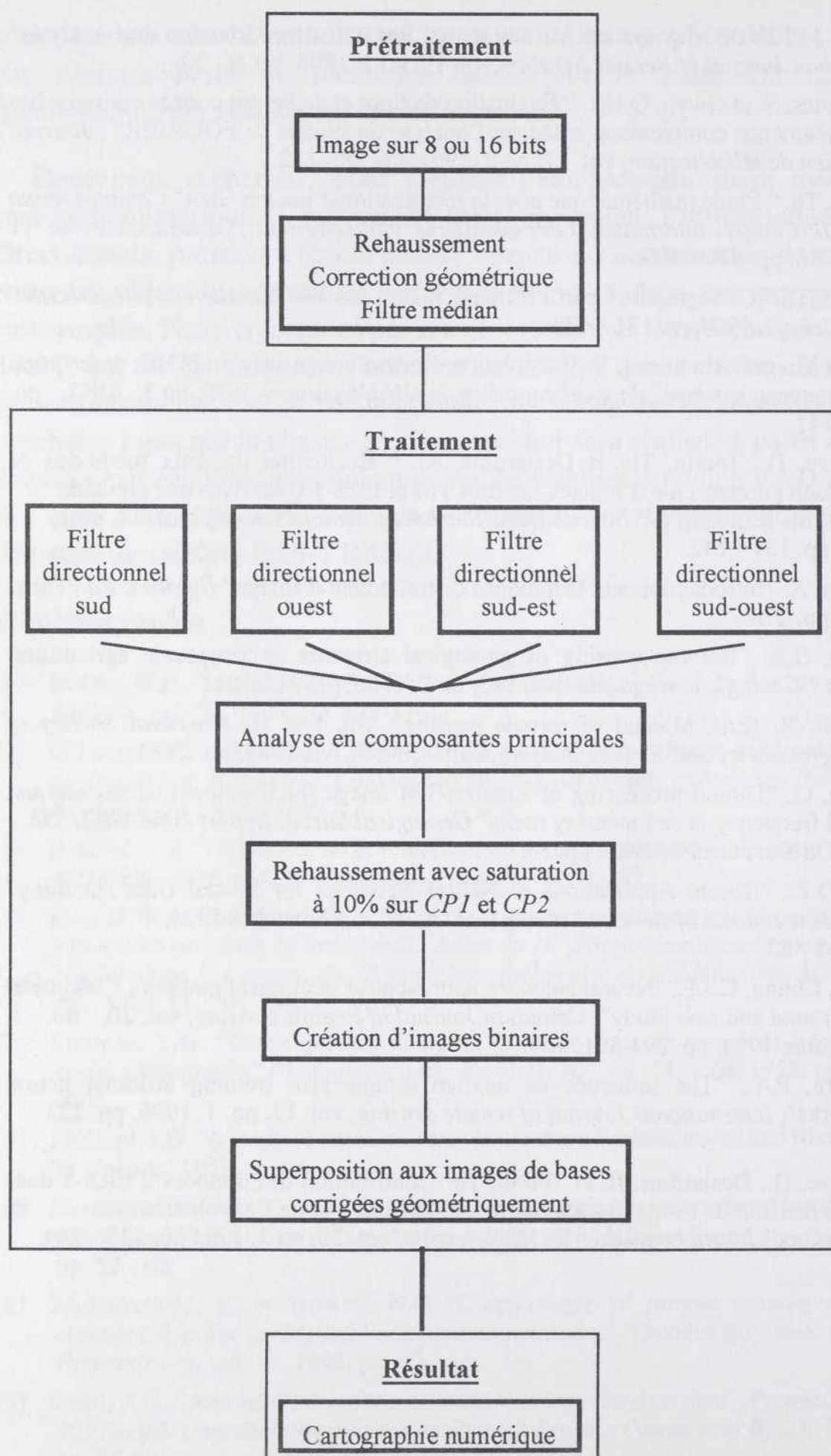


Figure 1: L'approche par filtres directionnels

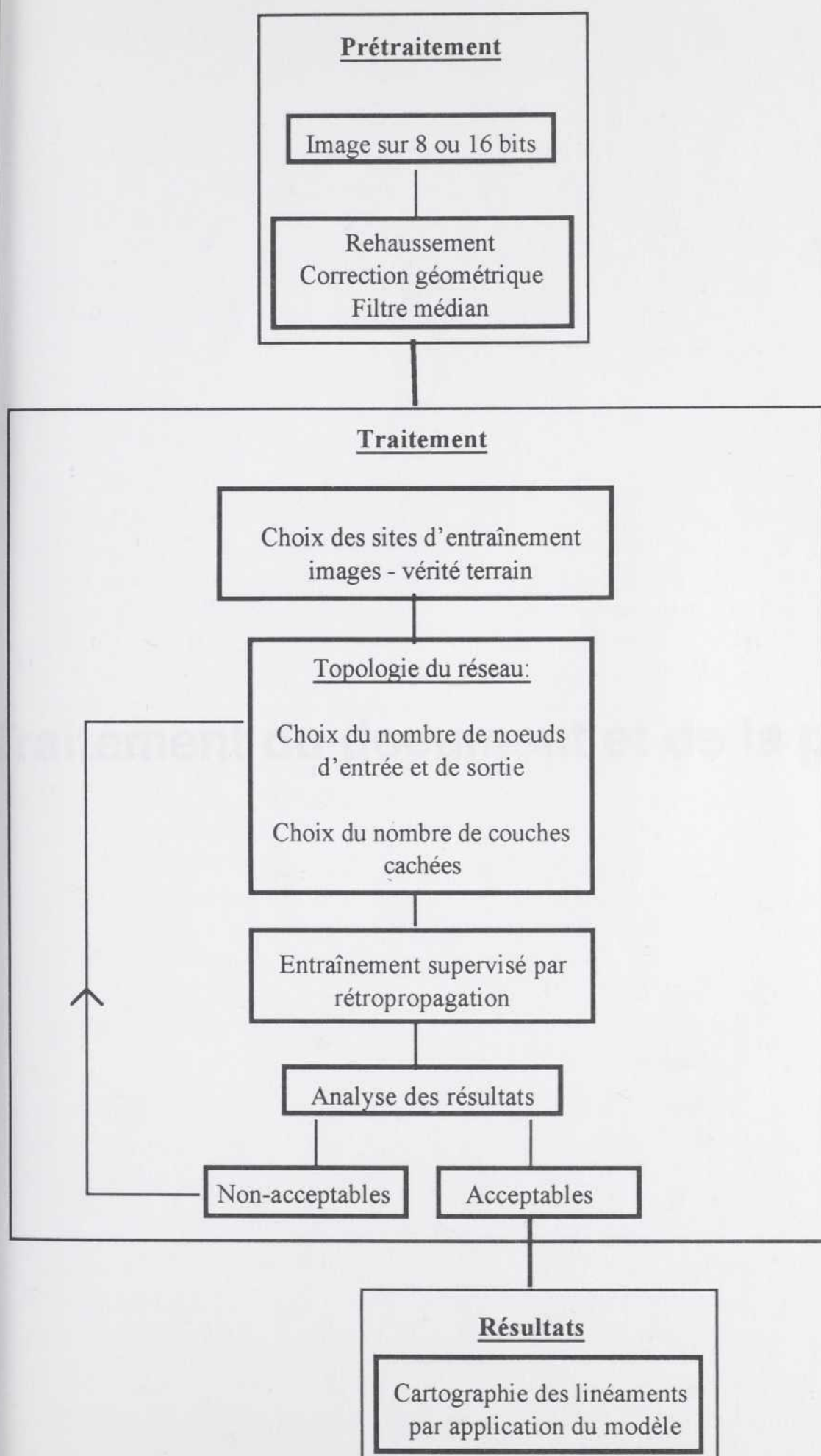


Figure 2: L'approche par réseaux de neurones

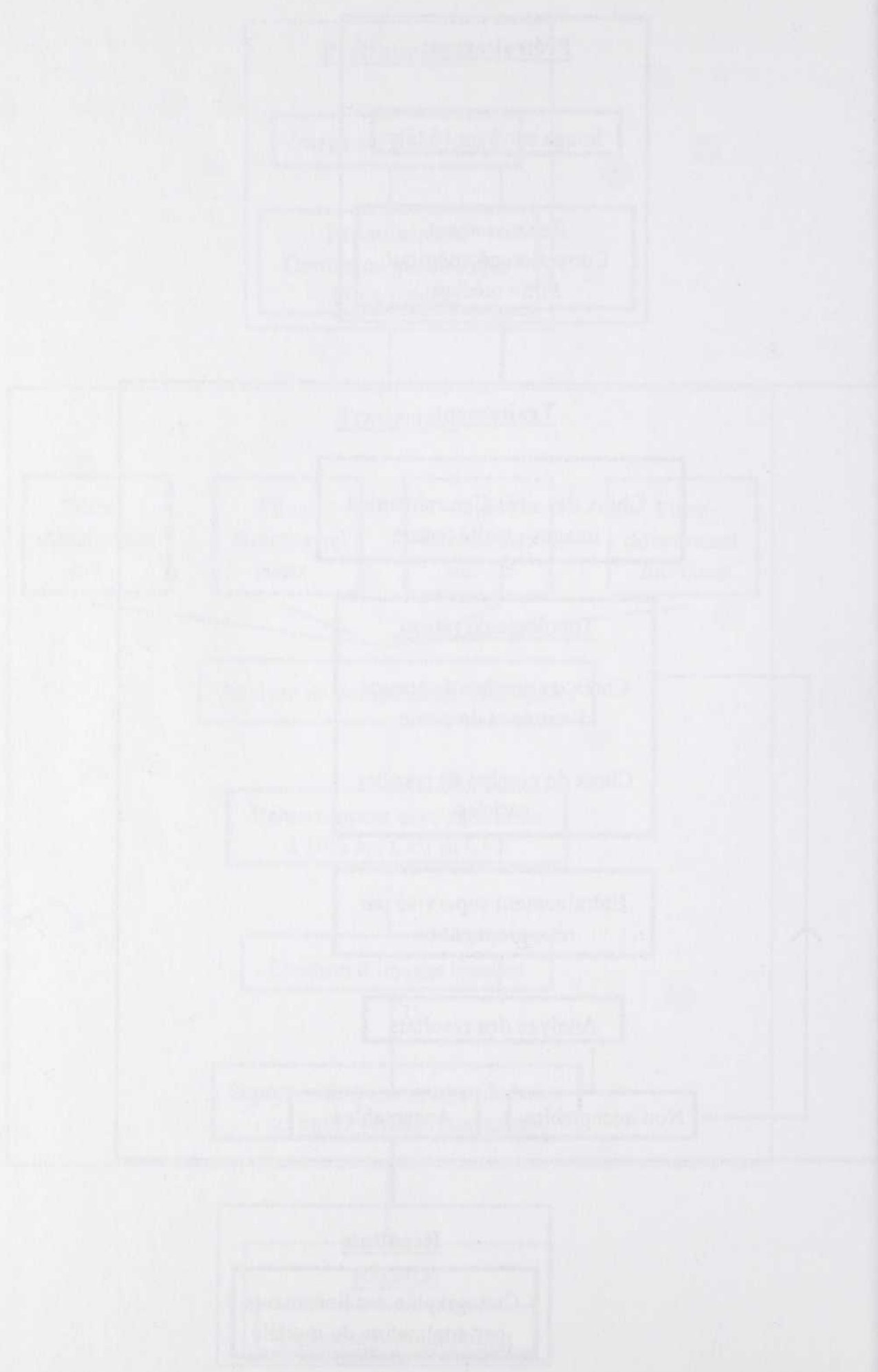


Figure 1. A schematic diagram of the experimental design.

Traitement du document et de la parole

Lect

12

Ecole

e-n

Résum

Cet

ursive

Rumelhe

poles. A

de lectu

trous le

ive. No

à l'écri

manusc

nous tra

ation fl

De plus

Cette m

été obt

tion et

une plu

1. Int

Il se

trop d

cette

chines

lexical

mais q

modél

la lect

quelq

Lecture automatique d'écriture cursive utilisant des concepts perceptifs

M. Côté¹⁻³, E. Lecolinet¹, M. Cheriet², C.Y. Suen³

¹ENST, Département signal, Paris, FRANCE.

²Ecole de Technologie Supérieure, département GPA, Montréal, (Qué.), CANADA.

³CENPARMI, Concordia University, Montréal, (Qué.), CANADA.

e-mail: cote@cenparmi.concordia.ca, elc@sig.enst.fr, cheriet@gpa.etsmtl.ca, suen@cenparmi.concordia.ca

Résumé

Cet article présente une nouvelle méthode de lecture automatique de l'écriture cursive utilisant des concepts perceptifs proposés par les psychologues McClelland & Rumelhart. L'étude proposée se limite à la reconnaissance hors-ligne de mots cursifs isolés. A partir d'observations et d'hypothèses utilisées dans l'élaboration de modèles de lectures, nous décrivons l'organisation de notre système connexionniste et montrons le rôle du mécanisme d'activation dans la perception et la lecture d'écriture cursive. Nous avons introduit dans notre méthode quelques caractéristiques spécifiques à l'écriture cursive. Premièrement, nous utilisons des primitives mieux adaptées au manuscrit telles que les ascendants, les descendants et les boucles. Deuxièmement, nous traitons les ambiguïtés relatives à la position des lettres en introduisant une position floue. La position et les lettres manquantes sont déduites à l'aide d'un lexique. De plus, notre modèle est capable de performer avec des mots de différentes longueurs. Cette méthode de reconnaissance a été implémentée. Des résultats préliminaires ont été obtenus et sont commentés. Notre travail est présentement axé sur la formalisation et la généralisation du modèle proposé et ce, en vue de son fonctionnement sur une plus grande base de données réelles.

1. Introduction

Il semble tout à fait naturel de pouvoir lire un texte manuscrit sans trop d'efforts. Etudier la manière dont nous parvenons à accomplir cette tâche complexe pourrait s'avérer utile pour apprendre aux machines la lecture de textes manuscrits. Déjà, plusieurs modèles d'accès lexical ont été proposés par des chercheurs en sciences cognitives [1] tels que le modèle du logogen [2], le modèle de vérification [3], le modèle d'activation interactive [4]. La plupart portent cependant sur la lecture de textes imprimés. Néanmoins, il est possible d'adapter quelques-uns de leurs concepts sous-jacents [5], [6], [7] à la lecture

automatique d'écriture manuscrite.

Nous présentons ici, une nouvelle méthode de lecture automatique de l'écriture cursive utilisant des concepts perceptifs proposés par les psychologues McClelland & Rumelhart [4]. Ce problème apparemment fort simple pour l'humain est cependant très complexe à réaliser automatiquement étant donné l'ambiguïté inhérente à ce type d'écriture. Notre modèle partage avec McClelland & Rumelhart, les caractéristiques suivantes: réseau connexionniste, traitement parallèle de l'information, mécanisme interactif d'activation (avec des processus ascendants et descendants). Nous démontrons comment nous intégrons les concepts sur la perception et la reconnaissance d'écriture cursive.

La section 2 présente le modèle de McClelland & Rumelhart, la section 3 introduit notre méthode, la section 4 définit les concepts d'activation, la section 5 décrit la technique d'étiquetage utilisée. Les résultats des simulations sont montrés à la section 6. La dernière section conclut sur la méthode proposée.

2. Modèle de McClelland & Rumelhart

Les humains reconnaissent mieux les lettres lorsqu'elles sont présentées dans un contexte que lorsqu'elles sont montrées isolément. Les chercheurs en sciences cognitives ont appelé ce phénomène "effet de la supériorité du mot". Afin de reproduire et modéliser cet effet, McClelland & Rumelhart, ont développés un modèle connexionniste de perception des lettres dans un contexte. Leur système repose sur trois hypothèses fondamentales: (1) la perception se déroule dans un système avec plusieurs niveaux de traitements, chaque niveau représentant un niveau d'abstraction différent, (2) la perception implique des traitements parallèles de l'information visuelle, et (3) les processus en causes sont interactifs c'est-à-dire qu'ils comportent des processus ascendants et descendants. Leur modèle connexionniste possède trois couches (primitives, lettres et mots) organisées hiérarchiquement. Les connexions entre les couches sont excitatrices et inhibitrices. L'expérimentation consiste à trouver une lettre manquante dans un mot de quatre lettres. Les mots stylisés présentés en entrée sont déjà codés avec

des primitives très simples (traits horizontaux, verticaux et obliques). Les lettres sont toujours représentées de la même façon et sont bien définies spatialement.

Comme nous avons en entrée des images de mots cursifs isolés, nous ne pouvons pas compter sur l'invariabilité de la forme, ni sur une localisation spatiale absolue des lettres. Aussi, nous avons introduit dans notre méthode quelques caractéristiques spécifiques à l'écriture cursive. Premièrement, nous utilisons des primitives mieux adaptées au manuscrit telles que les ascendants, les descendants et les boucles. Deuxièmement, nous n'avons pas recours à l'inhibition. Celle-ci est très utile pour discriminer entre les mots semblables lorsque l'information d'entrée est fiable (comme dans le modèle de McClelland & Rumelhart). En effet, l'augmentation d'activation d'une cellule associée à un mot aura pour effet d'inhiber les cellules associées aux mots inappropriés. Cependant, l'inhibition n'a pas que des avantages, particulièrement lorsque l'information extraite est bruitée et peu fiable comme c'est souvent le cas pour les mots cursifs. Dans ce cas, il vaut mieux conserver cette information le plus longtemps possible plutôt qu'en éliminer une partie rapidement en prenant le risque que cette information perdue ne compromette le succès de la reconnaissance. Troisièmement, nous traitons les ambiguïtés relatives à la position des lettres en introduisant une position floue. La position et les lettres manquantes sont déduites à l'aide d'un lexique. Quatrièmement, comme nous travaillons avec des mots de différentes longueurs, le nombre de lettres dans le mot inconnu est estimé.

3. Méthode proposée

Notre système comporte trois niveaux de détecteurs organisés hiérarchiquement: les détecteurs de primitives, les détecteurs de lettres et les détecteurs de mots. Il existe un groupe de neurones-traits et un groupe de neurones-lettres pour chacune des fenêtre d'observation dans l'image.

Ces détecteurs sont actifs ou passifs. Lorsqu'un détecteur est activé, c'est qu'il a détecté un stimuli et qu'il peut agir sur ses voisins, c'est-à-dire les détecteurs des niveaux adjacents auxquels il est con-

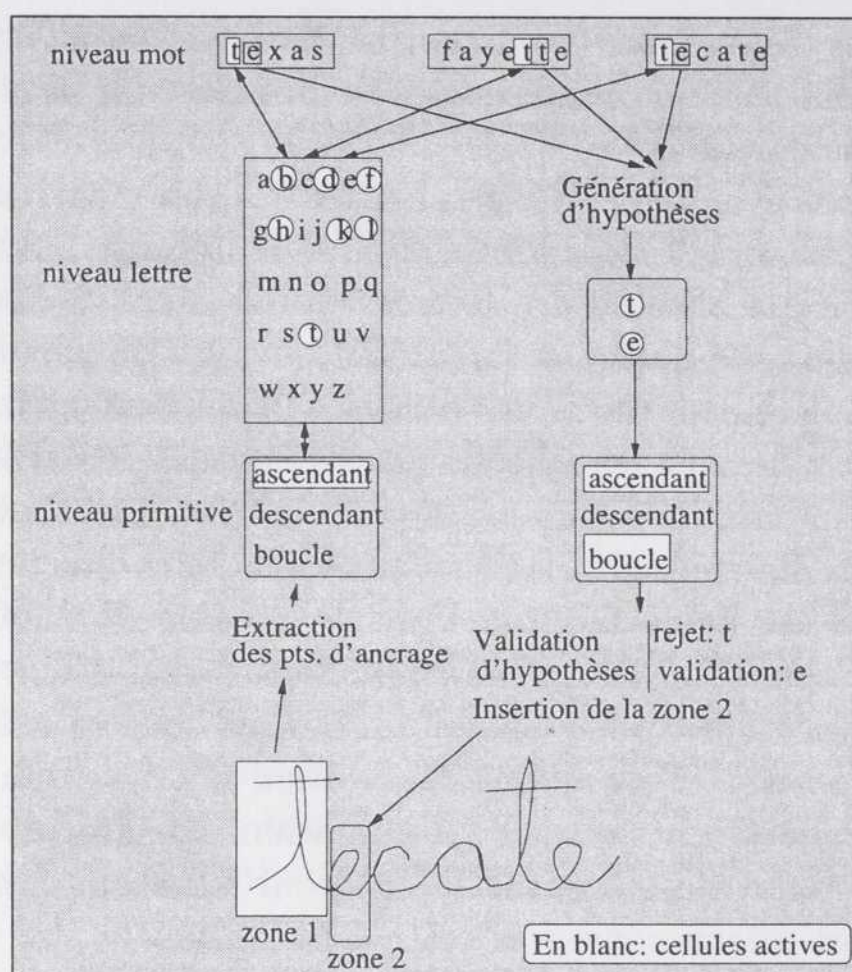


Figure 1: Propagation d'activation entre les trois couches du modèle pour le nom de ville "Tecate".

necté. Deux processus complémentaires vont permettre à l'information de bas niveau (les primitives) de se propager vers le niveau supérieur (les mots), puis à l'information de haut niveau (information lexicale) de se propager vers le bas niveau afin de réaliser une segmentation implicite du mot. Il s'agit du processus ascendant et du processus descendant. A partir de l'image du mot cursif inconnu, les primitives les plus significatives sont d'abord extraites, soit les ascendants, les descendants et les boucles. Ces primitives sont les points d'ancrage du processus tels que les lettres-clés décrites par Cheriet et Suen [8]. Lors du processus ascendant, une fenêtre d'observation est créée pour chaque primitive d'ancrage et les lettres connectées aux primitives d'ancrage sont activées. A ce stade, les lettres du mot inconnu ne sont pas nécessairement toutes reconnues. Puis, toujours lors du processus ascendant, les lettres activées vont stimuler les mots auxquels elles sont connectées. A la fin d'un processus ascendant, on obtient

donc des mots du lexique qui sont activés et d'autres qui ne le sont pas. Les primitives détectées initialement ont donc permis d'éliminer des possibilités et d'en favoriser d'autres. Lors du processus descendants, les mots activés vont générer des hypothèses permettant de donner des indices sur l'identité des lettres inconnues présentes dans l'image de départ. Ces hypothèses sont des lettres. Si on peut retrouver dans l'image initiale, les primitives définissant ces lettres, les hypothèses sont validées, sinon, elles sont rejetées. Ainsi, à la fin d'un processus descendant, des hypothèses sont validées et les détecteurs de primitives qui leurs correspondent sont activés. Ces nouvelles primitives activées sont ajoutées à celle du départ avant de propager leur activation lors d'une autre succession d'un processus ascendant et d'un processus descendant. Le système s'arrête lorsqu'il atteint la saturation après plusieurs cycles de processus ascendants et descendants. Le résultat obtenu est une liste de mots candidats triée par ordre décroissant d'activation. Le mot ayant obtenu la plus grande activation sera alors considéré comme le résultat de la reconnaissance.

Comme les positions des lettres dans le mot inconnu sont relatives les unes aux autres, nous modélisons une position par une fenêtre floue contenant plus ou moins de lettres (pas nécessairement une) et de taille variable. Nous appelons cette "fenêtre floue" une zone qui peut être imaginée comme une fenêtre d'observation sur le signal.

A la Figure 1, **après** la zone 1, on retrouve les hypothèses "e" et "t" (aucune hypothèse n'est proposée **avant** la zone 1). Les primitives trouvées après la zone 1 correspondent aux primitives de la lettre "e" et ne correspondent pas aux primitives de la lettre "t". Conséquemment, l'hypothèse "t" est rejetée tandis que l'hypothèse "e" est vérifiée. Une nouvelle zone, (zone 2) sera donc insérée après la zone 1. A l'intérieur de cette nouvelle zone, la lettre "e" est activée. Au cycle suivant, la nouvelle liste de zones en entrée sera: zone 1, zone 2.

Deux lexiques relient les différentes couches de détecteurs: un lexique primitives-lettres et un lexique lettres-mots. Pendant tout le processus, pour chacun des mots du lexique, un tableau d'étiquetage relie chacune des lettres du mot avec les zones correspondantes dans l'image.

4. Activation

Une cellule est active si son activation est positive. Une cellule inactive n'a aucune influence sur ses voisins. Lorsqu'une cellule ne reçoit aucune activation de ces voisins, son niveau d'activation baisse lentement jusqu'au niveau repos (défini = 0). A un instant t , chaque cellule a une activation $A_i(t)$. L'activation d'une cellule varie entre 0 et 1 selon l'équation suivante:

$$A_i(t + \Delta t) = A_i(t) - \Theta_i A_i(t) + E_i(t) \quad (1)$$

où: $A_i(t + \Delta t)$ activation de la cellule i au temps $(t + \Delta t)$.
 Θ_i constante de décroissance de l'activation de la cellule i .
 $E_i(t)$ contribution des j cellules voisines de i au temps t .

Contribution des j cellules voisines de i au temps t , $E_i(t)$, est représentée par:

$$E_i(t) = n_i(t)(M - A_i(t)) \quad (2)$$

avec: $n_i(t)$ somme des excitations et des inhibitions sur la cellule i au temps t .
 M borne supérieure de l'activation (= 1).

où le facteur $n_i(t)$ est défini comme suit:

$$n_i(t) = \sum_j \alpha_{ij} a_j(t) \quad (3)$$

avec: $a_j(t)$ activation de la cellule j voisine de la cellule i au temps t .
 α_{ij} poids pour l'excitation de la liaison j vers i .

Le facteur $(M - A_i(t))$ module la contribution des voisins $n_i(t)$, afin que l'entrée reste bornée. Quand l'activation d'une cellule atteint sa valeur maximale (un), la contribution de l'entrée est annulée. Ainsi, l'activation est toujours bornée. A partir de l'éq. (2), l'éq. (1) devient:

$$A_i(t + \Delta t) = (1 - \Theta)A_i(t) + n_i(t)(M - A_i(t)) \quad (4)$$

On peut démontrer que l'éq. (4) converge si Θ est positif et inférieur à 1. Nous posons $\Theta = 0.07$ comme dans [4] parce que cette valeur donne une constante de temps raisonnable pour atteindre la saturation.

Les poids du système ne sont pas appris. Ils s'adaptent durant le traitement suivant les équations suivantes:

$$\alpha_{pl} = \frac{1}{NP}, \quad \alpha_{lm} = \mathcal{F}(\Delta)_{lm} \frac{1}{NZ}, \quad \alpha_{ml} = \frac{1}{NM}$$

où NP est le nombre de primitives trouvées dans l'image pour la lettre L , $\mathcal{F}(\Delta)_{lm}$ est le coefficient-position de la lettre L dans le mot M (cf. Section 5), NZ est le nombre de zones trouvées dans le signal au temps t , et NM est le nombre de mots dans le lexique contenant la lettre L .

5. Etiquetage

Lors du processus ascendant, nous cherchons d'abord à mettre en correspondance chaque lettre activée par les primitives avec les mots du lexique et au bon endroit à l'intérieur du mot. Pour ce faire, une technique d'étiquetage a été développée qui tient compte de l'ordre relatif des lettres et de la position floue d'une lettre à l'intérieur d'un mot du lexique.

L'étiquetage des zones se fait des extrémités du mot vers l'intérieur jusqu'au centre mimant ainsi certains mécanismes de la lecture chez l'humain [9]. Les lettres contenant des primitives signifiantes sont instanciées en priorité.

L'estimation de la position relative des lettres est un facteur important pour bien faire l'étiquetage des lettres dans chacun des mots du dictionnaire. Lors de l'étiquetage, nous comparons toujours la position de la lettre testée avec la position correspondante dans le signal. Etant donné la grande incertitude sur le positionnement des lettres, nous introduisons ce que nous appelons un coefficient de position (réel compris entre 0 et 1). La valeur 1 signifie que la lettre du mot correspond exactement avec la lettre du signal. Par opposition, la valeur 0 signifie que la lettre du mot ne correspond pas du tout avec la lettre

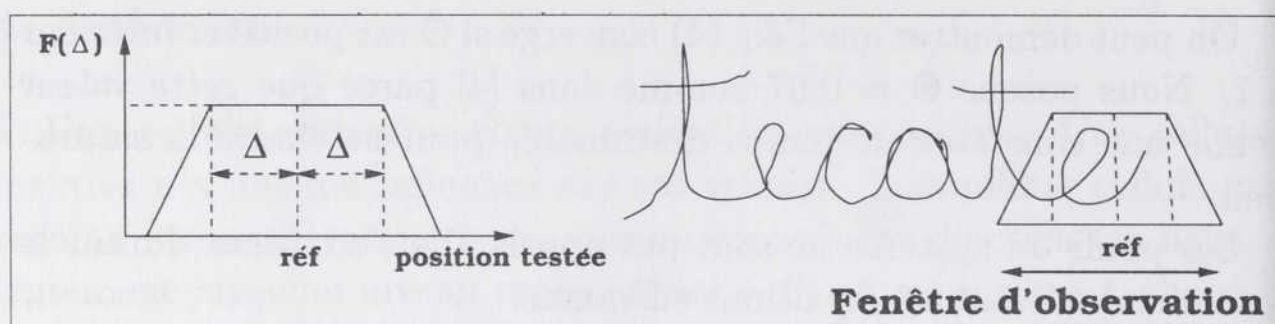


Figure 2: Fonction du coefficient-position.

du signal. Entre ces deux limites, ce coefficient varie selon la fonction $\mathcal{F}(\Delta)$ de la Fig. 2.

Dans cette figure, la référence est la position de la lettre dans une zone donnée, Δ correspond à la largeur de la lettre dans l'image (longueur du mot inconnu divisée par le nombre estimé de lettres dans le mot inconnu) et la position test est tout simplement une estimation de la position approximative de la lettre. Cette fonction peut être imaginée comme une fenêtre d'observation de largeur 4Δ glissant sur l'image.

6. Simulations et résultats préliminaires

Actuellement, le module d'extraction de primitives fournit au module de reconnaissance, toute l'information sur les primitives détectées dans l'image. Dans la méthode utilisée en [7], le lexique contient 27 primitives de base. A la suite d'observations, ces primitives ont été classées en deux catégories: les primitives significantes et les autres. Il y a 4 primitives de la première catégorie: les ascendants, les descendants, les boucles et les e_traits. La deuxième catégorie inclue les boucles instables comme celle des lettres "d, p, b, q" et les traits comme dans "m, n, w". Parce que ces primitives sont souvent absentes, leur présence est identifiée uniquement lorsqu'elles sont trouvées à proximité des primitives de première catégorie. L'identification des primitives s'effectue donc en deux étapes: d'abord, les primitives significatives sont trouvées ensuite, les primitives conditionnelles sont identifiées relativement aux primitives significatives existantes.

Le lexique contient 60 mots de 3 à 9 lettres chacun. Il est relié à une base de données de 60 images de mots cursifs (noms de ville) provenant d'adresses postales.

Le module estimant le nombre de lettres dans le mot inconnu n'est pas encore disponible. Cette information est donc donnée en paramètre d'entrée au programme pendant les simulations. L'influence de ce paramètre peut être étudiée en variant sa valeur.

Les résultats obtenus avec le mot "Tecate", "Ithaca", et "Ten" sont présentés à la Figure 3. Dans les images de la deuxième rangée, les segments sont des représentations virtuelles de traits ou d'ascendants. Les boucles sont schématisées par des cercles. Dans les images de la troisième rangée, chaque zone est délimitée par un rectangle. Comme on peut observer, les primitives les plus significatives telles que les ascendants et les boucles ont été détectées et validées.

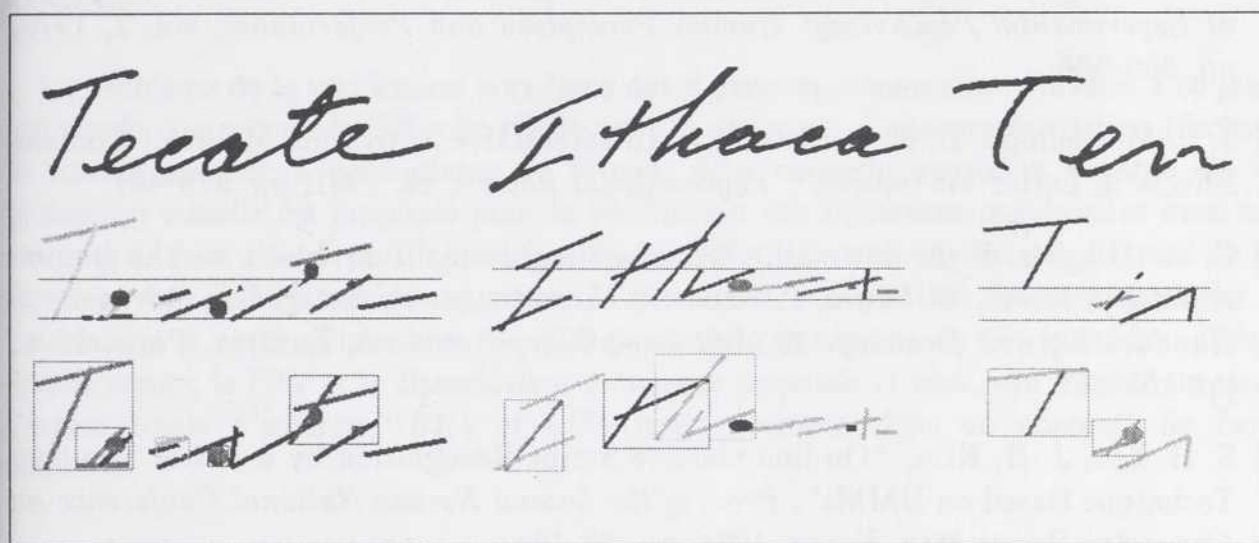


Figure 3: Les mots "Tecate", "Ithaca", et "Ten" avant traitement, après extraction des primitives, et les zones créées à la suite des processus ascendants et descendants.

7. Conclusions

Une méthode de reconnaissance de mots cursifs isolés utilisant des concepts perceptifs a été proposée. Après avoir présenté l'organisation de notre système, nous avons montré le rôle du mécanisme d'activation dans la perception et la lecture d'écriture cursive. Ce modèle a été implémenté et testé sur des données simulées. Nous travaillons actuellement à la validation de cette méthode sur des images réelles.

Remerciements

Cette recherche a été supportée en partie par les organismes suivants: CRSNG du Canada, FCAR du Québec et le réseau des Centres d'Excellence IRIS.

Références

- [1] M. Taft, *Reading and the Mental Lexicon*, Lawrence Earlbaum, Hillsdale, New Jersey, 1991.
- [2] J. Morton, "Interaction of Information in Word Recognition", *Psychological Review*, vol. 76, 1969, pp. 165-178.
- [3] C. A. Becker, "Allocation of attention during visual word recognition", *Journal of Experimental Psychology: Human Perception and Performance*, vol. 2, 1976, pp. 556-566.
- [4] J. L. McClelland, D. E. Rumelhart, "An Interactive Activation Model of Context Effects in Letter Perception", *Psychological Review*, 88, 1981, pp. 375-407.
- [5] C. A. Higgins, P. E. Bramall, "Handwriting recognition based on the human reading process", C. Faure, P. Keuss, G. Lorette, et A. Vinter, Ed., *Advances in Handwriting and Drawing: A Multidisciplinary Approach*, Europia, Paris, 1994, pp. 155-167.
- [6] S. H. Lee, J. H. Kim, "On-line Cursive Script Recognition by a Letter Spotting Technique Based on HMMs", *Proc. of the Second Korean National Conference on Character Recognition*, Korea, 1994, pp. 93-104.
- [7] E. Lecolinet, "Cursive Script Recognition by Backward Matching", C. Faure, P. Keuss, G. Lorette, et A. Vinter, Ed., *Advances in Handwriting and Drawing: A Multidisciplinary Approach*, Europia, Paris, 1994, pp. 117-135.
- [8] M. Cheriet, C. Y. Suen, "Extraction of Key Letters for Cursive Script Recognition", *Pattern Recognition Letters*, vol. 11, 1993, pp. 1009-1017.
- [9] I. Taylor, M. M. Taylor, *The Psychology of Reading*, Academic Press, New York, 1983.

Une approche perceptuelle pour la définition d'une représentation adaptée au problème de la vérification hors-ligne des signatures manuscrites

Robert Sabourin, Ph.D., ing.

*Laboratoire d'Imagerie, de Vision et d'Intelligence Artificielle (LIVIA)
École de technologie supérieure, département de génie de la production automatisée,
4750 ave. Henri-Julien, Montréal (Québec), CANADA, H2T 2C8.
e-mail: sabourin@gpa.etsmtl.ca*

Résumé

Le problème de la vérification hors-ligne des signatures manuscrites demeure à ce jour non résolu. La principale difficulté réside dans la définition d'une représentation (facteur de forme) adaptée à cette classe de formes. Une nouvelle approche inspirée de la perception visuelle est proposée pour la vérification des signatures manuscrites dans un contexte de faux aléatoires. L'objectif est de quantifier les variations locales du tracé de la signature sans l'obligation de segmenter en primitives. L'évaluation de cette approche a été réalisée à partir d'une base de 800 signatures provenant de 20 individus. Deux classificateurs, le PPV et le classificateur à distance minimale et seuil, ont montré un taux d'erreur totale d'environ 0.02% et 1.0% respectivement, dans un contexte de faux aléatoires.

1. Introduction

Le problème de la vérification hors-ligne des signatures manuscrites demeure à ce jour non résolu. Ce sujet fut l'objet d'intenses recherches depuis les vingt dernières années [1,2]. La principale difficulté réside dans la définition d'une représentation (facteur de forme) adaptée à cette classe de formes. L'humain est en mesure de distinguer facilement la signature authentique et le faux aléatoire défini comme la signature authentique d'un autre individu. En vérification automatique, cette tâche demeure à ce jour très complexe à réaliser. Comment expliquer la difficulté d'automatiser ce processus avec succès ?

Une revue de la littérature dans le domaine de la vérification hors-ligne des signatures manuscrites dans un contexte de faux aléatoires fait ressortir plusieurs approches [1,2] : les transformées 2D, les histogrammes de données directionnelles ou de courbure, les projections horizontales et verticales du

tracé de la signature, les approches structurelles, les mesures locales effectuées sur la trace de la signature, la position de points caractéristiques localisés sur le squelette de la signature.

Une analyse des principales approches proposées dans la littérature indique que le problème de la vérification des signatures manuscrites doit être reformulé pour être en mesure de proposer une solution originale à l'élimination des faux aléatoires. Nous devons porter une attention particulière aux caractéristiques intrinsèques de la signature authentique. Plamondon [2] avance le fait que la signature manuscrite est un cas particulier de l'écriture étant donné qu'elle est le résultat d'un mouvement spontané. En conséquence, l'allure générale de la signature authentique demeure sensiblement la même lorsque celle-ci est apposée dans un cadre pré-établi comme c'est habituellement le cas pour les chèques bancaires.

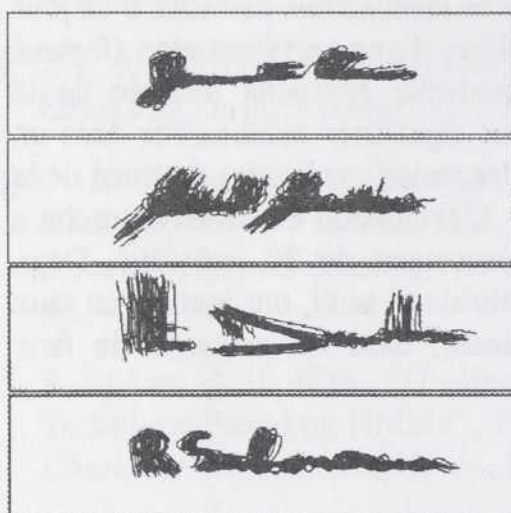


Figure 1

Malgré tout, un problème fondamental subsiste, celui de la définition d'une représentation (facteur de forme) adaptée au problème de la vérification. En fait, le tracé de la signature authentique est caractérisé par une grande variation intra-classe [Figure 1]. Ces variations locales du tracé de la signature représentent une caractéristique intrinsèque de l'identité du scripteur [3]. A titre d'exemple, 20 squelettes provenant de signatures authentiques du même scripteur ont été centrés et superposés à la Figure 1.

Pour les quatre scripteurs à l'étude, il ressort nettement que les proportions et l'orientation générale des signatures dans le plan sont relativement stables. Cependant, les signatures sont caractérisées par de grandes variations locales du tracé. L'amplitude de ces variations change en fonction de la position des traits à l'intérieur de la signature et différemment d'un scripteur à l'autre.

2. Critères pour la définition d'une représentation adaptée à la signature

Nous avons vu à la section précédente que les approches locales, en accord avec la méthodologie de l'expert, semblent mieux adaptées que les approches

globales. Une étude attentive de la superposition des squelettes de la Figure 1 permet d'avancer les hypothèses suivantes :

- 1) *les proportions et l'orientation générale des signatures authentiques apposées dans un contexte pré-établi sont relativement stables pour chaque scripteur, et que*
- 2) *les variations locales du tracé de la signature sont des caractéristiques intrinsèques de l'identité du scripteur et elles doivent également être considérées dans la définition d'une représentation adaptée. Ce phénomène est caractérisé par un déplacement local d'amplitude variable des traits suivant l'axe principal (horizontal) de la signature.*

Si ces hypothèses sont vérifiées sur l'ensemble de la population, il résulte une certaine invariance de la signature authentique pour ce qui est de l'échelle et de l'orientation. En conséquence, seule une correction en translation est requise pour la définition d'une représentation adaptée au problème de la vérification des signatures.

Un aspect important à considérer lors de la définition d'une représentation de la signature est l'absence systématique d'activité dans des zones de l'image entourant la silhouette de la signature. C'est en relation avec l'hypothèse 1 avancée précédemment et qui stipule que les proportions et que l'orientation générale sont relativement stables. En d'autres termes, l'étude du fond et de la forme doivent être considérés conjointement pour représenter la signature manuscrite.

3. Une représentation de la signature inspirée de la perception visuelle

Une nouvelle approche inspirée de la perception visuelle est proposée pour la vérification des signatures manuscrites dans un contexte de faux aléatoires. L'originalité de la technique proposée est que les mesures locales ne sont pas effectuées sur des points (primitives) localisés sur le tracé de la signature, mais à partir de points d'attention localisés dans le plan image. Cette technique permet la prise en compte du fond et de la forme lors de l'analyse de la signature à l'étude [Hypothèses 1 & 2]. En plus, l'identification et la localisation des points d'intérêts n'est plus nécessaire.

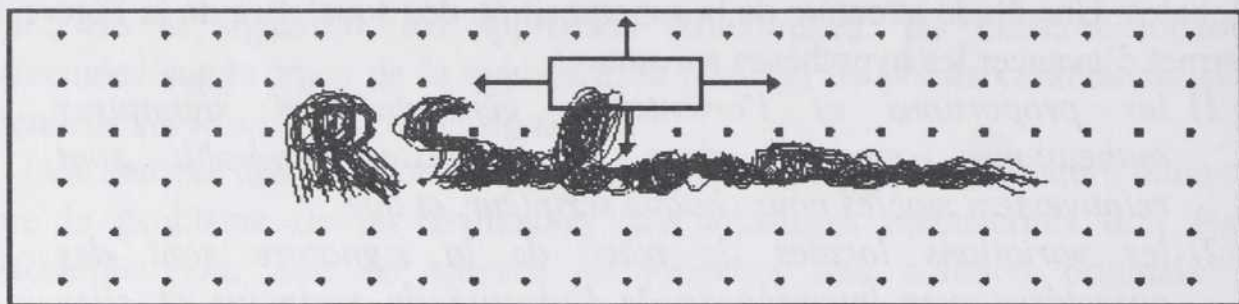


Figure 2

Une image de 512x128 pixels est centrée sur une grille constituée de points d'attention définis arbitrairement dans la scène [Figure 2]. Une rétine est déplacée successivement d'un point d'attention à l'autre. À chaque point d'attention, la rétine est excitée par une portion de l'image de la signature à l'étude. Une transformation morphologique appliquée sur la portion d'image perçue permet une mesure locale du niveau d'activité du signal. Le spectre de forme, résultat d'une suite d'ouvertures morphologiques appliquées dans l'espace-image occupé par la rétine, a été choisi pour caractériser la structure du tracé perçu par la rétine.

4. Le spectre de forme comme facteur de forme local

4.1 Définition des principaux opérateurs morphologiques

Les opérations de base en morphologie mathématique sont l'érosion et la dilatation [4]. Soit B un élément structurant (ES) de forme quelconque. L'érosion de l'ensemble X par rapport à B , noté $X \ominus B^s$, est défini par tous les points $z \in E$ pour lesquels les translatés B_z sont totalement inclus dans X :

$$X \ominus B^s = \{z \in E : B_z \subset X\} \quad (1)$$

où B^s correspond au transposé de B dans le plan E (c'est-à-dire son symétrique par rapport à son origine) et $\ominus$ représente la soustraction de Minkowski. La dilatation de l'ensemble X par rapport à l'ES B est défini par tous les points $z \in E$ pour lesquels le translaté B_z touche l'ensemble X :

$$X \oplus B^s = \{z \in E : B_z \cap X \neq \emptyset\} \quad (2)$$

où $\oplus$ représente l'addition de Minkowski. L'ouverture de l'ensemble X par l'ES B , noté par X_B , est une opération morphologique résultant de l'application d'une érosion suivie par une dilatation. Cette opération morphologique correspond à l'union de tous les translatés B_z de B inclus à l'intérieur de l'ensemble X :

$$(X \ominus B^s) \oplus B = X_B = \bigcup \{B_z : B_z \subset X\}. \quad (3)$$

Enfin, l'opération de *fermeture* X^B est le résultat de la dilatation suivie par une érosion, une opération duale de l'ouverture parce qu'elle peut être réalisée par l'ouverture du complémentaire X^c de X :

$$(X \oplus B^S) \ominus B = X^B = [(X^c)_B]^c. \quad (4)$$

Une *granulométrie* est le résultat de l'ensemble des ouvertures X_{rB} de l'objet X avec les éléments structurants rB , avec $r \in \mathcal{R} \rightarrow X_{rB}$. En d'autres termes, la transformation d'un ensemble X résultant de l'application de la transformation $\Psi(\cdot)$ suivie par la mesure $m(\cdot)$, c'est-à-dire $X \rightarrow \Psi(X) \rightarrow m(\Psi(X))$ est une granulométrie [4,5,6].

Le *pecstrum* (ou spectre de forme) est le résultat de l'application successive d'opérateurs morphologiques sur un ensemble quelconque X en utilisant un élément structurant rB qui augmente en taille. L'ouverture d'un signal avec un élément structurant de dimension r conduit à éliminer du signal des détails plus petits que l'ES rB . La version discrète du pecstrum est formulée par [5] :

$$p(n) = \frac{m(X_{nB}) - m(X_{(n+1)B})}{m(X)}, \quad n = 0, 1, \dots, k-1, \quad (5)$$

avec

$$\sum_{n=0}^k p(n) = 1. \quad (6)$$

Chaque valeur de $p(n)$ représente la fraction de la surface totale de l'ensemble X rejetée par l'ouverture de cet ensemble par l'ES $(n+1)B$, étant donné que l'ouverture avec nB a déjà été effectuée. L'ES utilisé est le multiple entier (nB) du prototype de base B . La valeur $n=k-1$ correspond à la valeur pour laquelle un objet binaire disparaît complètement après l'ouverture avec l'ES kB . En plus, un spectre caractérisé par une impulsion $p(n)=1$ signifie que l'objet binaire X a été complètement éliminé par une ouverture avec l'ES $(n+1)B$. Le pecstrum est une transformation invariante en translation et en rotation (dans les limites d'une trame carrée); cependant cette transformation ne possède pas la propriété d'invariance par rapport à l'homothétie. Cette dernière propriété n'est pas essentielle dans notre méthodologie parce que les granulométries sont toutes évaluées sur le même type de rétines dans l'image, c'est-à-dire que les rétines sont toutes de même taille ce qui procure une normalisation locale de l'image et de la mesure effectuée [Figure 2].

La partie négative du pecstrum est obtenue en mesurant le changement de l'aire de l'objet après l'application successive de fermetures [5] :

$$p(-n) = \frac{m(X_{nB}) - m(X^{(n-1)B})}{m(X)}, \quad n = 1, \dots, k. \quad (7)$$

La partie négative du spectre de forme est intimement associée à une mesure normalisée de la dimension des trous et des cavités de l'objet à l'étude. Les premières composantes des spectres de forme positifs et négatifs contiennent de l'information associée à la rugosité de l'objet. Le principal inconvénient du spectre négatif sur le plan de la mise en oeuvre tient dans la difficulté d'estimer correctement la valeur maximum de n , e.g. la valeur de k [5,6].

Le *pseudospectre* (*pseudospectrum*) a été introduit par Anastassopoulos [5] comme une solution au problème de la mise en oeuvre du spectre de forme négatif. Le pseudospectre p_s d'un objet binaire X correspond au spectre positif appliqué sur le complémentaire de X , soit X^c , borné par le cercle circonscrit à l'objet et ayant son centre de gravité correspondant à celui de X ; il est clair que le cercle circonscrit contient l'objet X . Les valeurs de $p_s(n)$ sont normalisées en fonction de l'aire de l'objet X et la forme de l'ES est supposée circulaire. Le processus d'évaluation de p_s termine lorsque l'aire de X^c devient nulle. L'information contenue dans le pseudospectre n'est pas identique à celle contenue dans le spectre négatif. Le pseudospectre contient de l'information pour ce qui est des trous et des cavités de l'objet X , mais cette transformation mesure également la ressemblance de l'objet à un cercle. Les premières composantes du pseudospectre peuvent être affectées par les effets de bords lesquels proviennent des régions où le cercle circonscrit touche l'extérieur de l'objet X .

4.2 Définition d'une représentation locale de la signature

Suivant le concept d'une représentation locale de la signature basée sur la perception visuelle [Section 3], le niveau d'activité du signal localisé dans le champ de vision d'une rétine peut être caractérisé de différentes façons à l'aide des opérateurs morphologiques. Définissons W , l'ensemble des pixels inclus dans la surface couverte par la rétine localisée à un point d'attention dans l'image [Figure 3]. Soit X , l'ensemble des pixels appartenant au signal qui excite une rétine. Par exemple, l'ensemble X montré à la Figure 3 est défini par les quatre portions de tracé de la signature localisées dans le champ de vision de la rétine rectangulaire W , $X = \bigcup_{i=1}^4 \{X_i\} \subseteq W$. Évidemment, $X = \emptyset$

lorsqu'il y a absence de signal dans le champ de la rétine. Enfin, l'ensemble E représente l'ensemble W augmenté utilisé dans le cas du pseudospectre dans le but d'éliminer les effets de bords caractérisés par la proximité des pixels de l'ensemble X localisés dans le voisinage de la bordure externe de l'ensemble W . Ici, il y a toujours absence de signal dans la zone extérieure. Plusieurs

stratégies ont été évaluées pour la définition d'une mesure $m(\cdot)$ représentant le niveau d'activité du signal X placé dans le champ de la rétine W .

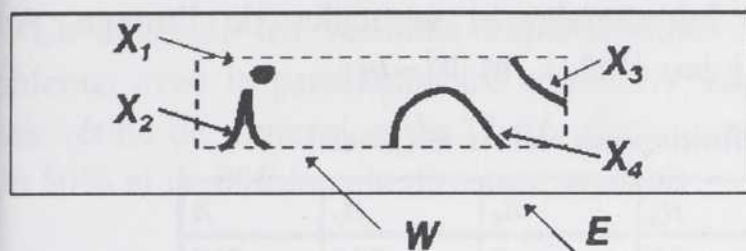


Figure 3

donné l'absence d'effets de bords dans ce cas, alors la rétine de base est considérée avec $W = E$. Ensuite, le *pseudospectre* $\Psi p_s(X^c)$ a été appliqué à l'ensemble X^c (le complémentaire de X par rapport au domaine défini par l'ensemble E , avec $W = E$). Lorsque l'ES utilisé pour la transformation $\Psi p_s(X^c)$ est un segment de l'ensemble $\{—, |, /, \backslash\}$, alors l'ensemble E est une surface rectangulaire telle que montrée à la Figure 3 avec $W = E$. Dans le cas d'une transformation $\Psi p_s(X^c)$ basée sur un ES circulaire, alors l'ensemble E correspond à la région circulaire définie par le cercle circonscrit à l'ensemble W [Figure 4b]. Dans les deux cas, la transformation $\Psi p_s(\cdot)$ est appliquée sur le complémentaire de l'ensemble X , i.e. sur $X^c \subseteq E$, peu importe le choix de l'ES.

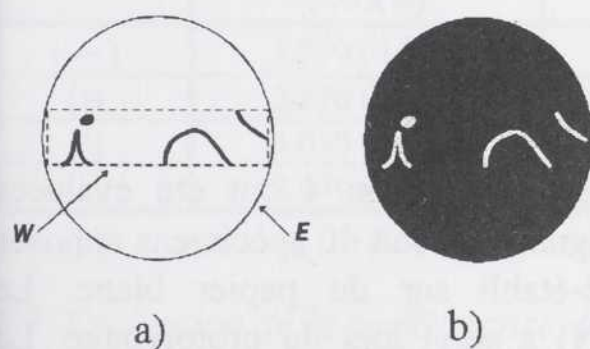


Figure 4

Enfin, le *pseudospectre augmenté* $\Psi ap_s(X^c)$ est simplement le pseudospectre basé sur l'un des ES de $\{—, |, /, \backslash\}$ et appliqué sur X^c ; le domaine du complémentaire étant défini par E , avec $X \subseteq W \subset E$, $(W \cap E)^c \neq \emptyset$, et $X^c \subseteq E$. Dans ce cas, un facteur d'échelle horizontal et vertical de 50% a été appliqué dans la définition de l'ensemble E [Figure 3]. Dans le cas d'un ensemble E de forme circulaire, l'augmentation de la surface n'est pas nécessaire parce que la forme de l'ES et celle de l'ensemble E sont identiques; en conséquence, il n'y a pas d'effets de bords dans ce cas.

Le champ de vision d'une rétine est directement proportionnel à sa dimension. La longueur horizontale (h) et verticale (v) d'une rétine (en pixels) permet l'évaluation de la cardinalité de l'ensemble W , $|W|$ [Tableau 1]. Les dimensions horizontales et verticales des images de signatures sont de 512 et

Premièrement, le *spectre positif* a été utilisé pour la définition d'une transformation $\Psi p(X)$ appliquée à l'ensemble X des pixels inclus dans W , avec $W = E$ et $(W \cap E)^c = \emptyset$. Étant

Finalement, le *pseudospectre augmenté* $\Psi ap_s(X^c)$ est simplement le pseudospectre basé sur l'un des ES de $\{—, |, /, \backslash\}$ et appliqué sur X^c ; le domaine du complémentaire étant défini par E , avec $X \subseteq W \subset E$, $(W \cap E)^c \neq \emptyset$, et $X^c \subseteq E$. Dans ce cas, un facteur d'échelle horizontal et vertical de 50% a été appliqué dans la

128 pixels respectivement; le nombre de rétines (n_r) peut être évalué à partir du produit ($n_h * n_v$), où n_h et n_v correspondent au nombre de rétines appliquées dans les directions horizontales et verticales de l'image. Au Tableau 1, nous avons $n_h = 512/h$, $n_v = 128/v$, et $|W| = h * v$.

Tableau 1: Définition des rétines W (pixels)

$h \times v$	n_h	n_v	n_r	n
16x16	32	8	256	768
16x32	32	4	128	384
32x16	16	8	128	384
32x32	16	4	64	192
64x32	8	4	32	96

Finalement, la représentation de la signature χ_i est un vecteur de caractéristiques de dimension n égale au produit de DL, le nombre de descripteurs locaux évalués sur une rétine et n_r , le nombre total de rétines appliquées sur l'image ($n = DL * n_r$). Nous avons utilisé les trois premiers moments granulométriques (la moyenne, la variance et le skewness de la distribution $p(n)$, alors $DL=3$) comme descripteurs locaux résultant de l'évaluation des mesures $m(\cdot)$ appliquées sur les transformations $\Psi(\cdot)$ de l'ensemble E , où $E \rightarrow \Psi(E) \rightarrow m(\Psi(E))$.

5. Expérimentation

5.1 Prototypage

Toutes les représentations χ_i définies à la section 4 ont été évaluées expérimentalement sur une base de 800 signatures, soit 40 spécimens apposés par 20 scripteurs dans un cadre pré-établi sur du papier blanc. Le classificateur du plus proche voisin (NN) a servi lors du prototypage. Le protocole expérimental est le même qui a servi pour caractériser plusieurs systèmes de vérification mis en oeuvre à notre laboratoire [7,8]. La performance du système de vérification est mesurée en terme des taux d'erreur de Type I (ε_1 , le rejet de signatures authentiques) et de Type II (ε_2 , l'acceptation de faux aléatoires) évalués pour les 20 scripteurs. Le taux d'erreur total moyen ε_t du système de vérification est exprimé en terme de la moyenne des taux d'erreurs de Type I et de Type II, c'est-à-dire $\varepsilon_t = (\varepsilon_1 + \varepsilon_2)/2$. Les valeurs de ε_t ont servies pour l'évaluation des facteurs de forme. Les paramètres investigués sont: le nombre et la dimension des rétines, le choix du spectre de forme utilisé pour quantifier le taux

d'activité du signal dans le champ de vision des rétines, le pourcentage de chevauchement entre les rétines dans les directions horizontales et verticales.

En résumé, les résultats expérimentaux les plus performants ont été obtenus avec le pseudospectre augmenté $\Psi_{ap_s}(X^c)$ et les ES $\{ |, —, /, \backslash \}$ dans une rétine de dimension de 32x16 pixels avec un chevauchement horizontal de 50% et de 0% dans la direction verticale.

Tableau 2: Résultats expérimentaux, pseudospectre augmenté $\Psi_{ap_s}(X^c)$ et le classificateur du plus proche voisin (25 itérations)

ES	$\bar{\varepsilon}_1(\%)$ ($\pm\sigma$)(%)	$\bar{\varepsilon}_2(\%)$ ($\pm\sigma$)(%)	$\bar{\varepsilon}_t(\%)$ ($\pm\sigma$)(%)
{—}	0.000 (0.000)	0.040 (0.037)	0.020 (0.019)
{ }	0.000 (0.000)	0.038 (0.046)	0.019 (0.023)
{/}	0.000 (0.000)	0.036 (0.044)	0.018 (0.022)
{\}	0.000 (0.000)	0.031 (0.039)	0.016 (0.020)
{•}	0.200 (0.141)	0.153 (0.092)	0.177 (0.081)

Tableau 3: Résultats expérimentaux, pseudospectre augmenté $\Psi_{ap_s}(X^c)$ et le classificateur à distance minimum et seuil (25 itérations) avec $N_{ref} = 6$.

ES	$\bar{\varepsilon}_1(\%)$ ($\pm\sigma$)(%)	$\bar{\varepsilon}_2(\%)$ ($\pm\sigma$)(%)	$\bar{\varepsilon}_t(\%)$ ($\pm\sigma$)(%)
{—}	1.550 (0.548)	0.145 (0.153)	0.848 (0.260)
{ }	2.670 (0.754)	0.480 (0.189)	1.575 (0.367)
{/}	3.010 (0.684)	0.318 (0.203)	1.664 (0.348)
{\}	2.070 (0.832)	0.305 (0.185)	1.188 (0.418)

5.2 Évaluation de la performance des classificateurs

Les représentations retenues sont maintenant évaluées sur la base de 25 itérations où les ensembles d'apprentissage et de tests ont été redéfinis aléatoirement. La moyenne des taux d'erreurs $\bar{\varepsilon}_t$ est évaluée sur la base des 25 expériences. Les résultats expérimentaux obtenus avec un classificateur de type plus proche voisin et avec un classificateur à distance minimum et seuil sont présentés aux Tableaux 2 et 3 respectivement. Pour ce qui est du classificateur à seuil, le nombre de signatures de référence a été fixé à 6 pour être en mesure de comparer ces résultats à ceux des travaux antérieurs [7,8]. En résumé, le taux d'erreur total moyen $\bar{\varepsilon}_t$ varie dans la plage des valeurs (0.016% - 0.20%) pour le plus proche voisin [Tableau 2] et dans la plage de

valeurs (0.848% - 1.664%) pour le classificateur à distance minimum et seuil [Nref=6 signatures de référence, Tableau 3]. Ces résultats expérimentaux sont du même ordre de grandeur que ceux obtenus avec un facteur de forme de type local évalué sur la même base de signatures [7,8].

6. Conclusion

Une nouvelle approche basée sur la perception visuelle a été proposée pour la définition d'un facteur de forme adapté au problème de la vérification des signatures dans un contexte de faux aléatoires. L'évaluation du taux d'activité du signal dans le champ de vision de rétines positionnées uniformément dans l'image permet la mise en oeuvre d'une analyse locale de la signature et de tenir compte des variations locales du tracé. Des travaux supplémentaires sont nécessaires pour être en mesure de choisir dynamiquement le nombre et la position des points d'attention dans l'image, de même que la taille des rétines en fonction de la complexité de la signature à l'étude.

7. Références

- [1] F. Leclerc et R. Plamondon, "Automatic Signature Verification : The State of the Art -- 1989-1993", *International Journal of Pattern Recognition and Artificial Intelligence, Special Issue on Automatic Signature Verification*, 1994, pp. 3-20.
- [2] R. Plamondon et G. Lorette, "Automatic Signature Verification and Writer Identification -- The State of the Art", *Pattern Recognition*, Vol. 22, No. 2, 1989, pp. 119-128.
- [3] E. Locard, "Les Faux en Écriture et leur Expertise", Payot, Paris, 1959.
- [4] J. Serra, "Image Analysis and Mathematical Morphology", New York, Academic Press, 1982.
- [5] V. Anastassopoulos et A.N. Venetsanopoulos, "The Classification Properties of the Pecstrum and Its Use for Pattern Identification", *Circuits Systems Signal Process*, Vol. 10, No. 3, 1991, pp. 293-326.
- [6] P. Maragos, "Pattern Spectrum and Multiscale shape representation", *IEEE Trans. on PAMI*, Vol. 11, No. 7, July 1989, pp. 701-716.
- [7] R. Sabourin et G. Genest, "An Extended-Shadow-Code Based Approach for Off-Line Signature Verification: Part -I- Evaluation of The Bar Mask Definition", *12th ICPR*, Jerusalem, Israel, October 9-13, Vol II, 1994, pp. 450-453.
- [8] R. Sabourin et G. Genest, "An Extended-Shadow-Code Based Approach for Off-Line Signature Verification: Part -II- Evaluation of Several Multi-Classifer Combination Strategies", *Third IAPR Conf. on Document Analysis and Recognition*, August 14-16, 1995, Montréal, Canada, pp. 197-201.

Éditeur de schémas de composants électroniques dessinés sur ardoise électronique

KODJO Dave Aimé-Désiré & BRAULT Jean-Jules

École Polytechnique de Montréal

Email: kodjo@scribens.polymtl.ca & brault@scribens.polymtl.ca

Résumé

L'éditeur proposé est un système interactif et convivial centré sur une ardoise électronique (tablette qui numérise la trajectoire du crayon). Il matérialise le concept de papier/crayon pendant le design d'un circuit. Cet éditeur permet de dessiner des symboles électroniques (transistors, ampli opérationnel, diode, résistance, inductance, condensateur, masse) et de les éditer avec des commandes gestuelles (sélectionner, effacer, copier, déplacer, grouper, dégroupier, tourner). Il doit analyser les données échantillonnées et transmises par la tablette et les structurer selon une hiérarchie "schéma/composant électronique/tracé curviligne". Plusieurs approches de reconnaissance des formes sont comparées dont deux approches connexionnistes (le réseau perceptron multicouches et le réseau de fonctions à base radiale), une approche syntaxique (comparaison de chaînes) et une approche statistique (classification par plus proches voisins). Le prototype a été évalué sur une population de dix personnes. Cette évaluation a porté sur trois aspects essentiels de l'éditeur à savoir la reconnaissance des composants électroniques, la reconnaissance des commandes gestuelles et la justesse de leur exécution.

1. Introduction

Un bloc-notes électronique (BNE) est généralement composé d'un crayon et d'une tablette transparente sur laquelle est superposé un afficheur. Le déplacement du crayon (écriture) à la surface de la tablette laisse une trace curviligne qui matérialise le concept d'encre électronique propre à cet outil. En plus des facettes standards d'édition et de calculs, le BNE intègre le caractère naturel de la rédaction d'un document sur un bloc-notes papier en permettant l'écriture manuscrite et l'émission de commandes gestuelles. Étant léger, le BNE est transportable partout où il existe un besoin de communication avec un outil informatique (interrogation d'une banque de données, saisie ou transmission de documents...). Il se présente sous plusieurs formes; les tablettes (exclusivement un médium d'acquisition), les PDA ("personal digital assistant"), les convertibles (possède un clavier)... Aujourd'hui, la quasi-totalité des BNE possède un module de reconnaissance d'écriture et de symboles graphiques simples (parallélogrammes, ellipses).

Cette communication présente un outil d'édition de schéma de composants électroniques à l'aide de commandes gestuelles. L'éditeur comporte une interface conviviale qui traite, reconnaît les symboles électroniques et les

commandes gestuelles, et restaure le schéma.

2. Problématiques

On fait face à quatre problèmes lors de la réalisation des tâches de l'éditeur, à savoir:

- différencier continuellement les types de tracés curvilignes dessinés par l'utilisateur: les symboles électroniques, les traits de liaisons entre ces symboles et les commandes gestuelles;
- segmenter ou localiser les symboles électroniques du schéma;
- reconnaître les symboles électroniques et les commandes gestuelles;
- restaurer le schéma électronique.

L'architecture d'un schéma électronique est complexe à déterminer du fait de sa variabilité. En effet, le schéma électronique possède une structure libre, car il peut être représenté (spatialement) de différentes façons en conservant son sens. Par contre, le document de textes est découpé en paragraphes, phrases, mots et caractères. Cette absence de contrainte spatiale explicite constitue une difficulté majeure de segmentation du schéma en composants électroniques.

Les approches de résolution de ce problème rencontrées dans la littérature sont de deux ordres selon que la segmentation est directe ou indirectement réalisée. Parmi les approches directes, on distingue deux avenues: la segmentation automatique et la segmentation assistée par l'utilisateur. Lorsqu'elle est automatique, souvent des contraintes de dessin sont imposées [3,13,14] pour réduire la complexité des algorithmes en jeu et par coup diminuer les délais engendrés par la segmentation. Elle peut être assistée [14], auquel cas l'utilisateur indique lui-même les limites de chaque symbole électronique. L'autre approche consiste à ne pas réaliser une segmentation explicite ou directe. Dans ce type de procédé la structure et les symboles électroniques du schéma sont reconnus conjointement [6,10]. Les tracés curvilignes sont associés à partir d'heuristiques ou de règles. De ces associations plusieurs candidats sont proposés. Et, l'association la plus vraisemblable est retenue.

L'absence de structure spatiale du schéma contraste avec celle des symboles électroniques. Ce qui justifie en partie le choix d'approches structurelles de représentation pour décrire les symboles graphiques [1,3,7,10,13,14].

La restauration ou la production d'une copie propre du schéma manuscrit est un processus à deux volets, fortement influencé par les performances de la segmentation et de la reconnaissance des symboles électroniques. Elle consiste d'abord à analyser les interactions entre les symboles électroniques (traits de liaisons, noeuds) et ensuite à embellir le schéma.

3. Description fonctionnelle de l'éditeur

La fonction d'éditeur suppose une interaction conviviale et une exécution des commandes dans des délais raisonnables et conformément aux attentes de l'utilisateur. Conséquemment, les choix à effectuer concernant la résolution des problèmes soulevés ci-dessus doivent mettre en jeu des techniques simples au sens algorithmique. Par rapport à cette idée, le fonctionnement de l'éditeur a été subdivisé en quatre modes. Le mode saisir dédié au dessin de symboles électroniques, le mode commande pour le dessin de commandes gestuelles, le mode segmenter réservé à la segmentation manuelle réalisée par l'utilisateur en encerclant ou en encadrant chaque symbole électronique du schéma, et le mode restaurer pendant lequel l'éditeur analyse les liaisons entre les symboles électroniques afin d'en dégager d'abord une représentation sous la forme d'un graphe non-orienté puis de produire une copie propre du schéma. Il comprend tous les éléments standards des interfaces graphiques: barres de menu et de boutons, zone de dessin et barre d'informations (figure 3.1). Une session de dessin commence par l'activation du mode saisir pour esquisser un schéma à l'encre noire. Puis, on initie la reconnaissance des symboles électroniques en les segmentant à l'encre rouge en mode segmenter. Les modifications sont toujours possibles en mode commande et à l'encre verte.

3.1 Symboles électroniques

Les symboles électroniques permis par cet éditeur sont: l'amplificateur opérationnel, la diode, le condensateur, la masse, les transistors BJT, la résistance, l'inductance et les sources de tensions et de courants continus et alternatifs de la figure 3.2.

3.2 Commandes

L'éditeur intègre les commandes d'édition standards (copier, effacer, défaire, grouper, dégrouper, déplacer) et des utilitaires tels que la sauvegarde, la lecture, l'impression de document. Il possède des commandes plus spécifiques permettant de choisir la technique de reconnaissance désirée. Seules, les commandes les plus couramment utilisées pour éditer un document ont été doublées de commandes gestuelles afin d'apposer à l'éditeur la marque chère au concept papier/crayon du BNE.

Les symboles des commandes gestuelles (figure 3.3) sont simples d'exécution et naturels lorsqu'on édite un document. Les commandes gestuelles implantées comprennent la copie, le déplacement, le regroupement des plusieurs symboles, le dégroupement d'un groupe de symboles, l'effacement et les commandes tourner à + ou - 90 degrés.

4. Caractérisation et reconnaissance des symboles

Les tracés curvilignes (ensemble de points entre une descente du crayon et sa levée de la surface de la tablette) de chaque symbole sont prétraités afin de limiter l'incidence du bruit sur son profil. Puis, on extrait des caractéristiques globales mais suffisamment pertinentes pour distinguer les symboles entre eux.

4.1 Prétraitements

Les tracés curvilignes reçus par la tablette sont empreintes d'irrégularités ayant diverses origines. Les imperfections d'ordre matériel sont dues au bruit d'échantillonnage. Les sources les plus importantes proviennent de facteurs humains: tremblements de la main, glissement du crayon sur la tablette en verre. Pour réduire ces imperfections, les tracés curvilignes subissent respectivement un filtrage passe-bas, une suppression des crochets aux extrémités (dûs au glissement) et un filtrage spatial.

Le bruit de haute fréquence est atténué à l'aide d'un filtre passe-bas de type auto-regressif du troisième ordre à coefficients entiers. Ce filtre, de fréquence de coupure d'environ 10Hz et de déphasage nulle, est obtenu par la double convolution d'un peigne à trois impulsions de Dirac [4]. Il est simple, rapide et réalise une bonne réduction du bruit de haute fréquence comparée à un filtre gaussien (coefficients réels). Une illustration de ce filtrage est donnée à la figure 4.1. La suppression des crochets aux extrémités du tracé curviligne s'appuie sur une méthode proposée sous différentes formes par Loy[11] et Mandler[12]. Le résultat de la suppression des crochets de la figure 4.1 est illustré à la figure 4.2. L'objectif du filtrage spatial à pas égaux implanté est de réduire la quantité d'information à traiter. Ce filtrage est délicat car son efficacité dépend du pas choisi. Il obéit aux équations ci-dessous. Le résultat de cette opération est illustré à la figure 4-3.

4.2 Caractérisation des symboles

Cette opération se fait en plusieurs étapes:

- l'information contextuelle des modes de fonctionnement et les spécificités des tracés curvilignes (forme, nombre) sont utilisées pour réaliser une première classification des symboles (figure 4.4 et tableau 4.1);
- les tracés sont ensuite réduits en segments de droite ou élémentaires à partir de l'algorithme de Kato [8] qui fixe les points caractéristiques;
- les segments élémentaires sont classifiés en deux groupes: segments long et court;
- on détermine l'orientation globale des symboles électroniques à partir des segments longs; ce qui réduit le nombre de symboles électroniques à

reconnaître par un facteur d'environ deux;

- puis, la chaîne de segments longs est ordonnée de façon à stabiliser leur succession et ce quelque soit la séquence des tracés curvilignes de départ [9];
- on représente séparément les séquences de segments longs et courts par un codage de Freeman[5]. Des tests préliminaires ont permis de constater que le codage de Freeman (A) assure une bonne discrimination des symboles des classes I et III. Cet alphabet doit être augmenté (alphabet B) pour les classes II et IV (figure 4.5).

5. Reconnaissance des symboles

Quatre méthodes de reconnaissance de formes ont été implantées: - deux techniques connexionnistes: le perceptron multicouches (PMC) et le réseau de fonctions à base radiale (FBR) - une technique structurelle de type comparaison de chaînes (CC) - et une technique de classification statistique par minimum de distance (MD). La mise en oeuvre de ces algorithmes impose certaines contraintes sur la taille du vecteur caractérisant les symboles. Le tableau 5.1 indique la répartition des algorithmes et les ajustements effectués en fonction des quatre classes de symboles.

L'entraînement des réseaux de neurones a conduit pour les symboles de la classe I à: un réseau PMC avec une couche cachée de 4 neurones et un réseau FBR avec la couche cachée de 6 neurones. Les meilleurs réseaux obtenus pour les symboles de la classe III sont: un réseau PMC de 20 neurones sur une couche cachée et un réseau FBR de 28 neurones sur la couche cachée. Pour les symboles de la classe IV des commandes gestuelles ont obtenu les meilleurs résultats avec un réseau PMC à une couche cachée de 20 neurones et un réseau FBR de 36 neurones sur la couche cachée. En général, la taille du réseau FBR est plus grand que celui du réseau PMC pour un taux de reconnaissance donné.

6. Restauration du schéma

Elle peut être décrite de prime abord comme un processus purement esthétique. Toutefois, on se rend bien vite compte qu'elle intègre en plus un volet d'interprétation [2,14]. Ce qui sous-entend une transcription du schéma en un codage formel réalisé à partir de l'analyse des liaisons entre les symboles électroniques.

Si, les symboles électroniques du schéma sont correctement segmentés, les tracés curvilignes présents constituent normalement l'ensemble des lignes de liaison entre ces symboles. Lorsque la commande restaurer est lancée, l'éditeur analyse la proximité des lignes de liaisons par rapport aux symboles

électroniques d'une part, et d'autres entre les lignes de liaisons elles-mêmes. On dégage ainsi la structure du schéma sous la forme d'un graphe non-orienté où les noeuds et les arcs représentent respectivement les symboles électroniques et les liaisons (figure 6.1). Après la construction de la hiérarchie "schéma/composant électronique/ tracé curviligne", l'éditeur produit une copie propre du schéma. Dans sa version actuelle, l'éditeur ne réalise pas de normalisation de la taille ni de re-disposition spatiale des symboles électroniques.

7. Évaluations de l'éditeur

Les évaluations ont porté sur trois aspects qui singularisent l'éditeur de composants de circuits électroniques à savoir: - la reconnaissance des symboles électroniques et des commandes gestuelles - et l'exactitude de l'exécution des commandes gestuelles. Pour ce faire, on a réalisé ces tests sur une population de 10 personnes choisies parmi des personnes familières ou non au concept de papier/crayon. L'évaluation du taux de succès des algorithmes de reconnaissance s'est fait sur 5 exemples de chaque symboles électroniques et de chaque commande gestuelle. L'exactitude de l'exécution des commandes gestuelles a été mesurée à partir d'un protocole consistant à transformer le schéma d'un amplificateur logarithmique en un intégrateur.

Les symboles électroniques sont reconnus avec un taux globale de 86%, 88% et 87% respectivement avec les algorithmes PMC, FBR et CC, et 82% avec l'algorithme MD. Les principales erreurs observées proviennent de la reconnaissance des transistors et des sources variables où souvent les dimensions de la flèche ou de la sinusoïde ne permettent pas de distinguer efficacement les segments longs des segments courts.

On observe un taux global de reconnaissance des commandes gestuelles de 90% avec les algorithmes PMC et FBR. Ces résultats (figure 7.2) sont influencés par le style de dessin des commandes. Les erreurs les plus élevées proviennent des commandes copier (79%) et tourner à -90 degrés (84%) où la normalisation par rapport à la direction la plus fréquente crée parfois une confusion. Pour être efficace, l'histogramme doit également tenir compte de la densité des directions.

Un taux de reprises global de 4% avant l'exécution exacte des commandes a été obtenu (figure 7.3). Les erreurs observées sont essentiellement des erreurs de reconnaissance des commandes. L'erreur la plus importante (8%) remarquée avec la commande grouper découle de la difficulté de dessin de cette commande autour des symboles ciblés.

8. Recommandations

Elles émanent pour la plupart des suggestions émises par les usagers. À terme, il serait intéressant de remplacer la segmentation manuelle par une segmentation automatique. Car, dans un contexte plus dense en symboles électroniques, la convivialité pourrait être dégradée par les délais plus longs d'édition. L'amélioration des taux de reconnaissance des symboles électroniques (88%) et des commandes gestuelles (90%) s'avère primordiale. La restauration, dans la version actuelle de l'éditeur, est limitée à l'extraction de la structure du schéma sous la forme d'un graphe et au remplacement des tracés curvilignes de liaisons par des lignes droites. La finition de la restauration consisterait à restructurer la disposition spatiale des symboles électroniques à partir de l'analyse des noeuds du schéma. Plusieurs ajouts de fonctionnalités ont été demandés par les usagers à savoir: - copier et de tourner plusieurs objets - avoir plusieurs groupes d'objets - changer le mécanisme de la commande grouper - permettre la télécopie du schéma - augmenter les niveaux de la commande défaire - implanter la commande refaire - associer une dénomination aux symboles électroniques reconnus.

9. Conclusions

L'éditeur de composants électroniques sur ardoise décrit dans cette communication est fonctionnel et utilisable. Il constitue une contribution viable au développement d'outils centrés sur le concept papier/crayon. Des études sont en cours pour améliorer les performances des algorithmes de reconnaissance et pour parfaire la restauration du schéma électronique.

10. Références

- [1] H. Bunke, "Experience with several method for the analysis of schematic diagrams", Intl. Conf. on Pattern Recognition, p710, vol.2, 1982.
- [2] H. Bunke, "Attributed programmed graph grammars and their application to schematic diagram interpretation", I.E.E.E. Trans. on P.A.M.I. p.574, No.6, 1982.
- [3] R. Cederberg, "Automatic interpretation of hand-drawn electrical schematic diagrams", Intl. Conf. on CAD and Manufacture of electronic components circuits and systems p.118, 1979.
- [4] Jean-Pierre Drouhard, "Filtrage numérique auto-regressif: théorie, implantation et application de filtres à coefficients entiers", rapport de l'institut de génie Biomédical de l'école polytechnique de Montréal, 1976.
- [5] H. Freeman, A. Saghri, "Generalized chain codes for planar curves",
- [6] M. Hiroshi, W. Toru, "On-line hand-sketched figure recognition", Pattern Recognition, vol.19, No.2, pp.147-160, 1986.
- [7] R. Kasturi, "A system for recognition and description of graphics", Intl. Conf. on

- Pattern recognition, p.255, vol.1, 1988.
- [8] O. Kato, H. Iwase, "Interactive hand-drawn diagram input", Proc. on Pattern Recognition and Image Processing, pp.544-549, 1982.
 - [9] Dave A.D. Kodjo, "Conception et réalisation d'un éditeur de composants électroniques sur ardoise", Mémoire de l'École Polytechnique de Montréal, 1995.
 - [10] H. Kojima H. Toida, "On-line hand-drawn line-figure recognition and its application", Intl. Conf. on Pattern Recognition, p.1138, vol.2, 1988.
 - [11] W.W Loy, I.D. Landau, "An on-line procedure of handprinted alphanumeric characters", I.E.E.E. Trans. on P.A.M.I., vol.4, no.4, pp.422-427, 1982.
 - [12] E. Mandler, "Advanced preprocessing technique for on-line script recognition of non connected-symbols", Proceedings of the 3rd Int. Conf. Handwriting and computer applications, pp.64-66, july 1987.
 - [13] K.J. Peng, K. Onda, T. Tamamoto, Y. Aoki, "Plex grammar for predicate logic programming and its application to the structural analysis of electronic circuits" Systems and computer in Japan, vol. 19, no. 9, 1988.
 - [14] C. Lin Wen, J.H. Pun, "Machine recognition and plotting of hand-sketched line figures", IEEE Transactions on Syst, Man, and Cybernetics, vol. 8, no.1, p.52, 1978.

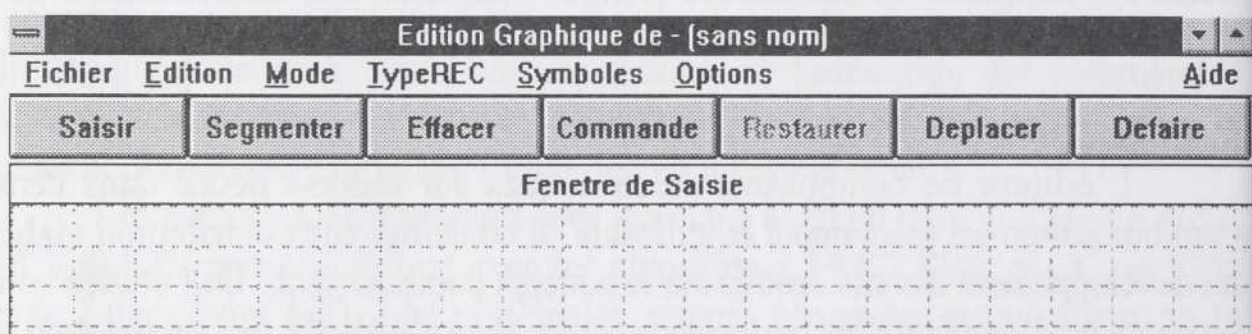


Figure 3.1 Interface graphique de l'éditeur

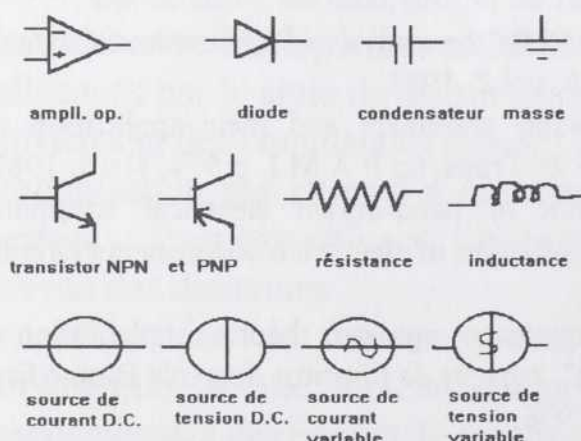


Figure 3.2 Symboles électroniques

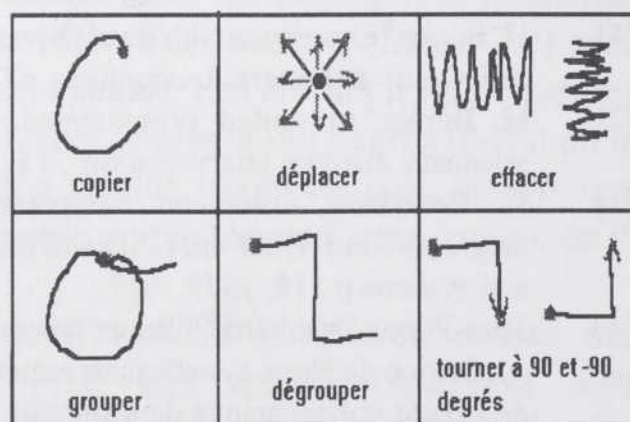


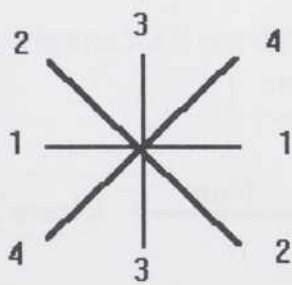
Figure 3.3 Commandes gestuelles

Tableau 4.1 Répartition des symboles en classes

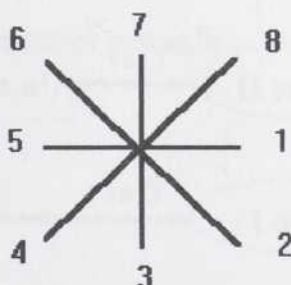
Classe	Symboles	critères de regroupement
I	-source de tension continue -source de tension variable -source de courant continu -source de courant variable	symboles contenant un cercle
II	-inductance -résistance	symboles dessinés en un seul tracé curviligne
III	-ampli. op. -condensateur -diode -masse -transistor NPN - transistor PNP	symboles ne contenant pas de cercle et dont le dessin nécessite plus d'un tracé curviligne
IV	-copier -effacer-tourner -grouper -dégrouper	les commandes gestuelles dessinées en un seul tracé curviligne

Tableau 5.1 Caractérisation des symboles en fonction des quatre classes

	PMC	FBR	CC	MD
Classe I	Décompte de l'occurrence des directions {1 2 3 4 13 31 }		comparaison des chaînes	
Classe II	histogramme des directions {1,2,3,4,5,6,7,8}			
Classe III	Décompte de l'occurrence des directions {1 2 3 4 13 31 34 32 42 24 23 43}		comparaison des chaînes	
Classe IV				histogramme



Alphabet A={1,2,3,4}



Alphabet B={1,2,3,4,5,6,7,8}

Figure 4.5 Codages de Freeman

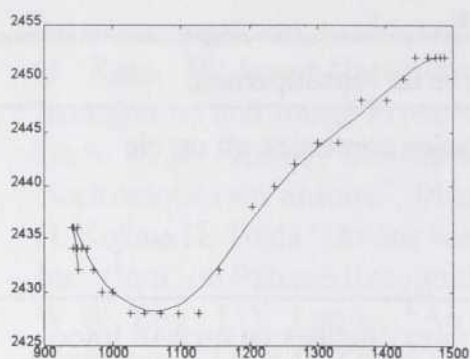


Figure 4.1 filtrage passe-bas

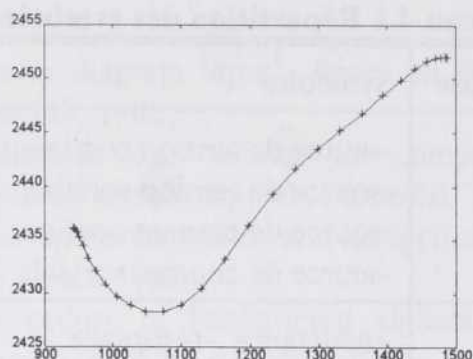


Figure 4.2 Suppression de crochets

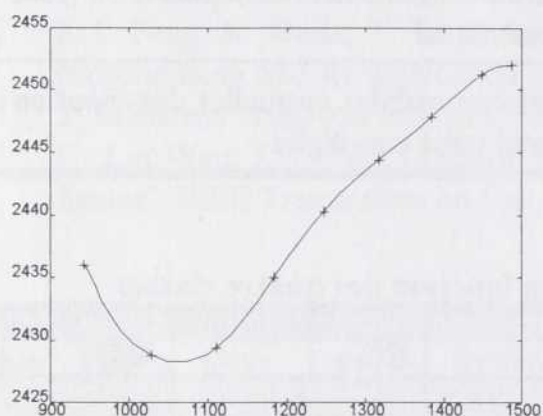


Figure 4.3 Filtrage spatial à pas égaux

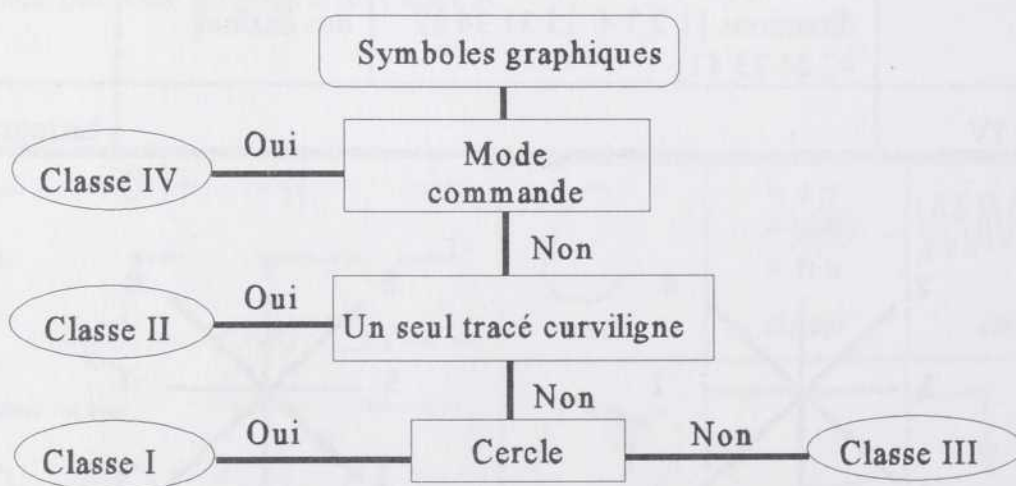


Figure 4.4 Classification des symboles

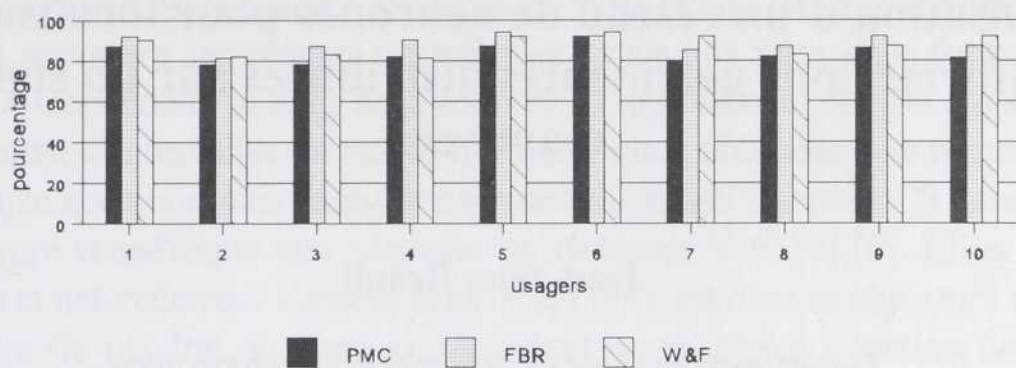


Figure 7.1 Taux de reconnaissance des symboles électroniques

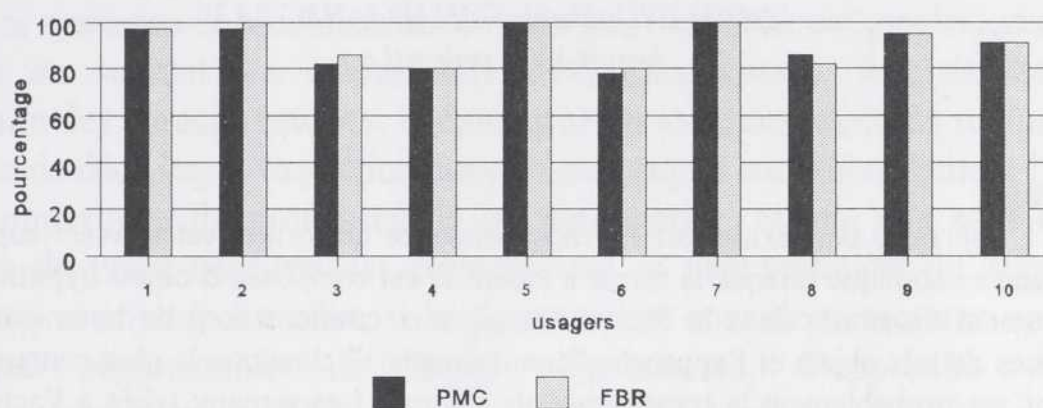


Figure 7.2 Taux de reconnaissance des commandes gestuelles

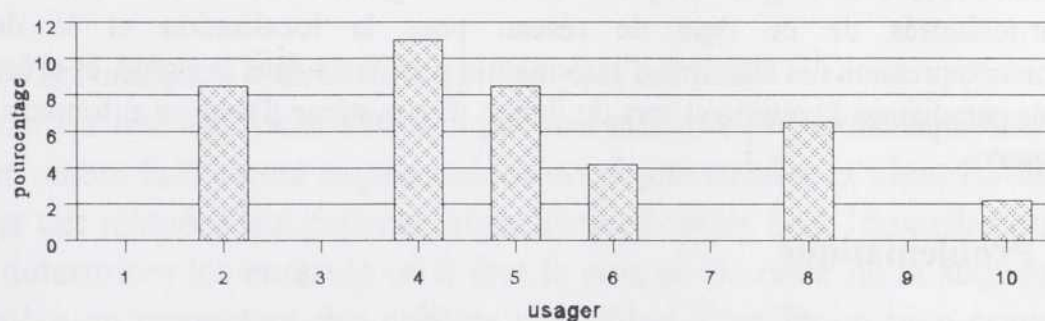


Figure 7.3 Taux de reprises des commandes gestuelles

Utilisation d'un réseau de neurones pour localiser les informations pertinentes distribuées sur un signal manuscrit

Jean-Jules Brault

*Département de génie électrique et génie informatique
École Polytechnique de Montréal
C.P. 6079, succ. centre-ville
Montréal (Québec), CANADA, H3C 3A7
brault@vlsi.polymtl.ca*

Résumé

L'approche connexionniste en reconnaissance de formes est souvent supérieure à l'approche symbolique lorsque la forme à identifier est composée d'objets hypothétiques et massivement distribués dans la forme. Les signaux continus sont de bons exemples de séquences de tels objets et l'approche "connexionniste" classique la plus connue pour les analyser est probablement la transformée de Fourier. Les signaux reliés à l'acte d'écrire devrait nécessiter le même genre d'approche. Nous proposons d'utiliser un réseau de neurones à délais adaptables (dérivé des travaux de Tank et Hopfield sur le traitement de la parole) pour localiser automatiquement les informations distribuées dans les signatures manuscrites. À l'aide de séquences synthétiques et de signatures réelles, nous faisons ressortir les particularités de ce type de réseau pour la localisation et le degré de dilatation/compression des éléments d'information distribués dans le signal, et présentons les différents paradigmes à considérer lors du design d'un système d'analyse autonome basée sur ces réseaux.

1 Problématique

1.1 Contexte

Il existe de nombreuses situations où il est nécessaire de vérifier l'identité d'une personne. Les méthodes utilisées peuvent être classées sommairement en deux catégories. La première classe est basée sur ce que la personne possède (clé, mot de passe, empreintes digitales), et la seconde sur ce qu'elle peut produire (voix, signature). Selon plusieurs études, l'analyse des signaux manuscrits provenant de l'exécution d'une signature est potentiellement une des meilleures méthodes pour authentifier automatiquement l'identité d'une personne [11], [15], [8].

1.2 *Taxinomie des systèmes existants en vérification des signatures (VdS)*

La signature se présente normalement sous nos yeux sous forme d'une image 2D. Cependant, si nous échantillonnons la pointe du crayon qui exécute une signature à intervalles de temps réguliers, nous obtenons une séquence de données que nous pouvons analyser comme tout signal temporel. Il existe dans la littérature scientifique une panoplie de systèmes VdS [9],[5]. Elles ont été analysées et naturellement classées selon que l'on considère la signature comme une image (le résultat de l'action de signer) ou un signal (l'action de signer comme tel). La signature sera considérée dans ce qui suit comme une séquence de coordonnées (x_t, y_t) en fonction du temps.

Les systèmes de reconnaissance de formes transforment généralement les données en suivant les étapes suivantes: l'acquisition, le prétraitement, l'extraction des caractéristiques, la comparaison avec des modèles mémorisés, et la prise de décision. L'extraction des caractéristiques est probablement l'étape la plus importante du processus de reconnaissance, et une des sous-étapes cruciales de cette dernière est certainement la segmentation du signal en segments (ou objets). En effet, si la signature est considérée comme un objet composite segmentable de façon déterministe, l'approche devrait fondamentalement être symbolique. Si elle est considérée segmentable mais de façon non déterministe, l'approche pourrait être symbolique mais aussi connexionniste, tout dépendant du degré d'incertitude associé à la localisation des "objets" dans la séquence.

1.3 *Les lacunes des approches symboliques en VdS*

Pour qu'une approche symbolique soit efficace [7], il faut que les sujets d'analyse soient facilement segmentables en objets stables et identifiables afin de trouver des relations (ou patterns) significatives entre eux. Identifier un objet signifie déterminer les endroits où il faut le couper du reste de la séquence, et cela s'opère en respectant des critères prédéfinis. Une façon bien connue, et générique, de segmenter une séquence de type signal est de comparer la valeur de chaque point de ce signal à un seuil quelconque. Les portions inférieures (ou supérieures) au seuil détermineraient les objets importants à analyser symboliquement. Cette façon de segmenter soulève une question importante. Comment déterminer la valeur de ce seuil critique ?

On pense que des connaissances sur la nature même de l'acte d'écrire [10], de signer ou d'imiter [1],[2] devraient normalement être utiles pour identifier ces éléments. Par exemple, on pourrait segmenter le signal aux endroits où le tracé passe par un *maximum de courbure* du tracé et/ou un *minimum de vitesse*. Cependant, en se basant sur d'autres principes, on pourrait tout aussi

bien utiliser les points d'inflexion sur le tracé, lesquels sont situés près des *minima de courbure*! De toutes façons, le fait de segmenter selon des critères locaux et souvent arbitraires conduit à une explosion combinatoire de séquences d'objets hypothétiques et une approche symbolique de type structurelle (ou syntaxique) souvent très complexe en résulte.

1.4 Approche connexionniste ou distribuée

Dans le contexte de la vérification d'identité, une information utile dans un signal devrait plutôt être, d'une part, tout élément de signal statistiquement présent dans les signatures d'une même personne, et d'autre part, statistiquement absent dans les signatures des autres personnes. L'approche automatique idéale devrait laisser le système déterminer lui-même ce qui est important pour lui en tentant de maximiser ces deux qualités que sont la stabilité et la discrimination des éléments du signal.

Pour qu'un système puisse localiser statistiquement les éléments d'information composant une classe de signaux et distribués dans ceux-ci, on suggère l'approche suivante: dégénérer le signal en le décomposant en signaux plus simples (application du théorème de Cover [3] sur la séparabilité des classes) et appliquer une méthode de régularisation pour stabiliser l'identification distribuée d'éléments d'information [4].

Pour une séquence donnée, si un signal est dégénéré en N signaux $D_n(t)$, valant "1" pour identifier la position potentielle d'un objet (ou élément d'information), et valant "-1" autrement (pour identifier une information dite négative), alors ce signal peut maintenant s'exprimer ainsi:

$$V(t) = \sum_{n=1}^N \left(\sum_{m_n=1}^{M_N} D_n(t) \cdot \delta(t - m_n \tau_B) \right), \quad (1)$$

où:

τ_B = délai de base

$m_n \tau_B$ = position temporelle de l'objet D_n

Pour une classe de séquences légèrement différentes, les valeurs des délais $m_n \tau_B$ ne seront probablement pas les mêmes. Il faut donc remplacer la fonction déterministe $\delta(t - m_n \tau_B)$ par une fonction de régularisation $G_m(t - m_n \tau_B)$, et faire apprendre au système connexionniste les M_N délais $m_n \tau_B$.

2 Réseau de concentration temporelle proposé par Tank&Hopfield

La figure 1 illustre l'architecture générale du réseau proposé par Tank et Hopfield [13], [14] pour l'analyse de la parole. La description sommaire faite ici ne fait référence à aucun type particulier de détecteur. Le signal à analyser est présenté à la gauche et passe à travers un ensemble de N détecteurs D_n . Les signaux $D_n(t)$ passent par des unités de délais flous, à réponse impulsionnelle gaussienne pour produire les signaux $x_n(m, t)$. Ces délais permettent de tenir compte des erreurs de positionnement des objets en fonction de la distance moyenne qui les sépare d'un point dit "certain" (comme le début ou la fin d'une séquence).

$$x_n(m, t) = \int_0^{\infty} G(m, t') D_n(t-t') dt', \quad \text{où:} \quad \begin{aligned} G(m, t') &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(t'-m\tau_B)^2} \\ \sigma &= \sigma_0 (m-c)^\alpha \\ \alpha &= \text{constante à déterminer} \\ c &= \text{position du point "certain"} \\ \infty &= t' \text{ tel que } G(m, t') \approx 0 \end{aligned} \quad (3)$$

Tous ces signaux sont par la suite pondérés par des poids synaptiques ω_{nm} et sommés à l'entrée d'une unité de reconnaissance (un neurone). Cette somme est filtrée par un passe-bas ayant τ_{rec} comme constante de temps produisant finalement un signal $V(t)$. Étant donné que nous avons affaire à des équations linéaires, nous pouvons exprimer $V(t)$ de la façon suivante.

$$V(t) = \left(\int_0^{\infty} q e^{-\frac{t'}{\tau_{rec}}} dt' \right) + \left(\sum_{n=1}^N \sum_{m=1}^M \omega_{nm} X_n(m, t) \right) \quad (4)$$

où:

$$X_n(m, t) = \int_0^{\infty} x_n(m, t-t') e^{-\frac{t'}{\tau_{rec}}} dt' \quad (5)$$

et où " q " est un seuil adaptable qui représente le niveau d'activité à l'entrée du neurone pour la classe de signal à reconnaître. La décision finale $y(t)$ sera donnée par la fonction logistique (non linéaire):

$$y(t) = \tanh(V(t)/2) \quad (6)$$

L'apprentissage permet de déterminer statistiquement la présence et l'importance de divers objets distribués dans le signal à différents moments. Les ajustement des poids et du seuil global sont donnés par les relations bien

connues de descente du gradient de l'erreur:

$$\Delta \omega_{n,m} = \varepsilon_1 \sum_{ex=1}^{N_{EX}} \sum_{t_{pa}=1}^{N_{PA}} [d^{ex}(t_{pa}) - y^{ex}(t_{pa})] X_n^{ex}(m, t_{pa}) \quad (7)$$

$$\Delta q = \varepsilon_2 \sum_{ex=1}^{N_{EX}} \sum_{t_{pa}=1}^{N_{PA}} [d^{ex}(t_{pa}) - y^{ex}(t_{pa})] \quad (8)$$

où N_{EX} est le nombre de signaux d'entraînement, N_{PA} est le nombre de points d'apprentissage situés le long du signal, et $d^{ex}(t_{pa})$ est la valeur désirée de la sortie d'un des signaux à un des N_{PA} points d'apprentissage. La convergence est assurée (en autant évidemment que ε_1 et ε_2 soient "petits") car la courbe d'erreur ne comporte pas de minima locaux [13].

3 Adaptation du réseau de Tank&Hopfield à la VdS

3.1 Choix des détecteurs

Le choix des détecteurs D_n est évidemment fortement fonction de l'application considérée. Le réseau de Tank et Hopfield a originellement été proposé pour traiter un signal vocal. Or, bien que la parole est un phénomène qui n'a pas le même contenu spectral qu'un signal manuscrit (contenu fréquentiel compris entre 0 et 15 Hz avec un maximum d'énergie autour de 5 Hz), une décomposition de celui-ci en fonction de la fréquence semble être une bonne approche à essayer.

3.2 Décomposition fréquentielle

La décomposition d'un signal $x(t)$ en fonction d'un spectre de détecteurs sinusoïdaux de différentes fréquences ω_n est bien connue des ingénieurs et des physiciens. La transformée discrète de Fourier (TDF) est calculée par:

$$X_n = \sum_{l=0}^{L-1} x(l) e^{-i l \omega_n}, \quad \text{où } \omega_n = \frac{2\pi n}{N} \quad (9)$$

où X_n donne la plus ou moins grande présence de chacune des fréquences ω_n (où $N=L/f_{éch}$) dans la fenêtre de largeur L . Si L est inférieure à la longueur du signal, alors en glissant la fenêtre sur tout le signal, on pourrait utiliser la variation des X_n en fonction du temps comme information à analyser.

Malheureusement, la résolution fréquentielle de la TDF est inversement fonction de la largeur de la fenêtre et comme la durée d'une signature typique dans sa totalité est relativement court, on risque de manquer de résolution.

3.3 *Décomposition en déplacement tangentiel*

Il n'est peut être pas nécessaire d'effectuer une décomposition du signal en fonction d'une base orthogonale (comme la TDF) car le principe de régularisation sous-jacent au réseau décrit ici implique l'utilisation d'une certaine redondance. Une approche multi-seuils de segmentation d'un signal se prête bien à ce réseau.

Certaines études corroborent l'idée que la vitesse tangentielle du tracé est un signal physiologiquement significatif du mouvement manuscrit [8],[10]. De plus, son évaluation à un point donné du signal ne demande pas une plage très grande autour de celui-ci (5 points suffisent dans notre cas).

3.4 *Paramètres du réseau à fixer*

Les équations 2 à 6 nécessitent la détermination de 10 variables, soit; le nombre de délais (M_N) et les variables associées (σ_0 , α , τ_B), le nombre de détecteurs (N), le nombre de neurones et les variables associées (τ_{REC} , ϵ_1 , ϵ_2) et le nombre de points d'apprentissage (N_{PA}). Nous avons conçu une série d'expériences exploratoires afin de nous orienter dans la détermination de ceux-ci.

4 **Expérimentation et discussion**

Trois séries d'essais ont été réalisées à l'aide d'un logiciel interactif implanté sur PC [12]. La figure 2 montre une des multiples fenêtres du logiciel. Les deux premières séries ont servi à déterminer la valeur des variables mentionnées ci-haut alors que la troisième a servi à valider ces variables.

4.1 *Signaux synthétiques et signatures réelles déformées*

Dans la première série d'expériences, nous avons entraîné un réseau, possédant un seul détecteur, avec des signaux synthétiques bipolaires (i.e. pouvant prendre que deux valeurs afin de déterminer les variables temporels du réseau. Puisque seule la position des objets dans le temps pouvait varier, cela nous a permis de limiter grandement l'espace des possibilités que peuvent prendre certaines des variables mentionnées ci-haut. Dans la deuxième série, nous avons utilisé deux signatures réelles pour tenir compte de la richesse des variations du mouvement humain. Néanmoins, nous tenions à limiter les artéfacts impondérables. Les signatures des deux personnes devaient être relativement semblables en durée et ne pas comporter d'hésitation. Nous avons également créé artificiellement des variations douces de chacune d'elles (en les étirant ou compressant) pour simuler un étalement intra-classe des objets d'une séquence.

Sachant que f_{ech} est la fréquence d'échantillonnage, f_{max} la fréquence maximale du signal et T , la durée en seconde de la signature, nous avons observé les faits suivants.

Les délais. La valeur du délai de base τ_B devrait être inférieure à $1/f_{max}$. Le nombre de délais devrait être supérieur à $T * f_{max}$. L'écart-type σ_0 devrait à peu près être égal à $\tau_B / (T * f_{max})^\alpha$ et la valeur de α devrait être inférieure à .5. Notez que si σ est petit et N_{EX} est grand, alors le réseau pourra déterminer expérimentalement la fonction de régularisation de chaque objet, et ce, pour chacun des niveaux D_n .

Les détecteurs. Le nombre de détecteurs doit être le plus grand possible, mais attention: le nombre d'exemples d'entraînement à fournir au système dépend entre autre du nombre de synapses à entraîner (i.e. $Nbr(neurones) * Nbr(détecteurs) * Nbr(délais)$).

Les neurones. Il doit évidemment y avoir au moins un neurone par classe de signatures. Si la signature d'une personne est très variable en durée, le système généralise trop et discrimine mal. Deux neurones pourraient donc appartenir à une même classe afin de diviser le travail de reconnaissance. La valeur de τ_{rec} doit être environ égale à τ_B . Le nombre N_{PA} devrait être supérieur au nombre de délais. Le nombre d'exemples N_{EX} dépend quant à lui très fortement de la variance intra-classe permise. Une façon de minimiser cette variance est de normaliser celle-ci en fonction de la durée moyenne des signatures d'un signataire. Il s'agit de ré-échantillonner les signatures à une fréquence telle que leurs durées totales soient identiques. Les valeurs de ϵ_1 et ϵ_2 devraient être assez petite pour assurer une descente du gradient... graduelle, soit $\epsilon_1 \leq (0.01 / N_{PA} * N_{EX})$, et $\epsilon_2 \leq (0.1 * \epsilon_1)$.

4.2 Deux classes de signatures réelles et leurs imitations

Dans la troisième série d'essais, nous avons utilisé deux classes de signatures réelles avec des imitations dites spécialisées. Cette expérience très sommaire de validation nous a permis de corroborer les hypothèses formulées à la suite des deux premières séries d'essais. Nous avons obtenu aucune erreur. Il ne faut cependant ne pas surestimer le réseau. Ceci n'est pas une expérience de validation très complète mais nous croyons être sur une voix intéressante.

4.3 Recommandations supplémentaires pour un système complet de VdS

Idéalement, et comme pour tout système statistique, on devrait avoir

beaucoup d'exemplaires par personne. On pourrait ainsi préclasser les signatures d'une même personne à partir de leur durée totale et entraîner autant de neurones qu'il y a de classes. Le nombre de classes serait évidemment déterminé en fonction du degré de dilatation/compression permis relié aux variables α et σ_0 du réseau. Malheureusement, il n'est pas aussi facile de ramasser des centaines d'échantillons d'une même signature et encore moins d'imitations spécialisées pourtant si essentielles [1]. Il est donc important d'être capable de se "fabriquer" une banque de signatures plausibles, de même que des faux "utiles" pouvant aider le réseau à ne pas généraliser idiotement.

Il serait profitable également d'ajouter d'autres types de détecteurs, en parallèle avec la vitesse (comme la direction du mouvement et la fréquence instantannée), afin d'augmenter le pouvoir discriminant du système.

5 Conclusion

Le but principal de ce papier était de montrer comment palier au problème du manque de stabilité d'une segmentation basée sur un seul seuil en divisant sa tâche parmi plusieurs seuils. Évidemment, d'autres types d'approches connexionnistes existent pour analyser des signaux spatio-temporels, tel le populaire TDNN. On peut consulter certains articles de revues [5], [6] ou livre [4] pour avoir des exemples d'application. Notons simplement que le principe de fonctionnement du TDNN est associé à celui du perceptron multi-couches (représentation globale), alors que celui présenté ici est plutôt du type fonctions à bases radiales (représentation locale).

La programmation dynamique est probablement une des meilleures approches pour comparer deux séquences particulières. Néanmoins, pour connaître la classe de signatures à laquelle une signature particulière appartient, il faut supposer que les paramètres de cette classe aient été déterminés statistiquement. Nous sommes sur le point de terminer des travaux dans cette direction afin d'établir une technique de comparaison de signatures qui exploite ni le principe de l'augmentation de dimensionnalité, ni celui de la régularisation. Nous serons alors en mesure de comparer les limites, les avantages et les inconvénients de chaque approche.

6 Remerciements

J'aimerais remercier Marc Lalonde et André Spilmann pour avoir programmé le réseau, et le CRSNG (OGP0089742) pour son support financier.

7 Références

- [1] J.-J. Brault, *Proposition et Vérification d'un Coefficient de Difficulté d'Imitation des Signatures Manuscrites*, Thèse de Doctorat, École Polytechnique, Montréal, 1988.
- [2] J.-J. Brault et R. Plamondon, "A Complexity Measure of Handwritten Curves: Modeling of Dynamic Signature Forgery", *IEEE Trans. SMC*, Vol. 23, No. 2, march/april, 1993, pp. 400-413.
- [3] T.M. Cover, "Geometrical and Statistical Properties of Systems of Linear Inequalities with Applications in Pattern Recognition", *IEEE Trans. Electronic Computers*, vol.14, 1965, pp. 326-334.
- [4] S. Haykin, *Neural Networks: A Comprehensive Foundation*, New York, NY: Macmillan College Publishing Compagny, 1994.
- [5] F. Leclerc et R. Plamondon, "Automatic Signature Verification: The State of the Art-1989-1993", *Int. Journal of Pattern Recognition and Artificial Intelligence*, World Scientific Publishing, vol.8, no.3, 1994, pp. 643-660.
- [6] R.P. Lippman, "Review of Neural Networks for Speech Recognition", *Neural Computation*, vol.1, 1989, pp. 1-38.
- [7] A. Newell, "The Serial Imperative", dans P. Baumgartner and S. Payr (eds.), *Speaking Minds*, Princeton, NJ:Princeton University Press, 1995.
- [8] R. Plamondon and M. Parizeau, "Signature Verification from Position, Velocity and Acceleration Signals: A Comparative Study", in *Proc. 9th Int. Conf. On Pattern Recognition*, 1988.
- [9] R. Plamondon and G. Lorette, "Automatic Signature Verification and Writer Identification-The State of the Art", *Pattern Recognition*, vol.22, no.2, 1989, pp. 107-131.
- [10] R. Plamondon, "A Kinematic Theory of Rapid Human Movements", *Biological Cybernetics*, vol.72, 1995, pp. 295-307.
- [11] D.E. Raphaël and J.R. Young, "Automated Personal Identification", *Long Range Planning Service*, Standford Research Institute, Menlo Park, CA, 1974.
- [12] A. Spilmann, *Vérification de l'Authenticité d'une Signature à l'Aide d'un Réseau de Hopfield et Tank*, Projet de fin d'étude, École Polytechnique, Montréal, 1995.
- [13] D.W. Tank and J.J. Hopfield, "Concentrating Information in Time: Analog Neural Networks with Applications to Speech Recognition Problems", in *Proc. IEEE First Int. Conf. Neural Networks*, San Diego, 1987, pp. 455-468.
- [14] K.P. Unnikrishnan, J.J. Hopfield D.W. Tank, "Connected-Digit Speaker-Dependent Speech Recognition Using a Neural Network with Time-Delayed Connections", *IEEE Trans. On Signal Processing*, vol.39, no.3, mars 1991, pp. 698-713.
- [15] T.K. Worthington, T.J. Chainer, J.D. Williford, and S.C. Gundersen, "IBM Dynamic Signature Verification", in *Computer Security, IFIP 1985*, J.B. Grimson and H.J. Kugler (Eds.), Elsevier Science Publ., 1985, pp. 129-154.

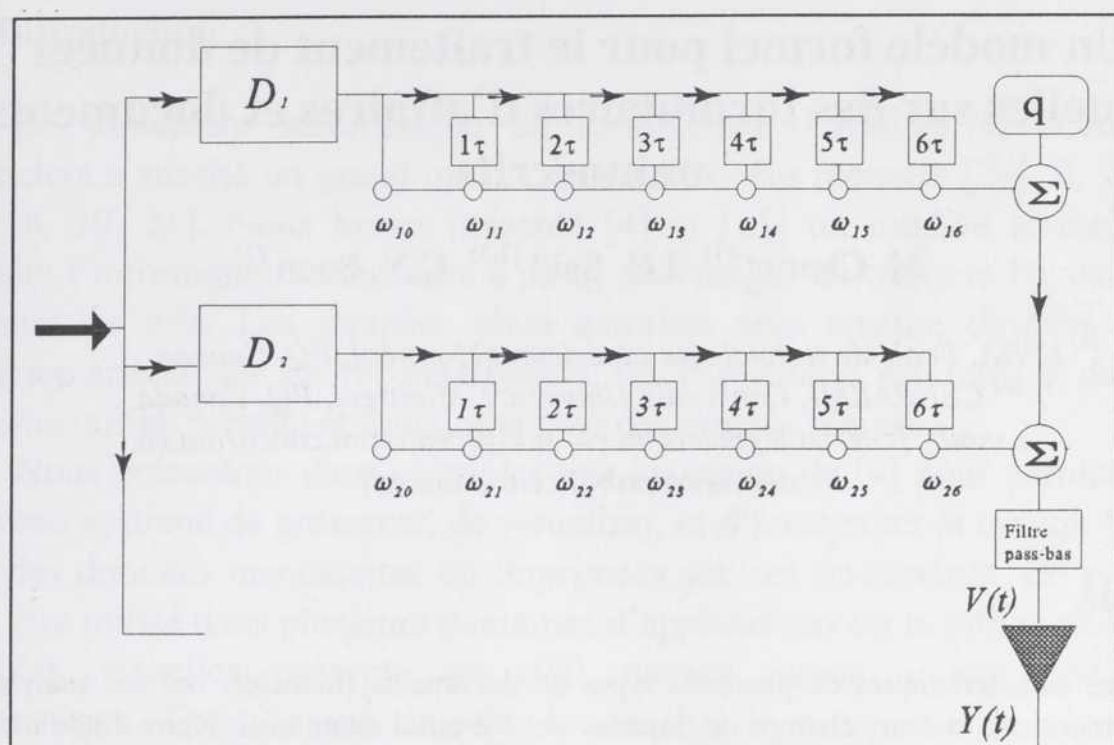


Figure1: Architecture du réseau de Tank et Hopfield

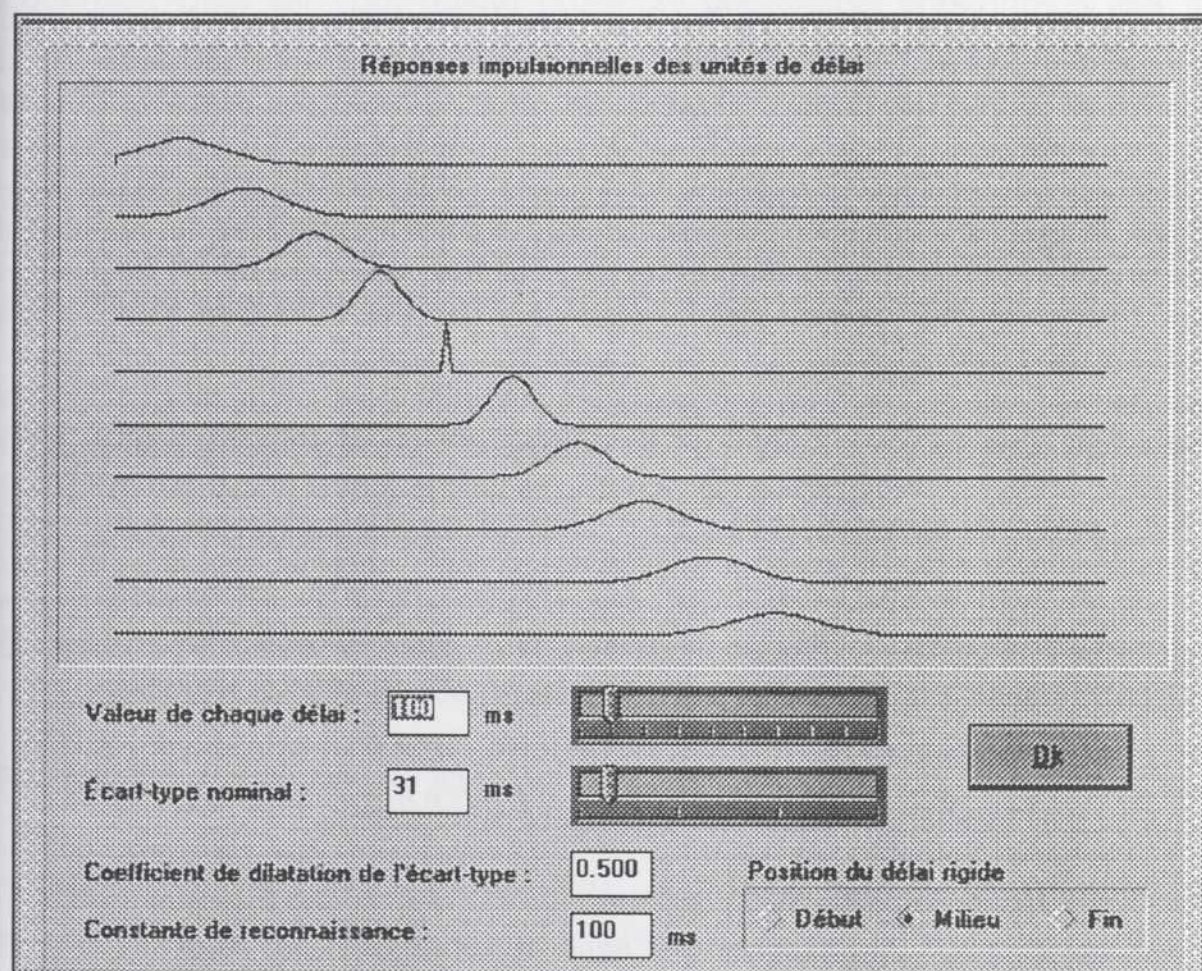


Figure2: Fenêtre montrant 10 fonctions de régularisation $G_m(t-m_n\tau_B)$.

Un modèle formel pour le traitement de données visuelles sur des formulaires d'affaires et documents manuscrits

M. Cheriet ⁽¹⁾, J.N. Said ^(1,2), C.Y. Suen ⁽²⁾

⁽¹⁾ LIVIA, École de technologie supérieure, Montréal, PQ, Canada

⁽²⁾ CENPARMI, Concordia University, Montréal, PQ, Canada

e-mail: {cheriet@gpa.etsmtl.ca, said@cenparmi.concordia.ca, suen@cenparmi.concordia.ca}

Résumé

Les caractéristiques de plusieurs types de documents financiers ont été analysées; leurs dimensions et leurs champs de données ont été aussi examinés. Notre étude indique que les chèques bancaires et les formulaires de retour de paiement sont les types de documents les plus fréquemment utilisés dans la vie de tous les jours (une grande quantité, à savoir des millions de ces documents sont utilisés et sont tous sujets de traitement par ordinateur). Le défi de trouver une solution pour lire ces documents d'une manière automatique à l'aide de l'ordinateur, nous a motivés d'orienter notre recherche sur la reconnaissance optique des données visuelles, en prenant les chèques et les bulletins de paye comme application.

Pour extraire les données visuelles, nous avons développé un modèle formel pour traiter des images de chèques en niveaux de gris. Nous avons introduit un algorithme efficace et rapide pour localiser et éliminer les lignes de base et extraire ensuite les données requises à partir du fond. Celui-ci peut contenir des images de scènes, des fleurs et des motifs très variés imprimés sur les chèques, rendant la tâche de traitement plus complexe. En outre, la morphologie mathématique est utilisée pour restaurer l'information perdue lors de l'extraction. Les parties des lignes de base faisant partie des données, perdues après l'élimination du fond et des lignes de base à partir des images de niveaux de gris, sont alors recouvertes en tout ou en partie. Les parties d'information non restaurées par le traitement morphologique, sont récupérées par un autre processus complémentaire faisant appel à des règles de topologie de l'écriture. Nous proposons dans cet article une structure nouvelle et dynamique puisant plusieurs connaissances dans le domaine de traitement d'images et d'intelligence artificielle qui permet de rendre le système robuste, efficace et productif.

Mots-clés: traitement et analyse de documents, images en niveaux de gris, segmentation d'images, morphologie mathématique, reconnaissance de caractères.

1. Introduction

Le traitement automatique de formulaires d'affaires et documents financiers a suscité un grand intérêt dans les années récentes [2-4, 9, 11, 13, 17, 18, 19, 21]. Nous avons présenté [4] et [11] un modèle formel pour extraire l'information manuscrite à partir des images de chèques bancaires en niveaux de gris. Les données ainsi extraites sont ensuite dirigées à des modules spécialisés de reconnaissance, où l'ensemble est intégré dans un système unifié, robuste et fonctionnel sur des données réelles.

Nous présentons dans ce papier une extension de [4] pour permettre au nouveau système de préserver, de visualiser, et d'interpréter la texture la plus fine des données manuscrites ou imprimées sur ces documents. Ce système peut être utilisé dans plusieurs domaines d'applications où la préservation des données visuelles présente un défi majeur quant à leur traitement automatique. Dans le cas de tels documents, le processus de séparer et d'éliminer l'information du fond de l'image à partir des objets d'intérêt tout en préservant la topologie de ceux-ci ne constitue pas une tâche facile à modéliser. Ceci est vrai, quand plusieurs documents contiennent des fonds variés formés des objets à différentes échelles de niveaux de gris, rendant ainsi la tâche de segmentation de la scène de l'écriture plus ardue, et où les méthodes de seuillage globales restent inefficaces [4,11]. La préservation des propriétés topologiques de l'information d'intérêt permettrait d'augmenter la productivité du système et possède un impact direct sur les résultats de reconnaissance obtenus aux étages suivants du système [4].

Seuiller une image de niveau de gris à l'aide d'une valeur t mène à convertir celle-ci en mode binaire. Les deux niveaux, ou plus précisément les deux classes, représentent les objets et le fond. Le processus de seuillage permet de classer les pixels de valeur supérieure au seuil t dans une catégorie, et le reste des pixels dans une autre catégorie. La valeur du seuil t est globale si la même valeur critique t est utilisée dans toute l'image, comme dans [4]. Plusieurs techniques ont été développées dans cet esprit pour déterminer la valeur du seuil d'une manière automatique [15, 1, 6, 8, 10, 14, 7, 16, 20]. Certaines méthodes utilisent simplement des histogrammes de densité, quant à d'autres utilisent l'information contextuelle telle que l'occurrence des niveaux de gris dans des pixels voisins. En dépit du bénéfice accru à cause de l'utilisation des méthodes contextuelles qui permettent au seuil t de varier à travers l'image, les méthodes à base d'histogrammes demeurent les plus généralement utilisées. Elles sont simples à interpréter et à implanter, et sont

plus rapides une fois l'histogramme calculé. De plus, dans cette classe restreinte de méthodes de seuillage, une panoplie de techniques a été proposée dans la dernière décennie. Lee, Chung et Park [12] ont comparé trois méthodes d'histogrammes avec deux autres méthodes contextuelles.

Dans ce papier, nous présentons une nouvelle approche locale à l'opposé de l'approche globale présentée dans [4], pour avoir un système intelligent capable de modéliser la perception humaine et d'interpréter les images de documents. La section 2 donne un aperçu sur la méthode récursive de seuillage à l'aide de la méthode étendue de Otsu que nous appliquons localement à chaque région d'intérêt. La section 3 présente le modèle formel en bref pour accroître la productivité du système. Plusieurs étapes sont nécessaires : (i) amélioration d'images (filtrage moyen), (ii) segmentation d'images (élimination du fond), (iii) détection d'objets (détection et élimination des lignes de base), (iv) traitement morphologique (opération de fermeture) et restauration de l'information perdue, et finalement v) segmentation et extraction des données requises. La section 4 présente les résultats de la nouvelle approche et conclut avec les directions futures de ce travail.

2. Seuillage dynamique par l'approche étendue de Otsu

Nous avons proposé dans [4] une extension à la méthode de seuillage de Otsu [13], qui est désignée dans la littérature parmi les méthodes de seuillage globales les plus efficaces [15]. En effet, dans la sélection du pic de segmentation, la contribution d'une petite région peut être masquée par d'autres régions plus larges. Une solution à ce problème est d'appliquer la procédure de seuillage d'une manière récursive. Les images seuillées sont considérées comme de nouvelles images, et la procédure de sélection du pic et du seuillage se poursuit jusqu'à ce qu'il n'y ait plus de nouveaux pics, ou que les régions restantes ne soient plus significatives. La procédure récursive de seuillage peut être formulée par :

$$\mathbf{f}_{\text{filt}}(i, j) = \mathbf{C}_1(t) \wedge \mathbf{C}_2(t+1) \wedge \mathbf{C}_p(t+i-1) \dots \wedge \mathbf{C}_p(t+p-1) \quad (1)$$

où $\mathbf{f}_{\text{filt}}$ désigne l'image filtrée et $\mathbf{f}_{\text{thr}} = \mathbf{C}_p(t+p-1)$ indique l'image à l'itération $t+p-1$, qui ne contient que les objets d'intérêt (les données visuelles).

En se basant sur les expériences que nous avons effectuées sur 505 chèques [5], le nombre d'itérations est souvent égal à 2 et rarement égal à 3. Ceci est illustré à la Figure 3.

3. Le modèle en bref

Soit N le nombre des entiers naturels, les couples (i,j) des coordonnées spatiales d'une image numérisée, et l'ensemble $G = \{0,1,...,l-1\}$ des valeurs entières représentant des niveaux de gris¹. La fonction image peut être alors définie comme une application $f : N \times N \rightarrow G$, où $f(i,j)$ est le niveau de gris du pixel de coordonnées (i,j) . Aussi, par convention, le niveau de gris 0 correspond à l'intensité la plus sombre et le 255 à la plus claire.

Comme nous l'avons déjà mentionné, nous envisageons par notre approche d'élaborer un modèle capable de préserver et de visualiser la texture la plus fine des données manuscrites ou imprimées sur des images de chèques. Pour poursuivre cette ligne, nous avons adopté en plus des étapes subséquentes, une technique de segmentation locale à l'opposé de la méthode globale développée dans [4]. Celle-ci permet d'améliorer la qualité de l'information extraite et d'accroître, en plus, la visibilité et l'interprétation du système. C'est une contribution majeure dans l'amélioration des résultats de reconnaissance de l'information extraite. Les Figures 1 et 2 illustrent cette nouvelle approche.

Dans ce qui suit, nous présentons l'algorithme détaillé de notre approche en prenant comme application l'extraction de la date, des montants numériques et littéraux à partir des chèques, désignés par $\alpha_i, \forall i \in \{1,2,3\}$, dans l'ordre.

ÉTAPE 1:

Soit une image de niveaux de gris f , produire l'image améliorée f_{filt} :

$$f_{\text{filt}}(i,j) = \frac{1}{N} \sum_{(n,m) \in S} f_D(n,m), \quad \forall i,j \quad (2)$$

Appliquer la procédure du seuillage dynamique telle que définie dans la section 2. Le résultat est l'image f_{thr} .

¹ Dans notre modèle nous considérons 256 niveaux de gris.

ÉTAPE 2 :

Déterminer $L_s = \{L_{s_1}, L_{s_2}, L_{s_3}\}$, l'ensemble fini des lignes de base à partir de l'image f_{thr} ainsi obtenue.

Établir la matrice M qui représente l'interprétation de la structure logique du document, en l'occurrence d'un chèque. Le système serait alors capable d'extraire les données visuelles $\alpha_{f_i} \forall i \in \{1, 2, 3\}$, à l'aide des relations qui peuvent être engendrées par cette matrice. Si on prend l'exemple du chèque de la Figure 3, la matrice M qui lui est associée pourrait être représentée par :

$$M = \begin{matrix} & L_{s_1} & L_{s_2} & L_{s_3} \\ \alpha_{f_1} & \left(\begin{matrix} I & A & A \end{matrix} \right) \\ \alpha_{f_2} & \left(\begin{matrix} B & R & A \end{matrix} \right) \\ \alpha_{f_3} & \left(\begin{matrix} B & B & I \end{matrix} \right) \end{matrix}$$

Le type de relations pouvant être engendré :

$$(\alpha_{f_i}, r, L_{s_i}) \quad \alpha_{f_i} \in \alpha_f, \quad r \in \mathcal{R}, \quad L_{s_i} \in L_s$$

et $\mathcal{R} = \{R, L, A, B, I\}$ désigne les relations (R : droite, L : gauche, A : dessus, B : dessous, I : coupe).

À l'aide de la matrice M , extraire $\alpha_f = \{f_d, f_c, f_l\}$ l'ensemble fini des données visuelles, à partir de l'image f_{fin} . f_d , f_c , f_l représentant respectivement les images de niveaux de gris de la date, des montants numériques et littéraux ainsi extraits, désignées ci-après par le nom générique $f_{d_{fin}}$.

ÉTAPE 3:

Appliquer localement l'approche étendue de Otsu, c-à-d à l'image $f_{d_{fin}}$, pour éliminer le fond :

$$f_{d_{fin}}(i, j) = f_{d_1}(t) \wedge f_{d_2}(t+1) \wedge \dots \wedge f_{d_p}(t+p-1) \quad (3)$$

où $f_{d_{thr}} = f_{d_p}(t+p-1)$.

Déterminer la position de α_{r_D} dans l'image $f_{D_{thr}}$ telle que désignée par la relation $(\alpha_{f_D}, r, L_{s_D})$ et éliminer la ligne de base associée pour produire l'image $f_{D_{lr}}$.

$$f_{D_{lr}} = \begin{cases} 255 & \text{si } f_{D_{thr}}(i, j) \in L_{s_D} \\ f_{D_{thr}}(i, j) & \text{sin on} \end{cases} \quad (4)$$

ÉTAPE 4:

Appliquer l'opération de fermeture, de la morphologie mathématique, à l'image $f_{D_{fil}}$, en utilisant l'élément structurant V. Le résultat est l'image $f_{D_{clos}}$.

ÉTAPE 5:

Restaurer l'information perdue en utilisant l'image $f_{D_{clos}}$ après seuillage.

$$f_{D_{res}} = \begin{cases} f_{D_{clos}}(i, j) & \text{si } f_{D_{clos}}(i, j) \leq \text{Mean} \\ 255 & \text{sin on} \end{cases} \quad (5)$$

$$f_{D_{AND}} = \begin{cases} f_{D_{res}}(i, j) & \text{si } f_{D_{res}}(i, j) \in L_{s_D} \\ f_{D_{lr}}(i, j) & \text{sin on} \end{cases} \quad (6)$$

ÉTAPE 6:

Calculer le gradient des pixels appartenant aux lignes de base, dans le but de restaurer l'information manquante après l'application de l'étape précédente.

$$f_{D_{grd}}(i, j) = \sqrt{(\Delta_x f_{D_{AND}}(i, j))^2 + (\Delta_y f_{D_{AND}}(i, j))^2} \quad (7)$$

$$f_{D_0}(i, j) = \tan^{-1} \left((\Delta_y f_{D_{AND}}(i, j)) / (\Delta_x f_{D_{AND}}(i, j)) \right) \quad (8)$$

Avec

$$\Delta_x f_{D_{AND}}(i, j) = f_{D_{AND}}(i+1, j+1) + 2 f_{D_{AND}}(i+1, j) + f_{D_{AND}}(i+1, j-1) + f_{D_{AND}}(i-1, j+1) - 2 f_{D_{AND}}(i-1, j) - f_{D_{AND}}(i-1, j-1) \quad (9)$$

$$\Delta_y f_{D_{AND}}(i, j) = f_{D_{AND}}(i-1, j+1) + 2 f_{D_{AND}}(i, j+1) + f_{D_{AND}}(i+1, j+1) + f_{D_{AND}}(i-1, j-1) - 2 f_{D_{AND}}(i, j-1) - f_{D_{AND}}(i+1, j-1) \quad (10)$$

ÉTAPE 7:

Appliquer les règles topologiques aux trous identifiés lors du processus précédent. Le processus topologique permet en fait de combler les trous que le processus morphologique était incapable de restaurer en tout. Mais il faut noter que le processus morphologique permet de rehausser et de mettre en évidence ces trous, rendant ainsi le processus topologique plus efficace et plus robuste. Le résultat est l'image $f_{D_{gaps}}$. Pour plus de détails se référer à l'article [4].

ÉTAPE 8:

Finalement appliquer le filtre médian à l'image résultat $f_{D_{gaps}}$, afin d'améliorer la qualité de l'image.

4. Résultats expérimentales et conclusion

Quatre cents vingt-deux chèques bancaires réels ont été utilisés pour entraîner notre système, et 203 autres chèques ont été utilisés comme ensemble de test. Des résultats très encourageants ont été obtenus, Figure 4. Pour plus de détails de cette étude expérimentale, se référer à l'article [4].

Pour conclure, nous avons présenté un modèle formel de traitement automatique de données visuelles sur des formulaires d'affaires et documents financiers. Nous avons proposé dans ce modèle une hiérarchie et une concurrence d'étapes qui permettent de préserver la topologie de l'information manuscrite et d'améliorer la qualité des données visuelles ainsi extraites, dont le but est d'accroître la robustesse du système global de lecture automatique de documents.

Nous travaillons actuellement sur la validation de notre modèle sur une large base de données réelles.

(9) Nous notons que cette approche peut être utilisée aussi efficacement dans d'autres applications.

Remerciements

(10) Cette recherche a été supportée en partie par les organismes suivants : CRSNG du Canada, FCAR du Québec et le Réseau des Centres d'Excellence IRIS.

5. Références

- [1] M. J. Carlotto, "Histogram analysis using scale-space approach," *Trans. Pattern Anal. Mach. Intell.*, vol. 9 (1), 1987, pp. 121-129.
- [2] R. G. Casey, D. R. Ferguson, K. M. Mohiuddin, and E. Walach, "An intelligent forms processing system," *Machine Vision and Applications*, vol. 5, no. 3, 1992, pp. 143-155.
- [3] D. Wang and S. N. Srihari, "Analysis of Form Images," *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 8 No. 5, 1994, pp. 1031-1052.
- [4] M. Cheriet, J. N. Said, and C. Y. Suen, "A formal Model for Document Processing of Bank Cheques," *3rd Inter. Conf. on Document Analysis and Recognition, ICDAR'95*, Montreal, Aug 24-26, 1995, pp. 210-213.
- [5] M. Cheriet, J. N. Said, and C. Y. Suen, "A Recursive Approach for Image Segmentation," *3rd IEEE Trans. Image Processing*, (submitted 1995), pp. 210-213.
- [6] L. Hertz and R.W. Schafer, "Multilevel thresholding using edge matching," *Compt. Vision Graphics Image Process*, vol 44, 1988, pp. 279-295.
- [7] S. L. Horowitz and T. Pavlidis, "Picture segmentation by a tree traversal algorithm," *J. Assoc. Comput. Mach.*, vol. 23, Apr. 1976, pp. 368-388.
- [8] J. N. Kapur, P. K. Sahoo, and A. K. Wong, "A new method for gray-level picture thresholding using the entropy of the histogram," *Compt. Vision Graphics Image Process*, vol. 29, 1985, pp. 273-285.
- [9] R. Kasturi et al, "A system for interpretation of line drawings," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. PAMI-12, Oct. 1990, no. 10.
- [10] R. Kohler, "A segmentation system based on thresholding," *Compt. Graphics Image Process*, vol. 15, 1981, pp. 319-338.
- [11] L. Lam, C. Y. Suen, D. Guillevic, N. W. Strathy, M. Cheriet, K. Liu, and J. N. Said, "Automatic processing of information on cheques," *Proc. 1995 IEEE Int. Conf. on Systems, Man and Cybernetics*, Vancouver, Canada, Oct. 22-25, 1995, Vol. 3, pp. 2353-2358.
- [12] S. U Lee, S. Y. Chung, and R. H. Park, "A comparative performance study of several global thresholding techniques for segmentation," *Comput. Vision Graphics Image Process.* {bf 52} 1990, pp. 171-190. *Machine Vision and Appl.*, Special Issue: Document Image Analysis Techniques, 1992, vol. 5, no., 3.

- [13] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Systems Man Cybernet*, vol. SMC-8, 1978, pp. 62-66.
- [14] S. S. Reddi, S. F. Rudin, and H. R. Keshavan, "An optimal multiple threshold scheme for image segmentation," *IEEE Trans. Systems Man Cybernet*, vol. 14 (4), 1984, pp. 661-665.
- [15] P. K. Sahoo, S. Soltani, and A. K. C. Wong, "SURVEY: A survey of thresholding techniques," *Compt. Vision Graphics Images Process*, vol. 41, 1988, pp. 233-260.
- [16] M. Spann and R. Wilson, "A quad-tree approach to image segmentation which combines statistical and spatial information," *Pattern Recognition*, vol. 18, 1985, pp. 257-269.
- [17] S. L. Taylor, R. Fritzson, and J. A. Pastor, "Extraction of data from preprinted forms," *Machine Vision and Applications*, vol. 5, no. 3, 1992, pp. 211-222.
- [18] Y. Y. Tang, C. D. Yan, M. Cheriet, C. Y. Suen, "Automatic analysis and understanding of documents." C. H. Chen, L.F. Pau, and P.S.P. Wang, editors, *Handbook of Pattern Recognition and Computer Vision*, World Scientific, (Singapore), 1993.
- [19] T. Y. Tang, C. Y. Suen, C. D. Yan, M. Cheriet, "Financial Document Processing Based on Staff Line and Description Language," *IEEE Trans. Systems, Man. and Cybernetics*, vol. 25, no. 5, May 1995, pp. 738-754.
- [20] S. Wang and R. M. Haralick, Automatic multithreshold selection, *Compt. Vision Graphics Image Process*, vol. 25, 1984, pp. 46-67.
- [21] C. D. Yan, Y. Y. Tang, and C. Y. Suen, "Form understanding system based on form description language," in *Proc. First Int. Conf. on Document Analysis and Recognition*, (Saint-Malo), France, Sept. 30-Oct2, 1991, pp. 283-293.

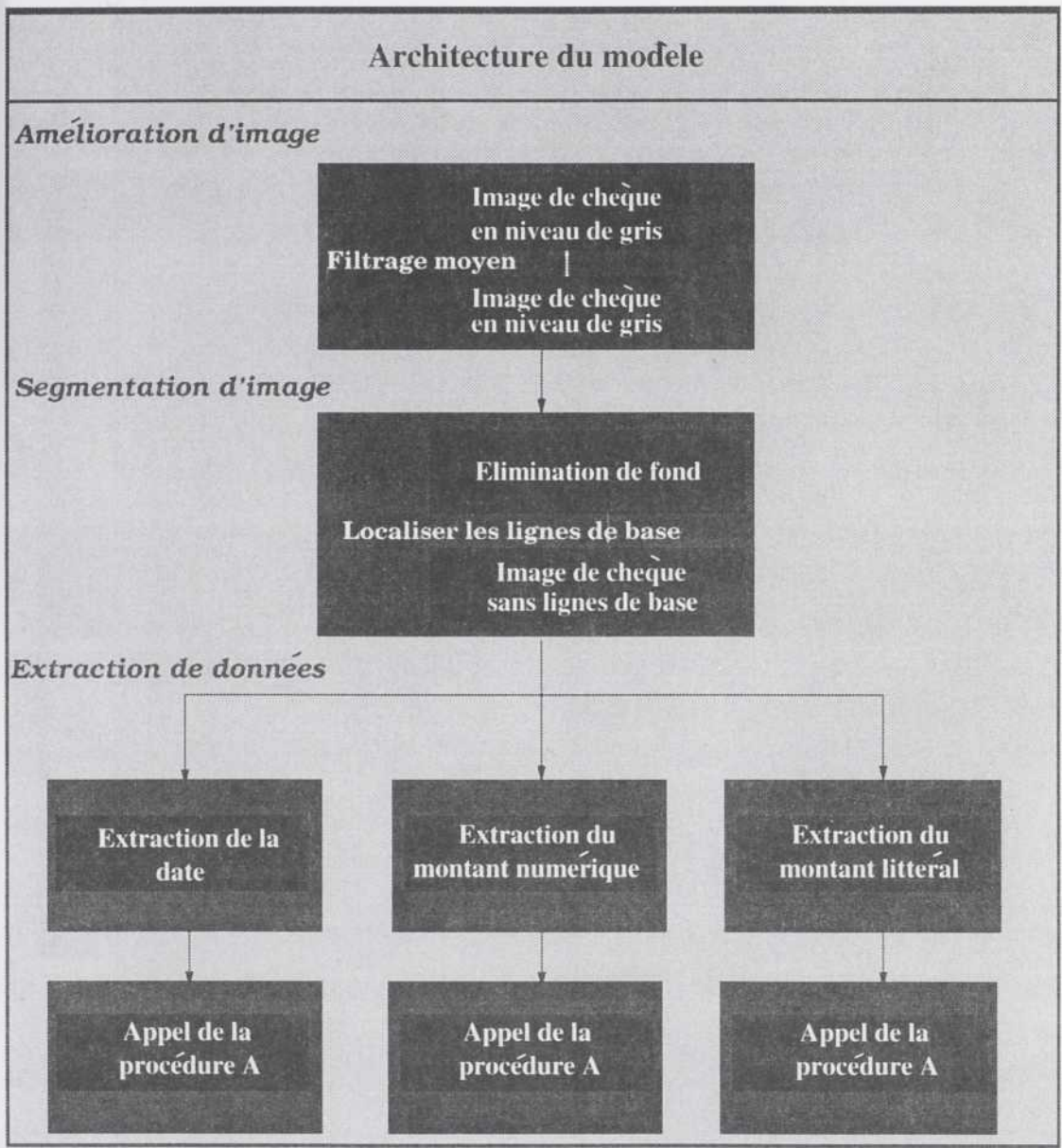


Figure 1: Désign du system de traitement de chèques.

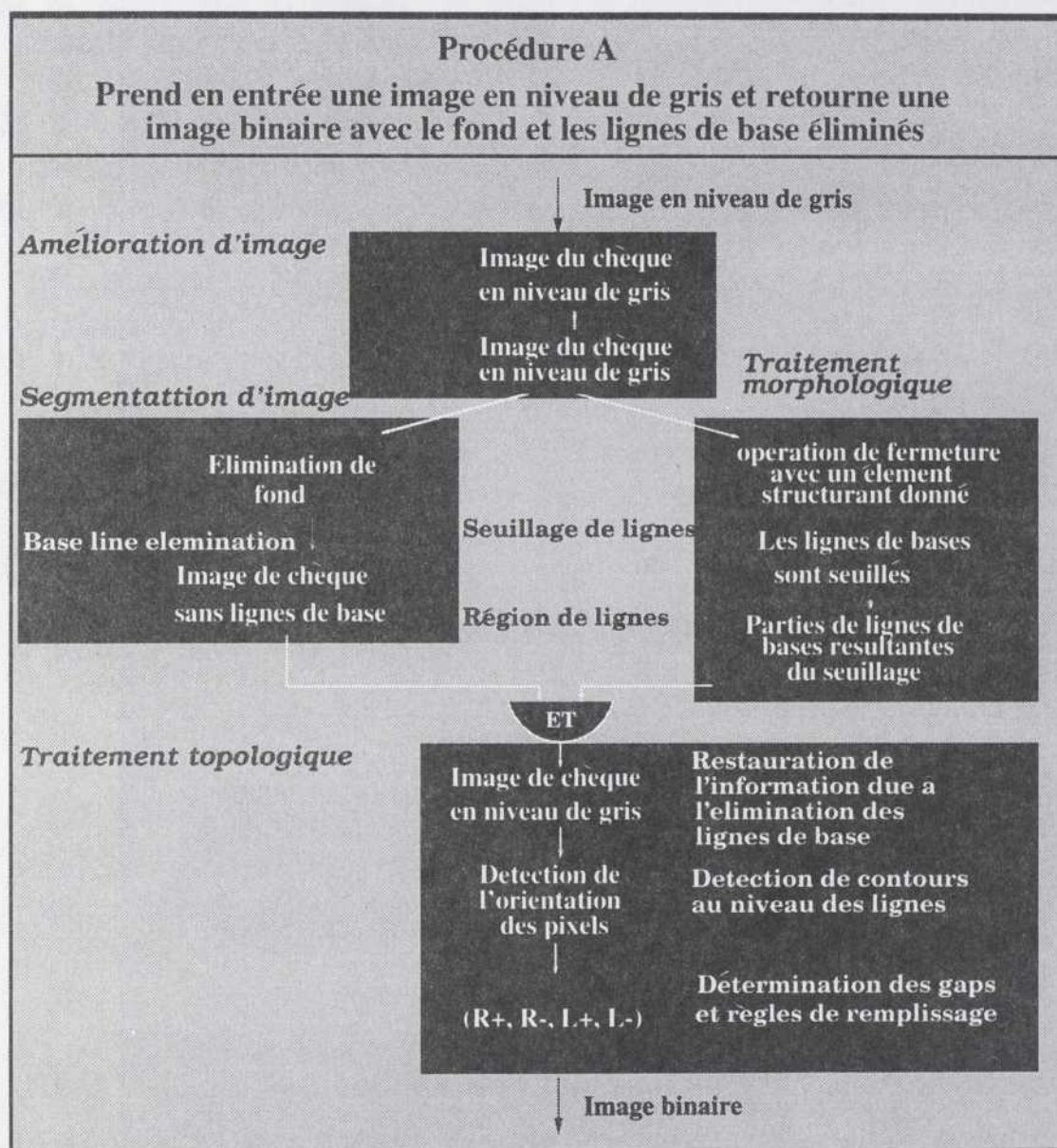


Figure 2: Procédure A.

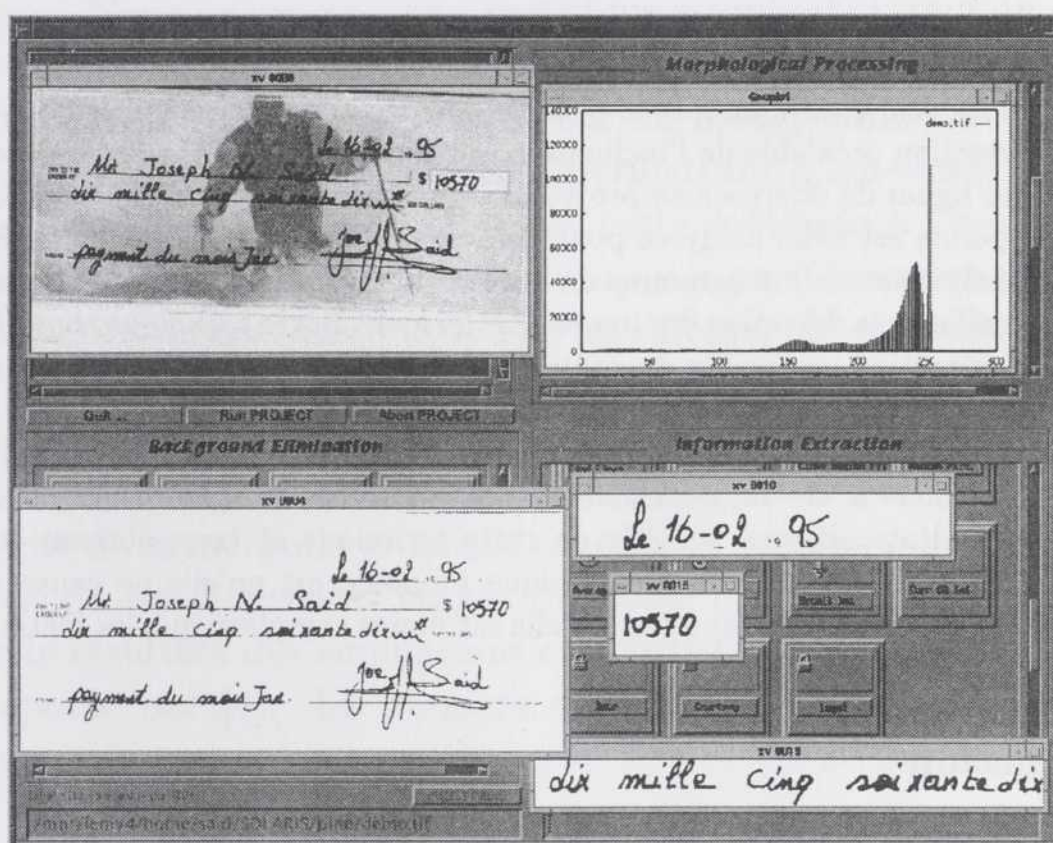


Figure 3: Extraction des données visuelles a partir de chèques.

Détection des lignes de référence de mots cursifs à l'aide de l'entropie

M. Côté¹⁻³, M. Cheriet¹, C.Y. Suen², E. Lecolinet³

¹ E.T.S., Département GPA, Montréal, (Québec), CANADA.

² CENPARMI, Concordia University, Montréal, (Québec), CANADA.

³ ENST, Département signal, Paris, FRANCE.

e-mail: cote@cenparmi.concordia.ca, cheriet@gpa.etsmtl.ca,
suen@cenparmi.concordia.ca, elc@sig.enst.fr.

Résumé

Nous proposons une technique simple d'extraction des lignes de référence de mots cursifs, sans correction préalable de l'inclinaison. La plupart des méthodes déjà connues trouvent les lignes de référence en projetant les points du mot sur l'axe vertical. La densité des points est alors analysée pour trouver ces lignes. Nous montrons que le calcul de l'entropie sur un histogramme de densité donne un indice supplémentaire permettant d'améliorer la détection des lignes de référence. Notre technique comporte quatre étapes: extraction des points du contour du mot, génération d'histogrammes des ordonnées des points, sélection de l'histogramme le plus compact à l'aide de l'entropie et déduction des lignes de référence à partir de l'analyse de l'histogramme choisi. Cette méthode a été implémentée et testée sur des images réelles. Nous comparons les résultats obtenus à l'aide de cette technique et ceux obtenus avec une autre méthode. L'avantage de la technique proposée est qu'elle ne cause aucune déformation de l'image de départ et qu'elle est moins complexe que les méthodes utilisant des régressions linéaires.

1. Introduction

Le succès de la reconnaissance de mots cursifs isolés dépend de l'efficacité de chacune des étapes en jeu. Afin de bien détecter les primitives significatives du mot comme les ascendants et les descendants, il est essentiel de pouvoir compter sur une technique performante d'extraction des lignes de référence qui sépare le mot en trois parties comme on peut le constater à la Fig. 5. Entre les deux lignes se trouve le corps du mot. Les ascendants sont localisés en haut de la ligne supérieure et les descendants, en bas de la ligne inférieure. Le problème est que l'écriture cursive est souvent inclinée et les méthodes habituelles de redressement du mot entraînent souvent des distorsions causant des pertes d'information importantes. C'est d'ailleurs pour

cette raison que des techniques de pré-traitement évitant la correction de l'inclinaison ont vu le jour [1], [2], [3] et [4]. Les méthodes déjà connues trouvent les lignes de référence, après correction de l'inclinaison du mot, en analysant les points critiques du contour du mot [5] ou en projetant les points du mot sur l'axe vertical [6]. La densité des points est alors analysée pour trouver les lignes. Dans [4], les lignes de référence sont déterminées en combinant 2 méthodes complémentaires. L'une calcule la distribution des traits verticaux et l'autre, l'histogramme des transitions sur l'axe des y. Nous proposons une technique simple d'extraction des lignes de référence sans correction préalable de l'inclinaison. Nous montrons que le calcul de l'entropie sur un histogramme de densité nous donne un indice supplémentaire permettant d'améliorer la détection des lignes de référence. De plus, cette méthode à l'avantage d'être moins complexe que la méthode proposée par Caesar et al. (1995) qui utilise des régressions linéaires. Par ailleurs, elle s'inscrit dans le prolongement des travaux de Cheriet et Suen (1993) qui ont développé une méthode de détection de lignes de référence suivant l'inclinaison du mot, fondée sur le principe de l'étalement ou "Smearing". Le mérite de la méthode proposée est qu'elle est plus robuste. La section 2 décrit les différentes étapes de notre méthode, la section 3 montre le résultats des simulations et compare notre technique avec celle proposée par [4]. La dernière section tire les conclusions de notre présentation.

2. Méthode proposée

Dans notre méthode, nous projetons les ordonnées du contour du mot selon plusieurs plans de rotation. Le calcul de l'entropie de l'histogramme des ordonnées dans chaque plan nous permet de trouver l'orientation selon laquelle le mot est le plus compact. La connaissance de cette orientation et l'analyse de la distribution de l'histogramme sélectionné nous permettent de tracer des lignes de référence selon l'orientation du mot. L'avantage de cette technique est qu'elle ne cause aucune déformation de l'image de départ. Notre méthode est réalisée en 4 étapes: i) extraction des points du contour du mot, ii)

génération d'histogrammes des ordonnées des points, iii) sélection de l'histogramme le plus compact à l'aide de l'entropie et iv) déduction des lignes de référence à partir de l'analyse de l'histogramme choisi. Les 3 premières étapes servent en fait à calculer l'inclinaison. Les sections qui suivent décrivent plus en détails, chacune de ces étapes.

2.1 Extraction des points du contour

Statistiquement, la majorité des points devraient se retrouver dans le corps central du mot. Comme nous ne voulons pas que les barres de 't' et les lettres comportant des projections ascendantes et descendantes aient un poids important dans cet ensemble de points, l'empreinte du mot sera constituée uniquement des points du contour. La Fig. 1 montre l'empreinte du mot "Elmira".

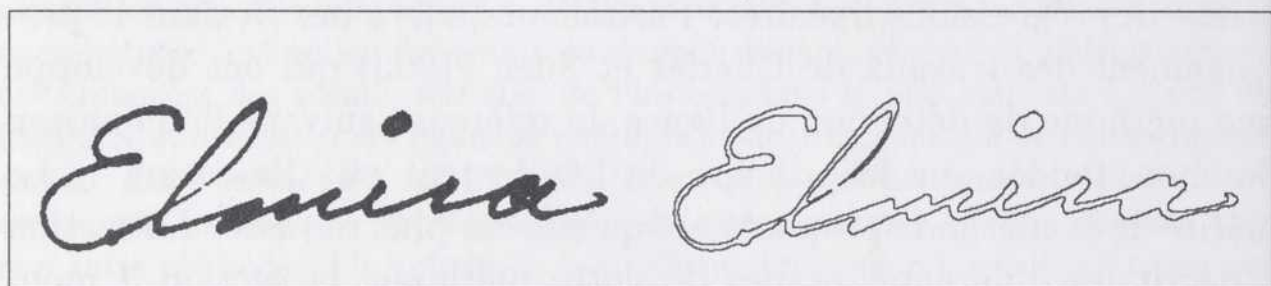


Figure 1: Le mot "Elmira" et son empreinte.

2.2 Génération d'histogrammes

Chaque ordonnée des points du contour est successivement projetée dans un repère dont l'axe d'inclinaison par rapport à l'horizontale varie selon un pas de 1° de -30° à $+30^\circ$. Le balayage de 60° autour de l'horizontale semble raisonnable puisqu'il est peu probable qu'un mot soit écrit avec une plus grande inclinaison. Cependant, si cela devait se produire, il est possible d'augmenter le balayage. Pour chacune des 61 projections, un histogramme de densité des ordonnées est calculé. On s'attend à ce que l'inclinaison réelle du mot donne la répartition la plus compacte de l'histogramme. La Fig. 2 montre l'histogramme de densité des ordonnées selon différents angles de projection. On remarquera que l'étalement diminue d'autant plus qu'on

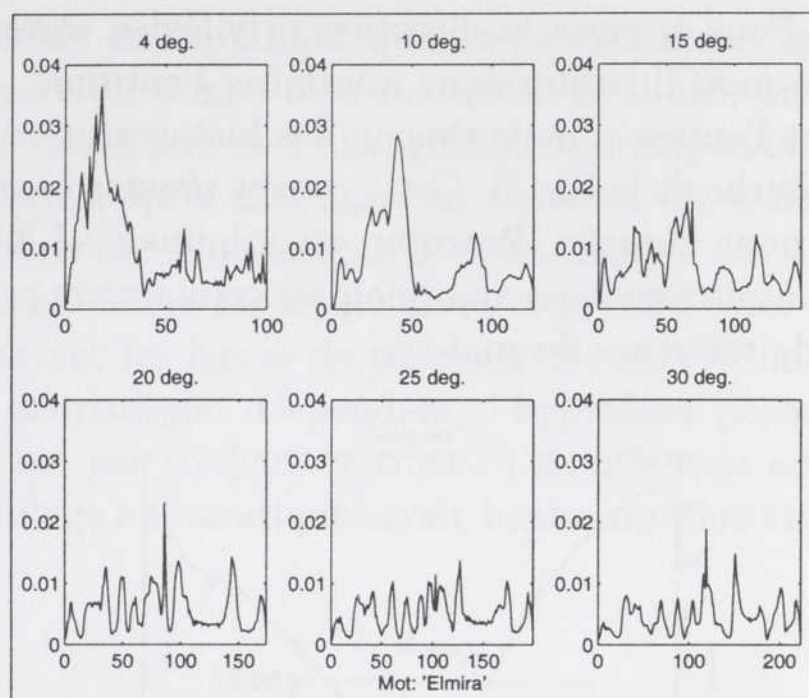


Figure 2: Exemples d'histogrammes de densité pour le mot "Elmira".

se rapproche de l'angle réel d'inclinaison du mot "Elmira", soit 4° . La section suivante montre comment l'entropie a été utilisée comme critère privilégié afin de choisir cet histogramme.

2.3 Choix du meilleur histogramme selon l'entropie

L'entropie est une mesure de l'information donnée par l'équation suivante:

$$E = - \sum_i p_i \log(p_i) \quad (1)$$

$$p_i = \frac{N_i}{N} \quad (2)$$

où N_i est le nombre de pixels ayant l'ordonnée y_i dans le repère de projection et où N est le nombre total de pixels ou de points dans le contour du mot.

La probabilité p_i de l'histogramme désigne la fréquence d'occurrence de l'ordonnée y_i ($\sum_i p_i = 1$). L'entropie E est maximale si toutes les probabilités p_i sont égales. Elle est minimale si toutes les probabilités p_i sauf une sont nulles. Ainsi, plus une distribution est hétérogène, moins elle est étalée et plus petite est son entropie. Au contraire, plus une distribution est homogène, plus elle est étalée et plus grande est

son entropie. Pour trouver la direction privilégiée correspondant à l'inclinaison du mot, il faudra donc minimiser l'entropie.

En calculant l'entropie pour chacun des histogrammes de densité, on obtient la courbe de la Fig. 3. Cette courbe montre bien qu'il existe une direction pour laquelle l'entropie est minimale. L'histogramme correspondant à cette entropie minimale est celui retenu pour déterminer les lignes de référence du mot.

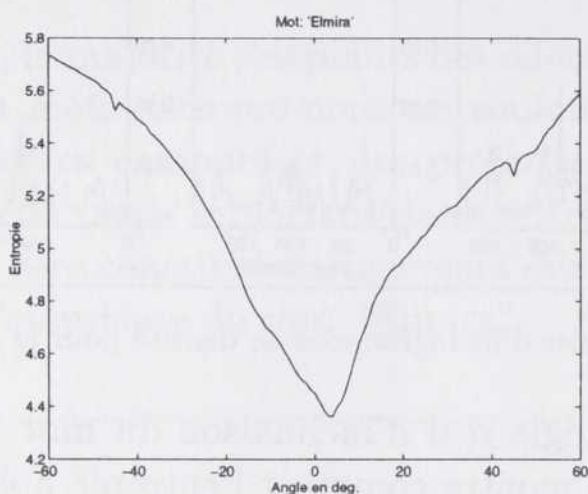


Figure 3: Courbe d'entropie en fonction de l'angle de projection pour le mot "Elmira".

2.4 Extraction des lignes de référence

L'analyse de l'histogramme sélectionné permet de déterminer les lignes de référence. Idéalement, l'histogramme devrait présenter au plus 3 pics: un pic central correspondant au corps du mot et 2 plus petits pics de chaque côté représentant respectivement la zone des ascendants et celle des descendants. L'absence de pics à gauche ou à droite du pic central peut être utilisée pour confirmer l'absence d'ascendants ou de descendants. C'est le cas pour l'histogramme de la Fig. 4 où il n'y a pas de pics correspondants aux descendants. Les lignes de référence sont par définition les droites séparant le pic central du mot des zones latérales. Il faut donc trouver les droites qui délimitent le pic central de l'histogramme et dont les ordonnées à l'origine sont les seuils. En traitement d'image, ce problème est couramment rencontré pour extraire l'information utile d'une image bruitée. Plusieurs méthodes permettent de trouver ces seuils [7]. Dans notre cas, une simple heuristique peut donner une approxima-

tion raisonnable en supposant que les ascendants et les descendants sont la plupart du temps bien marqués. D'abord, on cherche f_{max} , la fréquence maximale de l'histogramme correspondant au pic central. On suppose qu'à $\pm 0.2f_{max}$, on trouvera les seuils associés aux lignes de référence. Connaissant l'inclinaison et ces deux seuils, on peut facilement tracer les lignes de référence du mot. Sur la Fig. 5, on peut observer les lignes de référence du mot "Elmira", calculées selon cette heuristique. Cependant, l'hypothèse posant les seuils à $\pm 0.2f_{max}$, n'est pas toujours vérifiée. Une méthode adaptative pour faire un seuillage automatique serait beaucoup plus élégante dans ce cas.

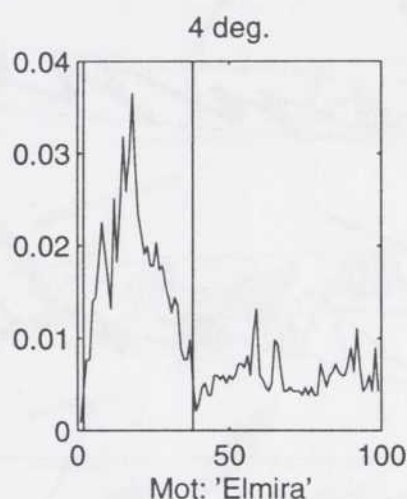


Figure 4: Histogramme correspondant à l'inclinaison du mot "Elmira".

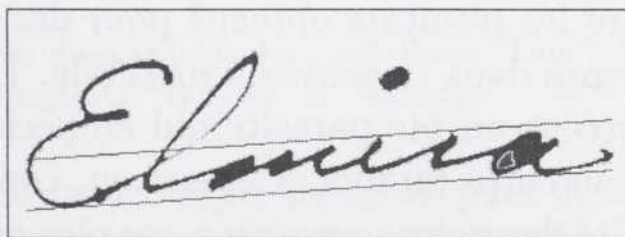


Figure 5: lignes de référence du mot "Elmira".

3. Simulations, résultats et discussion

Cette méthode a été implémentée. Les tests portent sur 63 images de mots cursifs isolés. La Fig. 6 montre les histogrammes choisis selon la courbe d'entropie pour les mots "Elmira", "Almaden", "Tennessee" et "Treadwell".

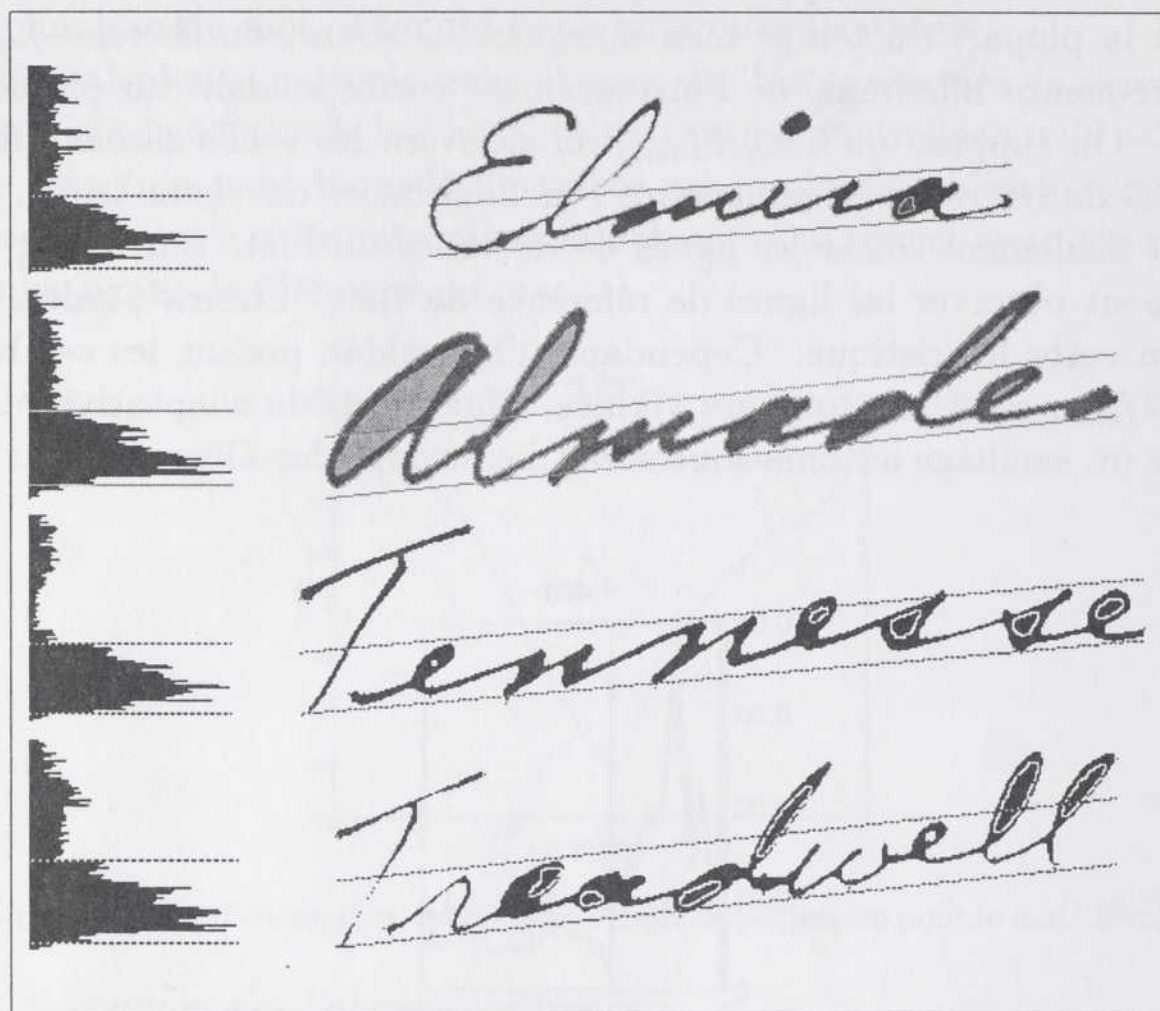


Figure 6: Histogrammes obtenus pour les mots "Elmira", "Almaden", "Tennessee" et "Treadwell".

La Fig. 7 montre les résultats obtenus pour des mots qui causent souvent des problèmes dans ce genre de méthode. En effet, les barres de 'T' et de 'F', créent un pic parasite qui empêche la détection du pic central associé au corps du mot. Cependant, comme la plupart du temps, la probabilité des points associés à ces pics parasites est faible, il est possible de les éviter. Cette technique fonctionne aussi pour des mots comme "Talmage" et "Tappan" qui possèdent des ascendants, des descendants et des barres de 'T'.

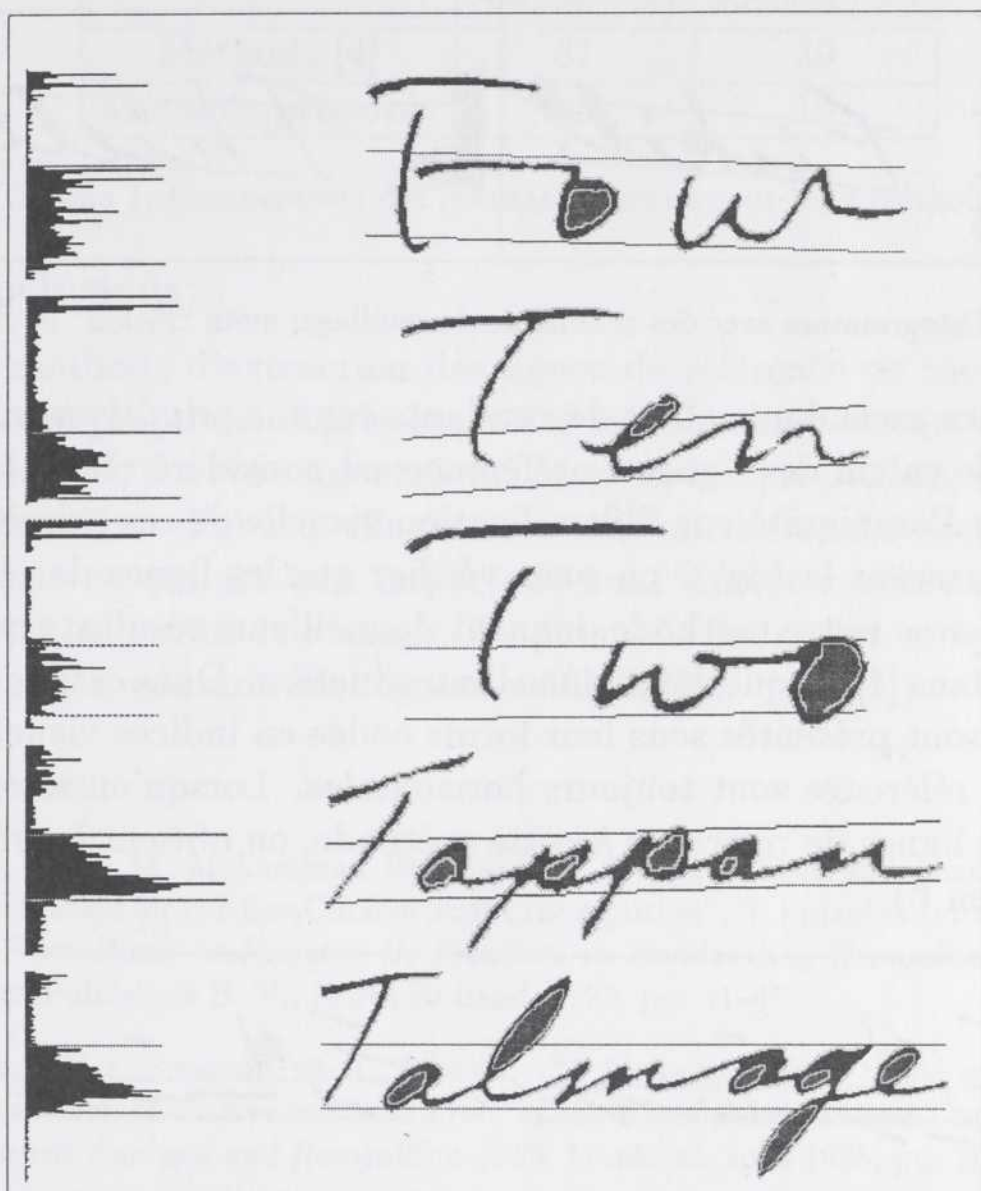


Figure 7: Histogrammes obtenus pour des mots courts "Four", "Ten", "Two" et pour des mots avec des descendants "Tappan" et "Talmage".

La Fig. 8 donne un exemple de mots pour lesquels les seuils ne sont pas déterminés précisément même si la direction est correctement évaluée. Dans ce cas, notre heuristique n'est pas adéquate. Il faudrait optimiser la recherche des seuils de façon à obtenir de meilleures lignes de référence. C'est pourquoi nous sommes à implémenter une recherche adaptative des seuils.

Comme le font remarquer judicieusement Caesar et al. (1995), l'évaluation de ce genre de technique est un problème en soi puisqu'il s'agit de juger un résultat subjectif. Puisque notre but est de bien

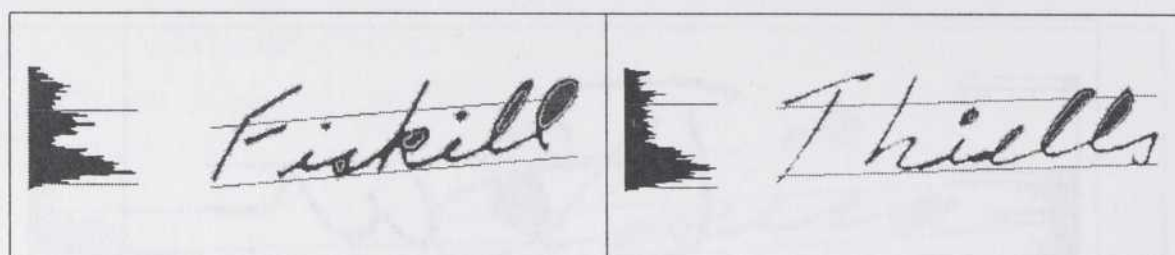


Figure 8: Histogrammes avec des problèmes de seuillage; mots "Fiskill" et "Thiells".

détecter les ascendants et les descendants comme primitives saillantes du mot, le calcul des lignes de référence est considéré réussi lorsqu'il n'y a pas d'ambiguïté sur l'identification visuelle de ces primitives.

En observant la Fig. 9 on peut vérifier que les lignes de référence obtenues avec notre méthode donnent de meilleurs résultats que ceux obtenus dans [4] tel que défini dans l'introduction. Dans cette dernière, les mots sont présentés sous leur forme codée en indices visuels et les lignes de référence sont toujours horizontales. Lorsqu'on intègre nos nouvelles lignes de référence à cette méthode, on obtient les résultats obtenus en b).

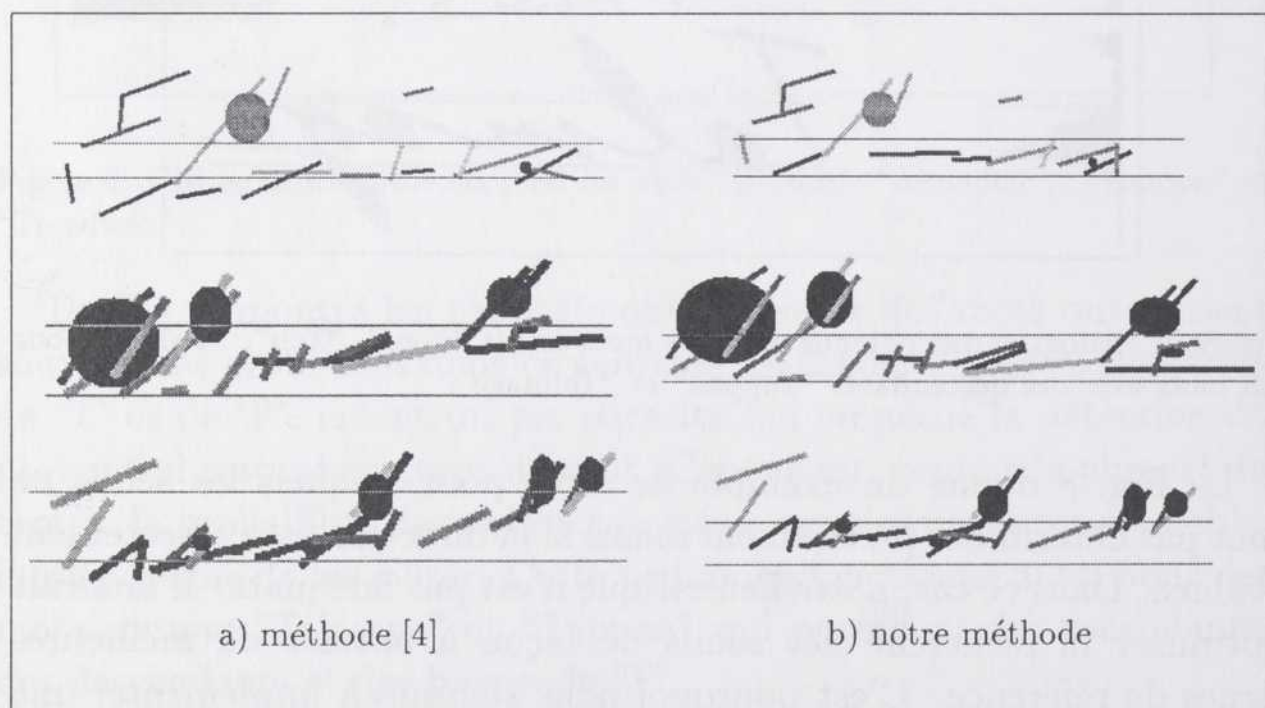


Figure 9: lignes de référence obtenues pour les mots "Elmira", "Almaden" et "Treadwell".

Le tableau 1 résume les statistiques pour les deux méthodes.

	Succès (%)	Echecs (%)
Méthode [4]	61	39
Méthode proposée	85	15

Tableau 1: Comparaison des résultats obtenus pour les 2 méthodes.

4. Conclusions

Une méthode d'extraction des lignes de référence de mots cursifs isolés a été proposée. Après avoir été implémentée et testée sur des images réelles, il a été démontré que cette technique présente plusieurs avantages. Tout d'abord, elle ne cause aucune déformation de l'image de départ, ce qui est très important pour détecter efficacement les primitives saillantes de l'image. De plus, elle est moins complexe que les méthodes utilisant des régressions linéaires.

Références

- [1] M.J.J. Holt, M. Mohammad Beglou, et S. Datta, "Slant-Independent Letter Segmentation for Off-line Cursive Script Recognition", S. Impedovo et J.C. Simon, Ed., *From Pixels to Features II: Frontiers in Handwriting Recognition*, Elsevier Science Publishers B. V., North-Holland, 1992, pp. 41-47.
- [2] A. Negi, K.S. Swaroop, et A. Agarwa, "A Correspondence based approach to segmentation of cursive words", *Proc. of the Third International Conference on Document Analysis and Recognition 1995*, Montréal, août 1995, pp. 1034-1037.
- [3] T. Caesar, J.M. Gloger, et E. Mandler, "Estimating the Baseline for Written Material", *Proc. of the Third International Conference on Document Analysis and Recognition 1995*, Montréal, août 1995, pp. 382-385.
- [4] E. Lecolinet, "Cursive Script Recognition by Backward Matching", C. Faure, P. Keuss, G. Lorette, et A. Vinter, Ed., *Advances in Handwriting and Drawing: A Multidisciplinary Approach*, Europa, Paris, 1994, pp. 117-135.
- [5] M.K. Brown, "Preprocessing Techniques for Cursive Ccript Word Recognition", *Pattern Recognition*, vol. 16, 1983, pp. 447-458.
- [6] R.M. Bozinovic, S.N. Srihari, "Off-line Cursive Script Word Recognition", *IEEE Trans. PAMI*, vol. 11, 1989, pp. 68-83.
- [7] P.K. Sahoo, S. Soltani, et A.K.C. Wong, "A Survey of Thresholding Techniques", *Computer Vision, Graphics, and Image Processing*, vol. 41, 1988, pp. 233-260.

Modélisation du langage pour la reconnaissance de la parole: une approche stochastique non-déterministe à contexte de longueur variable

Marie-Hélène Talon[†], Pierre Dumouchel[‡], Douglas O'Shaughnessy[†]

[†]INRS-Télécommunications, Ile-des-Soeurs (Québec).

[‡]Ecole de technologie Supérieure, département de génie électrique, Montréal (Québec).
e-mail: talon@inrs-telecom.ubec.ca

Résumé

Nous présentons dans cet article un modèle de langage stochastique non-déterministe à contexte de longueur variable intitulé VN_{SA}: Variable N-gram Stochastic Automaton. Les modèles de langage traditionnels sont déterministes, en ce sens qu'ils ne permettent qu'une seule décomposition possible pour chaque phrase donnée: une fenêtre temporelle de longueur fixe est supposée, c'est-à-dire que l'on calcule la probabilité conditionnelle d'un mot connaissant les N-1 mots précédents, N étant fixe. Le modèle VN_{SA} introduit le non-déterminisme et ne fixe plus la longueur de la fenêtre temporelle. Les probabilités conditionnelles du modèle sont estimées par la méthode du maximum de vraisemblance et pondérées pour tenir compte de la probabilité de rencontrer des mots nouveaux.

Ce modèle sera testé sur le corpus ATIS et les résultats seront comparés à ceux d'un modèle classique d'interpolation. La mesure de qualité du modèle de langage utilisée est la perplexité. Il s'agit d'une mesure standard qui permet de quantifier la complexité du langage.

1. Introduction

1.1 Pourquoi un modèle de langage

L'être humain, pour bien comprendre un message vocal, n'utilise pas que les informations acoustiques du signal de parole reçu. Il tient également compte du contexte syntaxique et sémantique du message, du contexte prosodique voire même du contexte sémiotique et pragmatique. Ces indices facilitent la bonne compréhension du message. Il est donc normal que l'ordinateur, dans une tâche de reconnaissance de la parole, essaie de prendre en considération certaines de ces informations supplémentaires, d'où l'idée de créer ce qu'on appelle un "modèle de langage", un outil qui intègre des informations syntaxiques et sémantiques au processus de reconnaissance de la parole dans le but d'obtenir de meilleures performances.

La tâche d'un système de reconnaissance de la parole est d'associer la bonne séquence de mots au signal acoustique reçu. Soit $M = m_1, m_2, \dots, m_N$, une séquence de N mots, et A les données acoustiques fournies au système de reconnaissance. Le système tentera de déterminer $\hat{M}$, la séquence la plus probable, c'est-à-dire celle qui maximise la probabilité d'une séquence étant donné le signal acoustique:

$$\hat{M} = \underset{M}{\operatorname{argmax}} P(M|A) \quad (1)$$

La formule de Bayes nous permet d'écrire:

$$P(M|A) = \frac{P(A|M)P(M)}{P(A)} \quad (2)$$

où $P(A|M)$: probabilité d'observer le signal acoustique A étant donnée M , la séquence de mots prononcée $\rightarrow$ modèle acoustico-phonétique.

$P(M)$: probabilité de la séquence de mots $\rightarrow$ modèle de langage.

$P(A)$: probabilité d'observer le signal acoustique A .

$P(A)$ n'intervenant pas dans la maximisation sur M , le critère de décision devient donc:

$$\hat{M} = \underset{M}{\operatorname{argmax}} \{P(A|M)P(M)\} \quad (3)$$

L'évaluation de $P(M)$ est donc un élément du processus de reconnaissance de la parole.

1.2 Le modèle n -gramme

La probabilité de rencontrer une séquence de N mots $M = m_1, m_2, \dots, m_N$ peut être calculée comme le produit des probabilités individuelles de chacun des mots, si on considère que les mots d'une phrase sont indépendants:

$$P(M) = P(m_1)P(m_2)\dots P(m_N) = \prod_{i=1}^N P(m_i) \quad (4)$$

Puisque les mots d'une phrase ne sont pas indépendants, mais qu'au

contraire ils suivent des règles syntaxiques et sémantiques d'enchaînement, on doit plutôt estimer $P(M)$ comme le produit des probabilités conditionnelles de chacun des mots, ce qui nous conduit à la règle en chaîne [1] :

$$P(M) = P(m_1)P(m_2|m_1)...P(m_N|m_1, m_2, \dots, m_{N-1}) = \prod_{i=1}^N P(m_i|m_1, \dots, m_{i-1}) \quad (5)$$

En pratique, il est difficile d'estimer de façon fiable les $P(m_i|m_1, \dots, m_{i-1})$, même pour des valeurs de i très petites, parce que les séquences $m_1, \dots, m_{i-1}$ peuvent n'être répétées qu'un très petit nombre de fois dans l'ensemble d'entraînement, voire ne pas se produire du tout.

Le modèle de langage n-gramme (un n-gramme est une suite de n mots consécutifs, un seul mot est donc appelé unigramme, deux mots consécutifs forment un bigramme) contourne ce problème en faisant l'hypothèse que la probabilité d'un mot dans une phrase ne dépend que des $n-1$ mots précédents:

$$P(M) = \prod_{i=1}^N P(m_i|m_{i-n+1}, \dots, m_{i-1}) \quad (6)$$

D'une manière générale, les estimés des $P(m_i|m_{i-n+1}, \dots, m_{i-1})$ sont obtenus par la méthode du maximum de vraisemblance, en comptant les occurrences des séquences de mots dans les données d'entraînement:

$$\hat{P}(m_i|m_{i-n+1}, \dots, m_{i-1}) = \frac{N(m_{i-n+1}, \dots, m_{i-1}, m_i)}{N(m_{i-n+1}, \dots, m_{i-1}, \Psi)} \quad (7)$$

$N(x)$ étant le nombre d'occurrences de la séquence x dans l'ensemble d'entraînement et Ψ , un mot quelconque de l'ensemble d'entraînement.

Cette estimation de la probabilité des n-grammes pose un problème de taille car même dans un très grand ensemble d'entraînement, plusieurs n-grammes $m_1, \dots, m_n$ ne se produisent qu'un très petit nombre de fois ou ne sont jamais observés, alors qu'on les retrouve dans l'ensemble de test. Une probabilité nulle pour un n-gramme rendrait impossible l'utilisation du critère

de décision (3).

Il existe deux approches classiques pour résoudre ce problème: l'interpolation et le "back-off".

1.3 Le modèle d'interpolation

Dans le modèle d'interpolation, la probabilité d'un n-gramme est une combinaison linéaire de la probabilité du n-gramme en question et des probabilités des n-grammes d'ordres moins élevés:

$$\hat{P}_I(m_i | m_{i-n+1}, \dots, m_{i-1}) = \lambda_n \hat{P}(m_i | m_{i-n+1}, \dots, m_{i-1}) + \dots + \lambda_2 \hat{P}(m_i | m_{i-1}) + \lambda_1 \hat{P}(m_i) + \lambda_0 \quad (8)$$

Les λ_i sont choisis de façon à maximiser globalement $P(M)$. Leur somme doit être égale à un. λ_0 est une probabilité de défaut pour les mots inconnus.

Par exemple, la probabilité de rencontrer le trigramme "la petite vie", c'est-à-dire la probabilité de rencontrer le mot "vie" suivant les mots "la" et "petite" par le modèle d'interpolation est:

$$\hat{P}_I(vie | la, petite) = \lambda_3 \hat{P}(vie | la, petite) + \lambda_2 \hat{P}(vie | petite) + \lambda_1 \hat{P}(vie) + \lambda_0$$

Les probabilités conditionnelles sont évaluées dans un ensemble d'entraînement comme suit:

$$\hat{P}(vie | la, petite) = \frac{N(la, petite, vie)}{N(la, petite, \Psi)}, \hat{P}(vie | petite) = \frac{N(petite, vie)}{N(petite, \Psi)} \text{ et} \\ \hat{P}(vie) = \frac{N(vie)}{N}, N \text{ étant le nombre total de mots du corpus d'entraînement.}$$

1.4 Le modèle "back-off"

Pour estimer la probabilité d'un n-gramme qui n'a jamais été rencontré dans l'ensemble d'entraînement, le modèle de langage "back-off" emploie la probabilité d'un n-gramme d'un ordre moins élevé, si elle existe. Mathématiquement, ce modèle se formalise comme suit:

$$\text{Si } \hat{P}(m_i | m_{i-n+1}, \dots, m_{i-1}) \neq 0$$

alors $\hat{P}_{BO}(m_i | m_{i-n+1}, \dots, m_{i-1}) = \hat{P}(m_i | m_{i-n+1}, \dots, m_{i-1})$;
 sinon si $\hat{P}(m_i | m_{i-n+2}, \dots, m_{i-1}) \neq 0$
 alors $\hat{P}_{BO}(m_i | m_{i-n+1}, \dots, m_{i-1}) = \alpha_1 \hat{P}(m_i | m_{i-n+2}, \dots, m_{i-1})$;
 ...

$$\text{sinon } \hat{P}_{BO}(m_i | m_{i-n+1}, \dots, m_{i-1}) = \prod_{j=1}^{n-1} \alpha_j \hat{P}(m_i) \quad (9)$$

On utilise une probabilité de défaut pour le cas des mots nouveaux. Appliquons le back-off à notre exemple:

Si $\hat{P}(\text{vie} | \text{la}, \text{petite})$ existe, alors $\hat{P}_{BO}(\text{vie} | \text{la}, \text{petite}) = \hat{P}(\text{vie} | \text{la}, \text{petite})$
 sinon si $\hat{P}(\text{vie} | \text{petite})$ existe, alors $\hat{P}_{BO}(\text{vie} | \text{la}, \text{petite}) = \alpha_1 \hat{P}(\text{vie} | \text{petite})$
 sinon si $\hat{P}(\text{vie})$ existe, alors $\hat{P}_{BO}(\text{vie} | \text{la}, \text{petite}) = \alpha_1 \alpha_2 \hat{P}(\text{vie})$

On évalue les probabilités conditionnelles de la même façon que pour le modèle d'interpolation. Les paramètres λ du modèle d'interpolation et α du back-off sont estimés de façon à maximiser $P(M)$ tout en conservant un modèle de langage probabiliste, c'est-à-dire que la somme des probabilités est égale à un.

1.5 Mesure de la qualité d'un modèle de langage

La perplexité est la mesure standard pour comparer la qualité des modèles de langage. C'est une mesure de la difficulté de la tâche de prédiction d'un texte, relativement au modèle de langage. Cette perplexité est calculée comme suit:

$$PP = P(m_1, m_2, \dots, m_N)^{-1/N} \quad (10)$$

Elle donne la taille moyenne de l'ensemble de mots parmi lesquels le système de reconnaissance doit choisir le prochain mot. C'est un facteur de branchement et le meilleur modèle de langage sera donc celui dont la perplexité est la plus faible.

2. Le modèle VNSA

2.1 Principe de base

Le modèle que nous présentons maintenant a été proposé par Giuseppe Riccardi des Laboratoires AT&T Bell [2]. C'est un modèle stochastique, de type back-off à longueur de contexte variable (Variable N-gram Stochastic Automaton).

Les états du modèle sont associés avec les séquences de mots, de longueurs variables, observées dans l'ensemble d'entraînement. Ces séquences sont de longueur $n-1$ au maximum, n étant l'ordre du modèle. Supposons que l'on se trouve dans l'état $m_1, \dots, m_{i-1}$ ($i \leq n$), c'est-à-dire que les mots $m_1, \dots, m_{i-1}$ ont été rencontrés et que le mot suivant est m_i . Si la probabilité conditionnelle $P(m_i | m_1, \dots, m_{i-1})$ est connue, alors elle sera employée. Sinon, on réduira le contexte de l'état à $m_2, \dots, m_{i-1}$, et on déterminera si $P(m_i | m_2, \dots, m_{i-1})$ existe, et ainsi de suite jusqu'à ce qu'il n'y ait plus aucun contexte ou que le mot m_i , n'ayant jamais été rencontré dans l'ensemble d'entraînement, on emploie une probabilité par défaut.

Le modèle VNSA apporte une nuance au back-off classique. Si la probabilité $P(m_i | m_1, \dots, m_{i-1})$ est connue, alors elle est employée, mais le contexte de l'état suivant peut être $m_1, \dots, m_{i-1}, m_i$ (la taille du contexte augmente) ou $m_2, \dots, m_{i-1}, m_i$ (la taille du contexte reste constante), selon ce qui est globalement le plus avantageux, c'est-à-dire qu'on choisit d'augmenter ou de ne pas augmenter la taille du contexte afin de maximiser la probabilité de la séquence, $P(M)$. Le tableau 1 résume les types de transitions possibles d'un état du modèle vers l'état suivant.

Tableau 1 : Types de transitions possibles

	contexte de l'état actuel	Mot courant	contexte de l'état suivant	Mot suivant
Type 1	$m_1, \dots, m_{i-1}$	m_i	$m_1, \dots, m_{i-1}, m_i$	m_{i+1}
Type 2	$m_1, \dots, m_{i-1}$	m_i	$m_2, \dots, m_{i-1}, m_i$	m_{i+1}
Type 3	$m_1, \dots, m_{i-1}$	m_i	$m_2, \dots, m_{i-1}$	m_i

Dans une transition de type 1 ou 2, le mot courant est lu. La taille du contexte de l'état suivant peut être augmentée par le mot courant (type 1) ou rester la même si on éjecte le mot le plus ancien du contexte actuel (type 2). Les transitions de type 1 ou 2 ne sont pas toujours possibles, contrairement à la transition de type 3, où on ne lit pas le mot courant et où la taille du contexte est diminuée.

Reprenons notre exemple: le mot du contexte actuel est "petite" et les mots suivants sont "vie nous", on peut représenter les transitions schématiquement comme sur la figure 1:

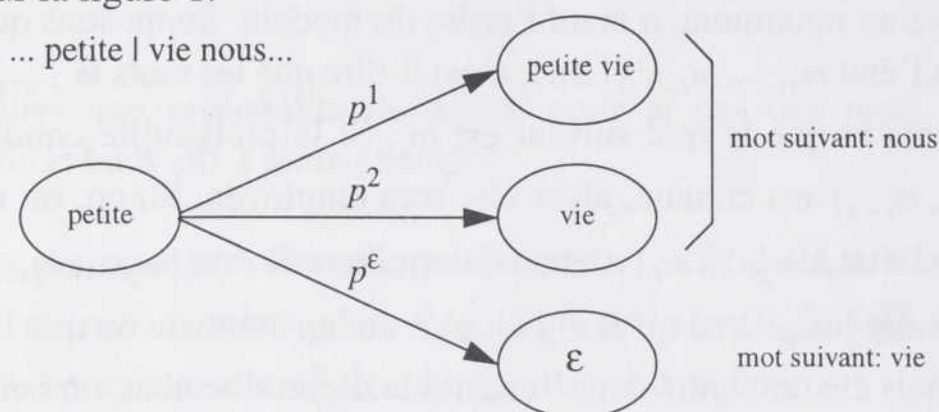


Figure 1 Schéma des transitions possibles pour les mots "petite" et "vie nous"

La probabilité de perdre un contexte peut être vue comme la probabilité de rencontrer un mot nouveau suite à un contexte connu. Ce nouveau mot peut faire partie du vocabulaire et tout simplement n'avoir jamais été rencontré suite au contexte présent. Pour chaque réduction de contexte, une "pénalité" sera attribuée. Il est également possible que le mot nouveau ne fasse pas partie du vocabulaire: le contexte sera alors réduit successivement jusqu'à zéro, et une valeur de probabilité par défaut sera attribuée à ce mot nouveau.

Si on représente maintenant une phrase selon tous les "chemins" possibles de décomposition, on obtient le treillis de la figure 2. La vraisemblance de cette phrase nous sera donnée par la décomposition qui maximise la probabilité. Pour trouver cette meilleure décomposition et sa probabilité, nous

utilisons l'algorithme de Viterbi.

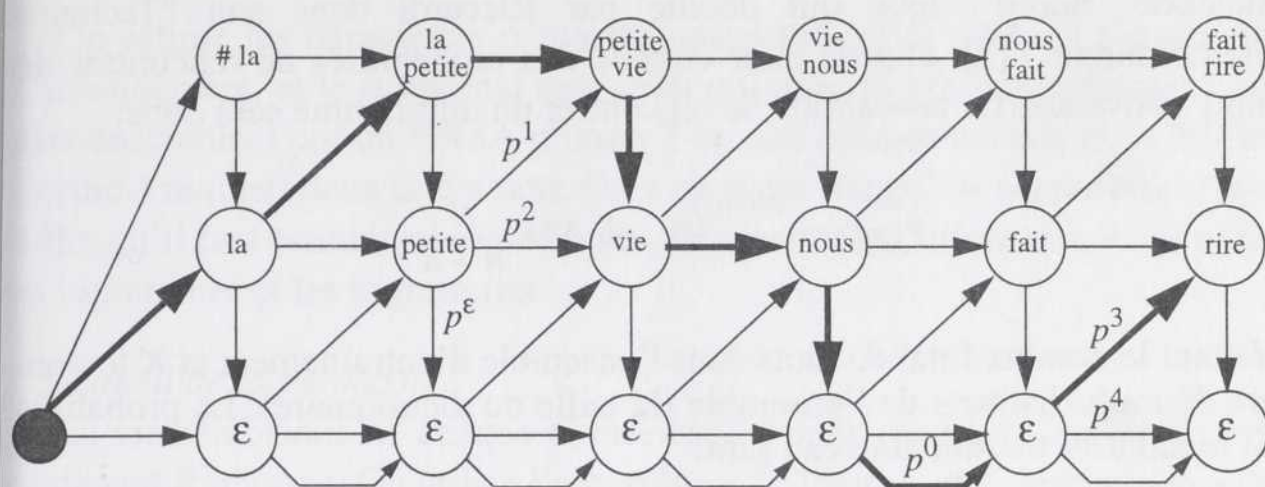


Figure 2 Tous les chemins possibles pour décomposer la phrase "la petite vie nous fait rire", avec un VNSA d'ordre 3.

Supposons que le meilleur chemin trouvé par Viterbi est la séquence d'états: "la" - "la petite" - "petite vie" - "vie" - "nous" - "ε" - "ε" - "rire". La probabilité selon ce chemin sera donc:

$$P(M) = P(la) \cdot \alpha'' P(petite|la) \cdot P(vie|la, petite) \cdot P(\epsilon|petite, vie) \\ (1 - \alpha'') \cdot P(nous|vie) \cdot P(\epsilon|nous) \cdot P(default) \cdot \alpha' P(rire).$$

Cette décomposition met en évidence les paramètres α du modèle, paramètres qui différencient p^1 de p^2 et p^3 de p^4 . Ce sont ces paramètres, variant entre 0 et 1 inclusivement, qui nous indiquent laquelle des options est la plus avantageuse entre augmenter la taille du contexte ou la garder constante. On peut en déduire de façon générale que:

$$p^1 = \alpha'' P(vie|petite) \text{ et } p^2 = (1 - \alpha'') P(vie|petite)$$

de la même façon:

$$p^3 = \alpha' P(rire|\epsilon) \text{ et } p^4 = (1 - \alpha') P(rire|\epsilon)$$

Le mot "fait" n'ayant jamais été rencontré dans l'ensemble d'entraînement, on utilisera p^0 , la probabilité par défaut.

2.2 Estimation des probabilités et des paramètres α

Nous utilisons la méthode du maximum de vraisemblance pour estimer les

probabilités de transition d'un état à l'autre, que nous pondérons par la méthode "add-c", telle que décrite par Riccardi dans son "Technical Memorandum" [3], afin de tenir compte des probabilités de rencontrer des mots nouveaux. La probabilité de rencontrer un unigramme sera donc:

$$\hat{P}(m_i) = \frac{N(m_i)}{N} \beta, \text{ où } \beta = \frac{N}{N+K} \quad (11)$$

N étant le nombre total de mots dans l'ensemble d'entraînement et K le nombre de mots distincts de l'ensemble (la taille du dictionnaire). La probabilité de rencontrer un mot nouveau sera:

$$\hat{P}(\theta) = 1 - \beta = \frac{K}{N+K} \quad (12)$$

On constate que plus K sera grand, en faisant l'hypothèse que N est constant, plus grande sera la probabilité de rencontrer un mot nouveau, n'appartenant pas au vocabulaire de l'ensemble d'entraînement.

On applique le même raisonnement aux n -grammes d'ordre supérieur et on obtient les estimations suivantes:

$$\hat{P}(m_i | m_{i-n+1}, \dots, m_{i-1}) = \frac{N(m_{i-n+1}, \dots, m_{i-1}, m_i)}{N(m_{i-n+1}, \dots, m_{i-1}, \Psi) + K(m_{i-n+1}, \dots, m_{i-1}, \Psi)} \quad (13)$$

$$\hat{P}(\theta | m_{i-n+1}, \dots, m_{i-1}) = \frac{K(m_{i-n+1}, \dots, m_{i-1}, \Psi)}{N(m_{i-n+1}, \dots, m_{i-1}, \Psi) + K(m_{i-n+1}, \dots, m_{i-1}, \Psi)} \quad (14)$$

$K(m_{i-n+1}, \dots, m_{i-1}, \Psi)$ est le nombre de n -grammes distincts, commençant par les mêmes mots $m_{i-n+1}, \dots, m_{i-1}$. θ représente ici un mot qui appartient au vocabulaire de l'ensemble d'entraînement, mais jamais rencontré à la suite d'un contexte connu, ou un mot nouveau n'appartenant pas à l'ensemble.

Pour estimer ces probabilités, on utilise un très grand ensemble d'entraînement. Cet ensemble est divisé en deux parties: 90% des mots servent à estimer les probabilités de transition, les 10% restant servent à estimer les

paramètres α .

On estime les paramètres α par tâtonnements sur le 10% de l'ensemble d'entraînement, et le α optimal sera celui qui minimisera la perplexité de ce sous-ensemble. Pour un VNSA d'ordre 2 on doit estimer un seul α , le VNSA d'ordre 3 requiert deux α distincts. On a un α par "étage" de n-gramme, c'est-à-dire qu'il faut estimer α' entre les unigrammes et les bigrammes et α'' entre les bigrammes et les trigrammes.

2.3 Calcul de la perplexité

La vraisemblance de chaque phrase est calculée sur un ensemble de test contenant R phrases. On utilise l'algorithme de Viterbi pour déterminer cette vraisemblance: on trouve le meilleur "chemin" possible à travers le treillis de toutes les possibilités de décompositions d'une phrase. Supposons qu'il existe Q chemins différents pour chaque phrase M_r et qu'à chaque chemin est associée une vraisemblance, on choisira la vraisemblance de cette phrase:

$$P(M_r) = \max\{P_1, P_2, \dots, P_Q\} \quad (15)$$

On calcule la vraisemblance totale de l'ensemble de test comme le produit des vraisemblances de chacune des R phrases de cet ensemble:

$$P(M_{test}) = \prod_{i=1}^R P_i \quad (16)$$

Finalement, on estime la perplexité de l'ensemble de test à l'aide de (10).

3. Expériences et résultats

3.1 Base de données

Nous avons appliqué notre modèle VNSA à la base de données ATIS 1994. Cette base est composée de phrases, en anglais, dont le sujet est une demande de renseignements aéroportuaires à propos des vols, des horaires, des services, des villes desservies, etc. Un exemple: "I'd like to book a flight from Cincinnati to New York City on United Airlines for next Saturday. "

3.2 Entraînement

Notre ensemble d'entraînement compte 208K mots et 21K phrases. Il a été

divisé en deux parties: 187K mots et 19K phrases pour estimer les probabilités de transition, et 21K mots et 2K phrases pour estimer les α optimaux. La taille du dictionnaire est 1,5K mots. La probabilité de rencontrer un mot nouveau dans notre ensemble de test, jamais vu à l'entraînement, est estimée à $7,75 \times 10^{-3}$ selon (10).

3.3 Test

Le corpus de test compte 10K mots et 1K phrases. Seulement 13 mots nouveaux apparaissent dans le texte de test, donc 99.87% des mots de l'ensemble de test apparaissent au dictionnaire d'entraînement.

Le tableau 2 présente les résultats de perplexité obtenus sur l'ensemble de test avec un VNSA d'ordre 1 (simple multiplication des probabilités de chaque unigramme), d'ordre 2 et d'ordre 3, avec les valeurs optimales des α .

Tableau 2 : Perplexité de l'ensemble de test en fonction de l'ordre du VNSA

	1	2	3
ATIS94.TEST	194.7	21.1	17.4
α optimaux	-	$\alpha = 0.96$	$\alpha' = 0.96$ et $\alpha'' = 0.09$

Nous avons également appliqué le modèle classique d'interpolation à notre ensemble de test, pour fins de comparaison. Les résultats obtenus sont présentés au tableau 3. Peu importe l'ordre, la probabilité de défaut pour un mot nouveau, λ_0 , est de $7,75 \times 10^{-3}$, comme pour le modèle VNSA:

Tableau 3 : Perplexité de l'ensemble de test avec le modèle d'interpolation

	1	2	3
ATIS94.TEST	194.7	21.8	18.7
λ optimaux	-	$\lambda_2 = 0.91, \lambda_1 = 0.09$	$\lambda_3 = 0.38, \lambda_2 = 0.53$ $\lambda_1 = 0.09$

4. Conclusion

Les résultats obtenus avec le modèle VNSA sont supérieurs à ceux qu'on obtient avec le modèle classique d'interpolation. Pour les deux modèles, l'ordre 1 signifie que l'on emploie (4) pour estimer la perplexité, il est donc

normal d'obtenir le même résultat puisque les probabilités des unigrammes sont estimées de la même façon pour les deux modèles.

C'est avec un modèle d'ordre 3 dans les deux cas qu'on obtient le meilleur résultat. Plus on allonge la fenêtre d'observation du passé, meilleure est notre capacité à prédire le mot suivant. Il serait intéressant de voir si un modèle d'ordre 4 pourrait faire une meilleure prédiction, même si les deux modèles d'ordre 3 accordent plus d'importance à la probabilité des bigrammes (λ_2 et α' sont dominants). Les mots de la base de données ATIS semblent être fortement liés deux à deux, par exemple: "I'd like", "to book", "what is", ... Peut-être que la distribution des λ et α est dépendante du type de vocabulaire employé, ainsi que de la longueur optimale de la fenêtre d'observation du passé.

Notre probabilité de rencontrer un mot nouveau a été surévaluée. Nous prédisions observer environ 8 mots inconnus sur 1000 mots dans notre ensemble de test. Nous n'en avons observé que 13, sur les 10 000 mots de l'ensemble de test.

Nous travaillons maintenant à appliquer le modèle de langage VNSA à une base de données de mots en français, avec un vocabulaire beaucoup plus étendu. Nous projetons également d'étudier les performances de ce modèle dans une tâche de reconnaissance de la parole. La perplexité nous permet de comparer les modèles de langage entre eux, mais la valeur réelle d'un modèle sera mieux évaluée si on le soumet à la tâche qui nous intéresse.

5. Références

- [1] F. Jelinek, "Self-Organized Language Modeling for Speech Recognition", in Alex Waibel and Kai-Fu Lee, editors, *Readings in Speech Recognition*, pp. 450-506, Morgan Kaufmann, San Mateo, CA, 1990.
- [2] G. Riccardi, E. Bocchieri, R. Pieraccini, "Non Deterministic Stochastic Language Models for Speech Recognition", *Proc. ICASSP 1995*, pp. 237-240, Détroit.

The first of these is the fact that the
 results of the experiments are in
 general agreement with the
 results of the experiments of
 other workers. This is a
 very important point, as it
 shows that the results are
 not due to any special
 conditions of the experiment.
 The second point is that the
 results are in good agreement
 with the results of the
 experiments of other workers.
 This is also a very important
 point, as it shows that the
 results are not due to any
 special conditions of the
 experiment. The third point
 is that the results are in
 good agreement with the
 results of the experiments of
 other workers. This is also a
 very important point, as it
 shows that the results are
 not due to any special
 conditions of the experiment.

The fourth point is that the
 results are in good agreement
 with the results of the
 experiments of other workers.
 This is also a very important
 point, as it shows that the
 results are not due to any
 special conditions of the
 experiment. The fifth point
 is that the results are in
 good agreement with the
 results of the experiments of
 other workers. This is also a
 very important point, as it
 shows that the results are
 not due to any special
 conditions of the experiment.

The sixth point is that the
 results are in good agreement
 with the results of the
 experiments of other workers.
 This is also a very important
 point, as it shows that the
 results are not due to any
 special conditions of the
 experiment. The seventh point
 is that the results are in
 good agreement with the
 results of the experiments of
 other workers. This is also a
 very important point, as it
 shows that the results are
 not due to any special
 conditions of the experiment.

The eighth point is that the
 results are in good agreement
 with the results of the
 experiments of other workers.
 This is also a very important
 point, as it shows that the
 results are not due to any
 special conditions of the
 experiment.

The ninth point is that the
 results are in good agreement
 with the results of the
 experiments of other workers.
 This is also a very important
 point, as it shows that the
 results are not due to any
 special conditions of the
 experiment.

The tenth point is that the
 results are in good agreement
 with the results of the
 experiments of other workers.
 This is also a very important
 point, as it shows that the
 results are not due to any
 special conditions of the
 experiment.

Imagerie biomédicale

Numé
morph

1. Potvi
1. Ecole
2. Centre
3. Ecole

La recons
numérissu
sur diver
reconstru
numérissu
proposé
morpholo
données
les dimen
des num
connaiss
la longue
d'intérêt
la définit
à l'impla
eaux. C
précision

1. Introdu

La scoli
comme
la scoli
Pour pa
mesure
l'hôpital
trois dir
et post
repères
et ligne
Présent
opérate

Numérisation d'images radiologiques par segmentation morphologique

J. Potvin ^(1,2), J. A. de Guise ^(1,2,3), Y. Ross ^(2,3), J. Dansereau ^(1,2)

1. École Polytechnique de Montréal, Montréal, Québec

2. Centre de recherche de l'hôpital Ste-Justine, Montréal, Québec

3. École de technologie supérieure, Montréal, Québec

La reconstruction tridimensionnelle de la colonne vertébrale de patients scoliotiques exige la numérisation d'un grand nombre de repères anatomiques (plateaux vertébraux et pédicules) sur diverses vues radiographiques (0°, 20° et latérale). Ce volet du processus de reconstruction est extrêmement long, fastidieux et sujet à erreurs. Afin d'optimiser l'étape de numérisation de radiographies, l'utilisation d'un traitement d'images assisté par ordinateur est proposé. Nous préconisons dans notre approche l'emploi d'outils de segmentation morphologique assisté d'informations issues d'une base de connaissance. L'accès à la base de données de l'Hôpital Ste-Justine a permis la réalisation de modèles statistiques représentant les dimensions et les positions des repères anatomiques dans l'image radiologique; les résultats des numérisations manuelles de plus de 200 radiographies furent mis à contribution. Une connaissance *a priori* de la position des plateaux vertébraux et des pédicules en fonction de la longueur de la colonne vertébrale permet de limiter l'application des algorithmes aux zones d'intérêts de l'image. De plus, l'information concernant les dimensions des vertèbres permet la définition de marqueurs interne et externe à l'objet d'intérêt. Ces marqueurs sont essentiels à l'implantation des algorithmes de segmentation morphologique par lignes de partage des eaux. Ce traitement informatique des images radiologiques permettra d'augmenter la précision des numérisations et d'en automatiser la réalisation.

1. Introduction

La scoliose est une anomalie de la colonne vertébrale qui a longtemps été définie comme une déformation latérale du rachis. Depuis quelques années, on aborde la scoliose comme une déformation tridimensionnelle de la colonne vertébrale. Pour parvenir à traiter de façon adéquate une telle déformation, il faut être en mesure de l'évaluer dans l'espace 3D. Le groupe de recherche sur la scoliose de l'hôpital Ste-Justine a développé des outils qui permettent une représentation en trois dimensions du rachis scoliotique. A partir d'images radiologiques (latérale et postéro-anérieure à 0° et 20°) de la colonne vertébrale, on numérise des repères anatomiques (centres des plateaux vertébraux, sommets des pédicules et lignes médianes des côtes) qui servent à la reconstruction tridimensionnelle. Présentement l'extraction des repères anatomiques s'effectue manuellement: un opérateur visualise les radiographies et marque les positions des repères à l'aide

d'un pointeur. Cette étape peut exiger jusqu'à trois heures de travail pour un seul patient.

Tirer avantages des méthodes d'imagerie afin de permettre l'extraction des repères anatomiques plus rapidement et plus précisément constitue l'objectif de ce projet. L'efficacité de la méthode de segmentation morphologique par ligne de partage des eaux a déjà été démontrée sur les pédicules des vertèbres de colonnes vertébrales scoliotiques [1]. L'utilisation de cette méthode de segmentation nécessite l'imposition de marqueurs interne et externe à l'objet d'intérêt. L'objectif spécifique de ce travail est de segmenter les dix-sept vertèbres utilisées pour les reconstructions tridimensionnelles des rachis scoliotiques. Ce projet se divise en deux problématiques distinctes: la segmentation de l'image et la définition des marqueurs dans l'image radiologique.

2. Méthodologie

La première partie de cette section présente la méthode de segmentation morphologique par ligne de partage des eaux et démontre l'intérêt d'utiliser des marqueurs afin d'optimiser les résultats de la segmentation dans notre contexte particulier. La seconde partie propose une méthode d'estimation des positions des centres des plateaux vertébraux dans l'image radiologique en limitant l'interaction de l'opérateur à l'introduction de quelques points spécifiques dans l'image. La connaissance des positions des centres des plateaux permet la création des marqueurs nécessaires à la segmentation.

2.1 Méthode de segmentation

La morphologie mathématique propose une méthode de segmentation dans laquelle l'on considère l'image en niveaux de gris comme une topographie, chaque niveau de gris est perçu comme une élévation du relief terrestre. La segmentation par ligne de partage des eaux est réalisée en considérant les zones d'intensité minimum d'une image comme des bassins vers lesquels convergent toute les eaux. La simulation de l'immersion de la représentation topographique de l'image permet de comprendre le fonctionnement de l'algorithme de ligne de partage des eaux. On effectue donc une immersion progressive du relief, voire de l'image, au fur et à mesure que se produit l'immersion, l'eau s'accumule dans les bassins et des barrages sont érigés aux endroits où l'eau communique d'un bassin à un autre (figure 1). A la fin du processus chacun des bassins de l'image

est cerné d'une ligne de partage des eaux. Cette procédure produit des contours fermés autour des minimums.

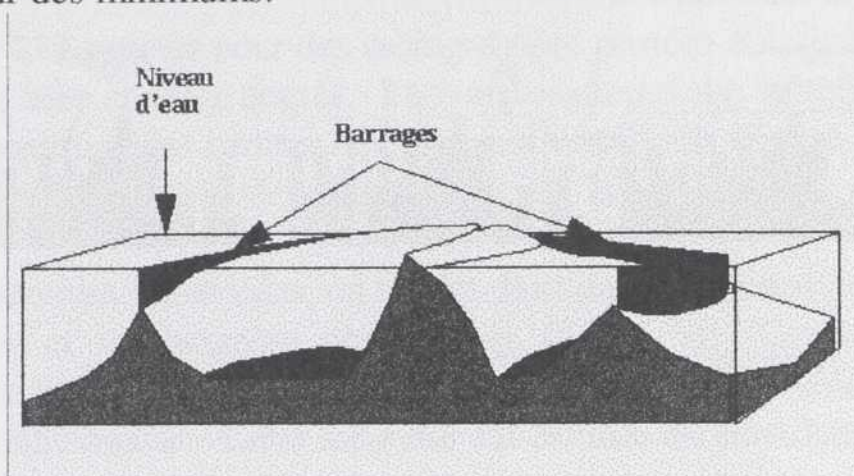


Figure 1. Simulation du processus d'immersion

Dans notre application les contours d'intérêts sont les frontières du corps vertébral dans la projection radiographique. L'application pure et simple d'un algorithme de ligne de partage des eaux à une image de vertèbre produit une segmentation excessive de l'image. Comme le démontre la segmentation des figures 2 et 3, tous les minimums locaux de l'image correspondent à des bassins versants et sont cernés de lignes de partage des eaux.



Figure 2 Image originale

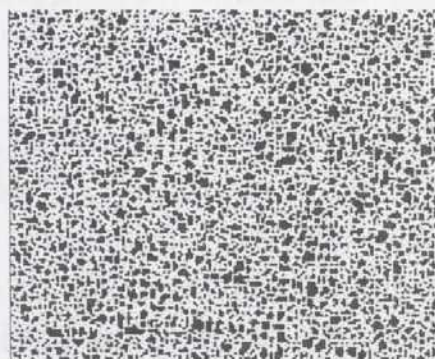


Figure 3 Lignes de partage des eaux

La littérature propose une solution à ce problème, soit l'imposition de minimums dans l'image[2][3]. On génère une image de marqueurs, dans laquelle on identifie les zones des minimums d'intérêts de l'image, c'est-à-dire dans notre contexte: un marqueur est placé à l'intérieur du corps vertébral et un autre à l'extérieur.

Une reconstruction par érosion de l'image des marqueurs permet d'éliminer les minimums locaux contenus dans l'image originale et de limiter la segmentation

à l'objet d'intérêt[4]. La figure 4 présente le principe de la reconstruction par érosion sur une fonction à une dimension.

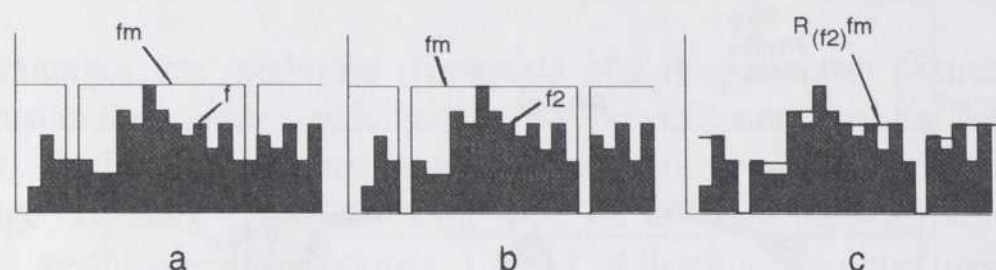


Figure 4 Reconstruction par érosion d'une fonction à une dimension

A partir de la fonction de marqueurs et de la fonction à segmenter (figure 4.a) on construit la fonction f_2 qui correspond à l'union des valeurs minimums de ces deux fonctions (figure 4.b). Par la suite, on réalise l'érosion de la fonction de marqueurs conditionnellement à la fonction f_2 , c'est-à-dire qu'on contraint la fonction reconstruite à être plus grande ou égale à la fonction f_2 . Cette nouvelle fonction est la reconstruction par érosion de la fonction de marqueurs, $R_{(f_2)}f_m$ (figure 4.c). Le principe de reconstruction s'applique par extension aux fonctions de deux dimensions. Si l'on applique l'algorithme de ligne de partage des eaux sur une image reconstruite, seulement les frontières de l'objet d'intérêt sont segmentées.

La méthodologie présentée jusqu'ici est directement proposée dans la littérature par différents auteurs. Plusieurs implémentations d'algorithmes d'extraction de contours par lignes de partage des eaux sont suggérées dans diverses publications. Nos outils de segmentation sont construits à partir des algorithmes proposés par Bleau et al.[5][6].

Le défi à cette étape est de définir les marqueurs dans l'image radiologique afin d'en extraire les contours des vertèbres. La section suivante présente la particularité de ce projet, soit la méthode de définition des marqueurs dans l'image radiologique.

2.2 Modèles Statistiques

L'existence d'une base de données contenant les résultats des numérisations des repères anatomiques motive la création d'une base de connaissances sur leurs positions dans l'image. L'utilisation de l'information contenue dans cette base de

connaissances permettra de générer les marqueurs dans l'image radiologique. La base de données utilisée est constituée des résultats des numérisations manuelles effectuées sur 222 patients pour des radiographies postéro-anterieur prises sous des angles de zéro et vingt degrés. La numérisation d'une vertèbre consiste à localiser les positions des centres des plateaux vertébraux (figure 5, points 1 et 6) et les sommets des pédicules (figure 5, points 2,4 et 3,5).

A l'aide des positions numérisées, on extrait dans un premier temps les hauteurs des vertèbres et les hauteurs des disques (figure 5). Ces mesures sont normalisées en fonction de la longueur de la projection de la colonne vertébrale dans l'image radiologique. Cette longueur est estimée en cumulant les hauteurs des vertèbres et des disques. Dans un second temps, on extrait les positions des centres des pédicules gauche et droit à l'aide de l'information relative à la position de leurs sommets (figure 5). Ces mesures sont dans ce cas normalisées en fonction de la hauteur de la vertèbre.

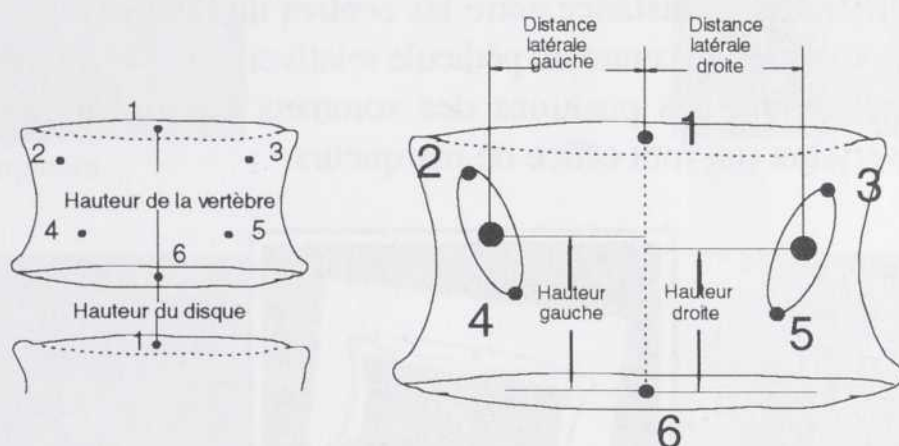


Figure 5 Extraction des mesures

Un modèle statistique est obtenu en effectuant la somme des moyennes des hauteurs des disques et des hauteurs des vertèbres pour chacun des niveaux vertébraux. Si on limite l'information introduite par un opérateur à la longueur de la ligne médiane de la projection de la colonne vertébrale dans le plan radiologique, le modèle statistique permet d'estimer les positions des centres des plateaux vertébraux le long de cette ligne médiane. La figure 6 illustre le fonctionnement du modèle. Sur l'image de la colonne vertébrale, l'opérateur introduit quelques points sur la ligne médiane du rachis, (représenté par les x sur la figure 6). Une attention particulière doit être apportée à la localisation du centre du plateau inférieur de la dernière vertèbre lombaire et à celle du centre du plateau supérieur de la première vertèbre thoracique, car le modèle est

constitué de mesure normalisées en fonction de la longueur de la colonne vertébrale. A l'aide du modèle statistique, on effectue une interpolation des positions des centres des plateaux vertébraux (représenté par les points sur la figure 6) le long de la courbe définie par les points introduits par l'opérateur.

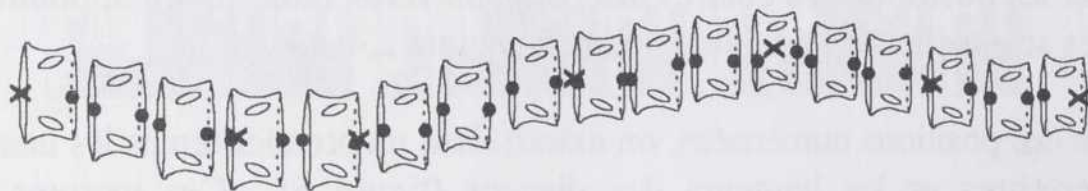


Figure 6 Estimation des centres des plateaux vertébraux

Connaissant les positions des centres des plateaux vertébraux on peut définir des marqueurs qui sont situés à l'intérieur et à l'extérieur de la vertèbre. On illustre, à la figure 7, l'exemple de la définition des marqueurs d'une vertèbre à partir des positions des centres des plateaux vertébraux. En se basant sur la hauteur de la vertèbre, estimée par la distance entre les centres de ses plateaux vertébraux, et sur la position latérale du centre du pédicule relativement à la ligne médiane de la vertèbre, on définit les positions des sommets des polygones interne et externe à la vertèbre qui font office de marqueurs.

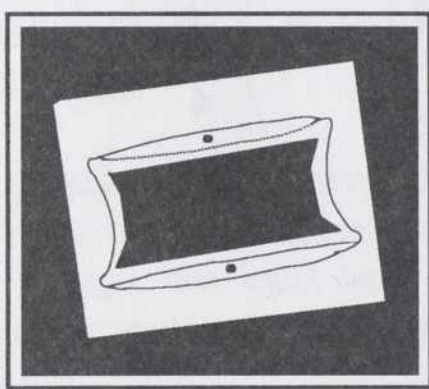


Figure 7 Définition des marqueurs

2.3 Évaluation de la méthode

On évalue la validité de la méthode proposée en trois temps, premièrement on vérifie la qualité des segmentations réalisées à l'aide de la méthode de ligne de partage des eaux en appliquant l'algorithme à partir de marqueurs définis manuellement. Deuxièmement, les positions estimées à l'aide du modèle statistique sont comparées aux positions numérisées par un opérateur. Par la

colonne
on des
sur la
teur.

Ensuite, la méthode de segmentation est évaluée dans son ensemble, soit en utilisant les modèles statistiques pour générer les marqueurs dans l'image.

3 Résultats

L'application de l'algorithme de ligne de partage des eaux à des images permet d'évaluer qualitativement le résultat de la segmentation morphologique. Les figures 8 et 10 illustrent la définition manuelle des marqueurs tandis que les résultats de la segmentation sont présentés aux figures 9 et 11.

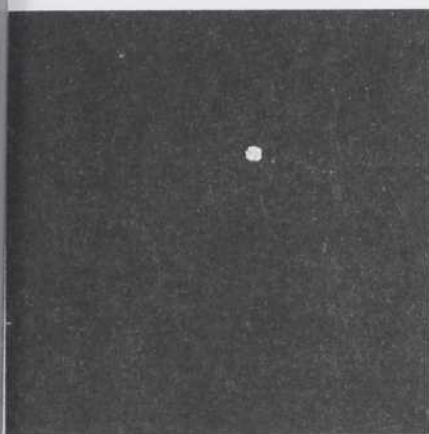


Figure 8 Marqueurs

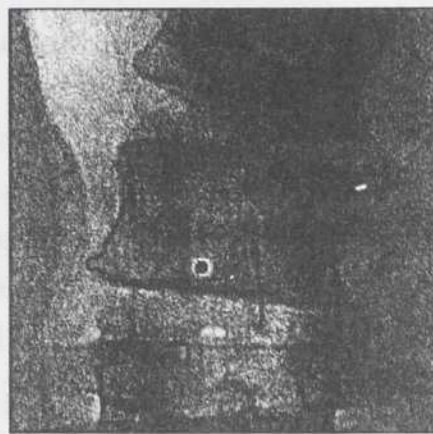


Figure 9 Segmentation

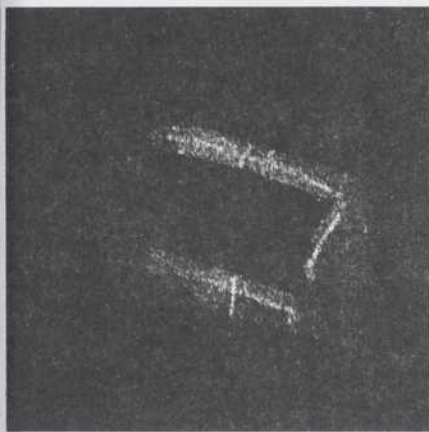


Figure 10 Marqueurs

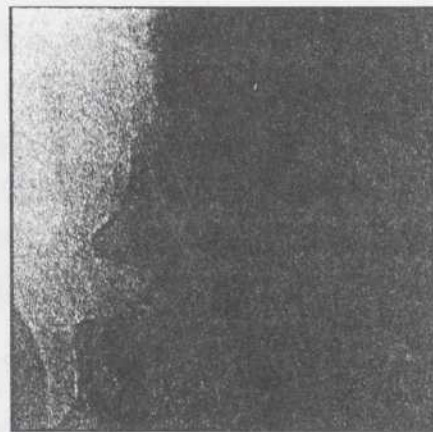


Figure 11 Segmentation

ent on
ne de
éfinis
modèle
Par la

L'application des modèles à 10 images numérisées manuellement à 0 et 20 degrés permet d'évaluer l'erreur de positionnement inhérent aux modèles statistiques en comparant les positions des repères estimés par les modèles aux positions numérisées par un opérateur. Le tableau 1 donne l'erreur moyenne de positionnement des centres des plateaux vertébraux exprimée en pourcentage de la longueur de la colonne vertébrale.

Tableau 1: Erreur moyenne de positionnement des niveaux vertébraux

Image	Moyenne	Écart-type	Maximum	Minimum
0°	1.20%	0.47%	1.93%	0.25%
20°	1.31%	0.64%	2.35%	0.17%

Les figures de la dernière page présentent le résultat d'une segmentation effectuée en utilisant le modèle statistique pour définir les marqueurs. La figure 12 correspond à l'image originale et la figure 13 au résultat de la segmentation.

4. Discussion et conclusion

Dans l'ensemble la segmentation morphologique par ligne de partage des eaux est relativement précise lorsque l'on impose les marqueur manuellement. On note cependant que la présence de structures à l'intérieur de la vertèbre tend à dévier les lignes de partage des eaux. En conséquence, les surface des marqueurs internes doivent être les plus grandes possibles car l'intérieur du corps vertébrale dans la projection radiographique contient plusieurs structures telles les pédicules, les apophyses, etc., qui peuvent nuire à la segmentation des frontières du corps vertébral.

Il en découle que la précision des modèles statistiques s'avère insuffisante pour permettre une segmentation adéquate des vertèbres. On observe pour plusieurs niveaux vertébraux de la figure 13 que les lignes de partage des eaux correspondent aux frontières des vertèbres adjacentes. La définition d'un marqueur interne de grande surface provoque régulièrement le chevauchement du marqueur et de la frontière de la vertèbre à segmenter.

Cependant l'utilisation de ce genre de modèle prédictif n'est pas sans intérêt. Dans notre contexte, les informations pertinentes à la segmentation des vertèbres sont limitées dans l'image; si l'on considère qu'une image radiographique est constituée de 2140 lignes par 880 colonnes de pixels, on peut évaluer approximativement la surface occupés par le rachis à 10 % de la surface totale de l'image. Peu importe la méthode utilisée pour segmenter l'image, il est extrêmement avantageux d'être en mesure de prédire la position des zones d'intérêts dans l'image.

En conclusion, l'utilisation spécifique de la méthode de lignes de partage des eaux assistée de marqueurs n'est pas applicable dans sa version actuelle. On peut tout de même envisager d'améliorer son utilisation soit: en augmentant la précision des modèles statistiques ou en effectuant un post-traitement sur les résultats des segmentation des vertèbres.

Tout compte fait, il faut être conscient que les résultats issus de notre approche sont les meilleurs obtenus jusqu'à ce jour pour la résolution de cette problématique spécifique.

5. Références

- Y. Ross, "Segmentation des structures osseuses d'une colonne vertébrale par les techniques de la morphologie mathématique", rapport interne école de technologie supérieure, 1995, 139 pages
- S. Beucher, "Segmentation Tools in Mathematical Morphology", SPIE vol. 1350 Image Algebra and Morphological Processing. 1990, pp.70-84
- J.F. Rivest, P. Soille, "Principles and Applications of Morphological Image Analysis", IAPR 11th IAPR International Conference on Pattern Recognition, 1992.
- L. Vincent, "Morphological Grayscale Reconstruction in Image Analysis: Applications and Efficient Algorithms", IEEE Transaction on Image Processing, vol. 2 No. 2, 1992, pp. 176-201
- A. Bleau, J. A. de Guise, A.-R. LeBlanc, "A New Set of Fast Algorithms for Mathematical Morphology: I. Idempotent Geodesic Transforms", CVGIP: Image Understanding, vol. 56, No. 2, 1992, pp. 178-209
- A. Bleau, J. A. de Guise, A.-R. LeBlanc, "A New Set of Fast Algorithms for Mathematical Morphology: II. Identification of Topographic Features on Grayscale Images", CVGIP: Image Understanding, vol. 56, No. 2, 1992, pp. 210-229

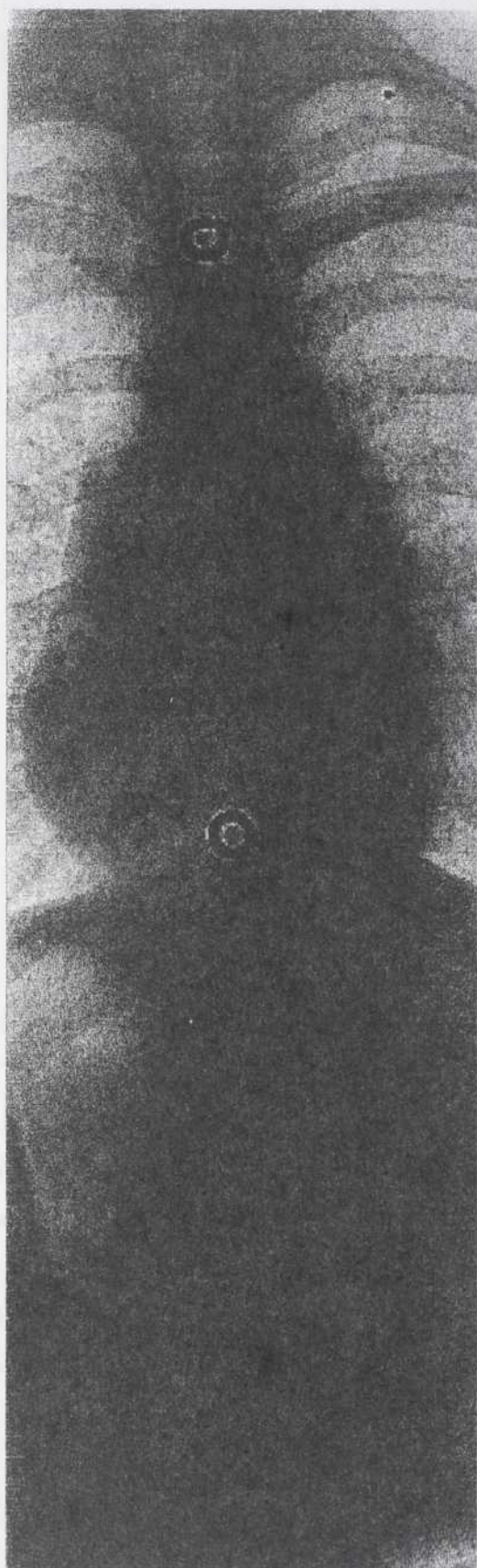


Figure 12 Image originale

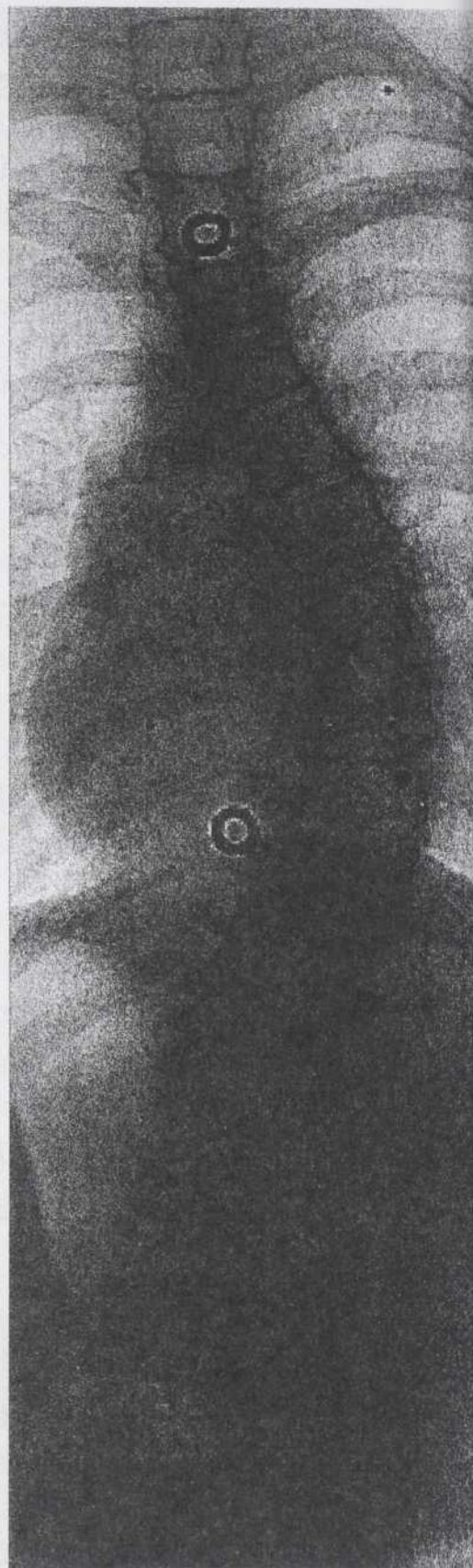


Figure 13 Image segmentée

Un sy

Bo

E-

Résum

E

(3-D) ut
reconstr
nombre
utilisées
traiteme
complex
des obje
nouvelle
modèle
modélis
avec l'in
sous-obj
à un ens
projecti
appariée
avec de
permett

1. Intr

E

pour le
ensembl
de vue
méthod
à phot
transfo
ou bie

Un système structurel de vision pour la reconnaissance d'objet dans des images radiographiques

Hail Mallouche, Jacques A. de Guise[†], et Yves Goussard

*Institut de génie biomédical, École polytechnique de Montréal
Boîte postale 6079, Succ. "Centre-Ville," Montréal, Québec, Canada, H3C 3A7*

[†]Génie de la production automatisée, École de technologie supérieure

1100, Notre-Dame West, Montreal, Quebec, Canada, H3C 1K3

E-mail: hail@grbb.polymtl.ca; deguise@gpa.etsmtl.ca; yves@grbb.polymtl.ca

Résumé

En imagerie médicale, les méthodes usuelles de reconstruction tridimensionnelle (3-D) utilisent un ensemble représentatif de projections 2-D. Cependant, le problème de reconstruction se transforme progressivement en un problème inverse mal posé quand le nombre de projections est limité et l'angle de vue est restreint. Les méthodes d'inférence utilisées en vision artificielle abordent également la même problématique. De plus, le traitement des images radiographiques représente trois difficultés principales: (i) la complexité causée par la superposition des structures; (ii) la nature variante et déformable des objets; et (iii) le faible rapport Signal/Bruit. Dans cet article, nous présentons une nouvelle méthode de reconnaissance d'objet à partir d'images radiographiques basée sur un modèle à contours dynamiques. Cette approche est constituée de deux étapes: (i) la modélisation géométrique de l'objet; et (ii) l'appariement de la projection 2-D du modèle avec l'image. Nous supposons que: (i) chaque objet peut être modélisé par un ensemble de sous-objets reliés suivant un graphe relationnel intrinsèque; (ii) les sous-objets appartiennent à un ensemble de formes géométriques de type pseudo-cylindrique ou pseudo-conique. La projection dans le plan de l'image radiographique du modèle vérifiant ces hypothèses est appariée avec l'image observée en optimisant une fonction d'énergie. L'approche a été testée avec des images synthétiques d'objets simples. Les résultats sont encourageants et nous permettent de proposer l'approche pour le traitement d'images radiographiques réelles.

1. Introduction

En imagerie médicale, la reconstruction 3-D est de grande importance pour le diagnostic clinique [3]. Les méthodes de reconstruction utilisent un ensemble représentatif de projections 2-D récoltées à partir de différents points de vue autour de l'objet à examiner. On retrouve des applications de ces méthodes en tomographie axiale, en résonance magnétique et en tomographie à photon unique. Cependant, le problème inverse de reconstruction se transforme en un problème mal posé quand le nombre de projections est limité ou bien quand l'angle de vue est restreint [9]. Les méthodes d'inférence

utilisées en vision artificielle permettent d'introduire des informations additionnelles afin de stabiliser l'espace de solutions [5]. La reconstruction 3-D des vaisseaux coronariens ou de la colonne vertébrale à partir de deux vues en stéréo ou de deux vues orthogonales nécessite l'utilisation de ces méthodes [2][3].

Le présent travail s'inscrit dans un projet de développement d'un système de vision pour reconstruire en 3-D la colonne vertébrale à partir d'un nombre limité d'images radiographiques. Toutefois, la reconnaissance d'objet en vision artificielle peut se faire directement à partir des images 2-D ou après l'inférence de la troisième dimension [7]. Les méthodes d'inférence 3-D sont basées essentiellement sur l'appariement de primitives extraites des images issues de différents points de vue. La disparité résultante de cet appariement est utilisée pour calculer la position 3-D de chaque primitive. Cependant, la détection et l'extraction de primitives dans les images radiographiques du corps humain doivent surmonter trois difficultés majeures: (i) la complexité causée par la superposition des structures et la nature de la scène; (ii) la nature variante et déformable des objets; et (iii) le faible rapport Signal/Bruit.

Dans ce cadre, nous présentons une nouvelle méthode de reconnaissance d'objet à partir d'images radiographiques en utilisant une structure basée sur un modèle à contours dynamiques [1][6][13]. L'approche est constituée de deux étapes: (i) la modélisation géométrique de l'objet; et (ii) l'appariement de la projection 2-D du modèle avec l'image. Nous supposons que: (i) chaque objet peut être modélisé par un ensemble de sous-objets reliés suivant un graphe relationnel intrinsèque; (ii) les sous-objets appartiennent à un ensemble de formes géométriques de type pseudo-cylindrique ou pseudo-conique. Par conséquent, la projection semi-transparente en 2-D de chaque sous-objet peut être décrite par un contour extrémal fermé et un ensemble de contours fermés situés à l'intérieur du contour extrémal. Nous supposons aussi que le point de vue dont est issue la projection radiographique est connu approximativement. Le problème se formule par l'optimisation d'une fonction d'objectif constituée par la somme de trois termes [8]: (i) la mesure de l'adéquation de la solution avec l'image observée; (ii) la mesure de l'adéquation de la solution avec le modèle connu *a priori*; et (iii) les contraintes de lissage.

Cette méthode a été implantée puis utilisée pour la segmentation d'images synthétiques auxquelles différents niveaux de bruit gaussien ont été ajoutés. Les résultats préliminaires démontrent l'intérêt de l'approche et nous permettent de proposer cette méthode pour le traitement d'images radiographiques réelles.

2. Modélisation géométrique

Les approches de vision basées sur un modèle nécessitent généralement la définition d'un langage représentatif permettant de décrire la scène et ses objets constituant. Nous basons ce langage sur la description géométrique de l'objet 3-D connu *a priori*. Dans ce projet, on s'intéresse plus spécifiquement à la reconstruction 3-D des vertèbres humaines à partir de deux ou trois vues radiographiques prises sous des angles différents.

D'autre part, les informations structurelles décrivant la scène sont de grande importance pour la reconnaissance en vision humaine [4][12][15]. Le modèle est donc basé sur une représentation schématisée de la structure d'une vertèbre et de ses projections dans différents plans radiographiques. Nous

supposons que chaque objet peut être décrit par l'union de sous-objets appartenant à un ensemble de formes simples et élémentaires. La structure peut être décrite par un graphe relationnel où un sommet et un arc représentent respectivement un sous-objet et une connexion entre deux sous-objets. Les attributs des sommets et des arcs décrivent la géométrie des sous-objets et leurs inter-connexions. Pour modéliser une vertèbre, trois sous-objets de forme pseudo-cylindrique, pseudo-conique et pseudo-planaire ont été créés. Dans la figure 1, nous montrons des exemples de projections de ces trois sous-objets, sous un angle quelconque.

La projection radiographique de chaque sous-objet est représentée par un contour fermé extrémal et un ou deux contours fermés internes. Le contour extrémal délimite l'ensemble des régions formant la projection du sous-objet considéré. Les contours internes délimitent les projections des bases de la forme pseudo-cylindrique et de la forme pseudo-conique.

En utilisant ce modèle simplifié, une radiographie d'une vertèbre est composée de la superposition des projections d'au moins deux de

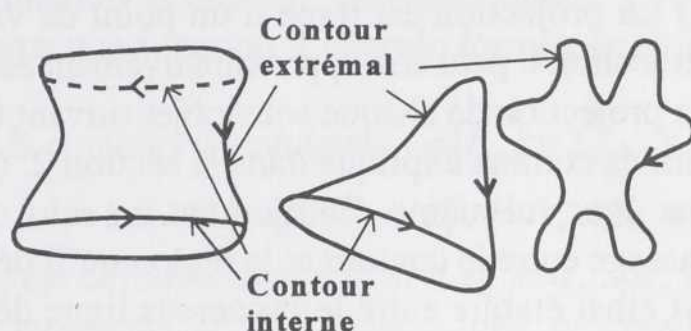


Figure 1. Projections sous un angle quelconque, des trois sous-objets constituant une vertèbre simplifiée.

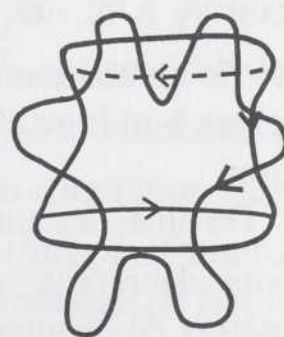


Figure 2. Radiographie d'un modèle simple d'une vertèbre comprenant deux sous-objets.

ces sous-objets (figure 2). La forme pseudo-cylindrique et la forme pseudo-planaire schématisent respectivement le corps vertébral et la structure apophysaire.

3. Représentation des connaissances

Nous notons par G_3 le graphe relationnel décrivant la structure 3-D de l'objet sous projection radiographique. Nous introduisons quatre hypothèses: (i) La projection est issue d'un point de vue non accidentel; (ii) L'angle de projection θ peut être approximativement estimé par un processus externe; (iii) La projection de chaque sous-objet suivant θ génère un ensemble de contours fermés comme expliqué dans la section 2; (iv) Un contour fermé ne se croise pas avec lui-même. En ajoutant un sens de balayage à chaque contour, le passage entre le contour et la région qu'il délimite devient intuitif. Une dualité est ainsi établie entre le processus ligne décrivant le contour et le processus région. Ces hypothèses permettent de décrire la projection de l'objet suivant θ à l'aide du graphe $G_{2,\theta}$ issue de G_3 . $G_{2,\theta}$ représente le graphe relationnel reliant les différentes régions définies par la superposition des différents processus régions.

Chaque sommet du graphe 2-D représente une région du modèle projeté [12]. Supposons que le modèle 2-D est défini par K contours fermés. Chaque région a_j identifie une combinaison des intérieurs superposés des K processus de région. Deux sommets a_i et a_j seront reliés par un arc b_{ij} s'ils ont un segment de contour $l_{k,ij}$ qui les sépare (figure 3). Nous notons que $l_{k,ij}$ appartient au contour l_k . Dans ce cas le passage de l'étiquette a_i à l'étiquette a_j se fait en ajoutant r_k à a_i . $G_{2,\theta}$ représente une importante

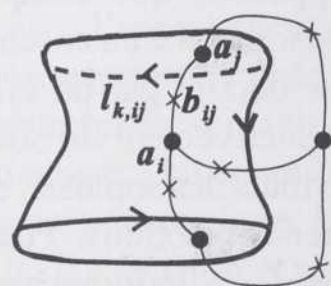


Figure 3. Le graphe 2-D d'un sous-objet pseudo-cylindrique.

source de connaissance *a priori* dérivant directement du modèle et pouvant participer à réduire l'espace de recherche et contraindre la recherche de la solution.

De plus, des attributs peuvent être ajoutés aux arcs du graphe $G_{2,\theta}$. Ces attributs décrivent, par exemple, la distribution statistique de certains paramètres de chaque segment de contour $l_{k,ij}$ identifié par l'arc b_{ij} de $G_{2,\theta}$. Soient s un site appartenant à $l_{k,ij}$ et ϕ_s l'angle de la tangente de $l_{k,ij,s}$. Nous désignons respectivement l'histogramme et l'histogramme filtré de ϕ_s par

$H_{b_{ij}}\{\phi_s, s \in l_{k,ij}\}$ et $H_{b_{ij},\varphi_\sigma} = \varphi_\sigma * H_{b_{ij}}$. σ mesure la variance de la forme et du positionnement du modèle. φ_σ est un filtre de lissage (gaussien, par exemple). $H_{b_{ij}}$ ou $H_{b_{ij},\varphi_\sigma}$ peuvent être associés à b_{ij} .

4. Critère à trois interactions

L'appariement du modèle 2-D avec l'image se fait en déformant le modèle en vue d'optimiser la fonction d'objectif correspondante. Cette fonction peut être interprétée comme une fonction d'énergie formée de trois termes [10].

Soit le modèle $M_{2,0}$ formé de K contours dynamiques $l = \{l_k, k=1, \dots, K\}$ définis sur K grilles discrètes $\{L_k^l, k=1, \dots, K\}$. Chaque $L_k^l, k=1, \dots, K$, est identique à la grille L^y sur laquelle est définie l'image y . Soit $r_k, k=1, \dots, K$, le processus de région associé au processus l_k . L'ensemble des processus $r = \{r_k, k=1, \dots, K\}$ est défini sur l'ensemble de grilles binaires $\{L_k^r, k=1, \dots, K\}$ où chaque $L_k^r, k=1, \dots, K$, est identique à L^y . Soit respectivement s et δs un site discret défini sur L^y et un système de voisinage centré sur s et défini sur L^y . La fonction d'énergie est donnée par la forme suivante:

$$E(l, r, y) = \sum_{k=1, \dots, K} \sum_{s \in l_k} [I_1(l_{k,\delta s}, y_{\delta s}) + I_2(l_{k,s}, r_{\delta s}) + I_3(l_{k,s}, l_{k,\delta s})] \quad (1)$$

La première sommation est définie sur les K contours, alors que la deuxième sommation représente l'intégrale curviligne sur chaque contour l_k . Nous soulignons que pour un site $s \in l_k$, les trois interactions I_1, I_2 , et I_3 sont fonctions de r, l , et y définis sur le voisinage δs . L'énergie globale est ainsi définie comme la somme des énergies locales à chaque site $s \in l_k, k=1, \dots, K$.

Le premier terme $I_1(l_{k,\delta s}, y_{\delta s})$ mesure l'interaction ou l'adéquation locale entre la solution locale $(l_{k,\delta s}, r_{k,\delta s})$ et l'image $y_{\delta s}$. I_1 permet d'ajuster les contours $\{l_k\}$ aux frontières de l'image y . I_1 est connu aussi comme la fonction de formation d'image. Nous pouvons formuler I_1 par une corrélation entre les gabarits d'intensité dérivant du modèle 2-D déformé et l'image. Toutefois, il est difficile et inefficace de produire ces gabarits car le processus de formation d'image dépend de la technique radiographique utilisée et des organes sous projection. Par conséquent, on choisit une mesure relative pour exprimer I_1 :

$$I_1(l_{k,\delta s}, y_{\delta s}) = \sum_{j=1, \dots, m} |\delta s^{(a_j)}| \left| \bar{y}(\delta s_{int(k)}^{(a_j)}) - \bar{y}(\delta s_{ext(k)}^{(a_j)}) \right| \quad (2)$$

où $\{a_j, j=1, \dots, m\}$ est l'ensemble des étiquettes de région définies sur $r_{-k, \delta s}$. $r_{-k} = \{r_1, \dots, r_{k-1}, r_{k+1}, \dots, r_K\}$ représente l'ensemble des processus des régions sans r_k . De plus, $\delta s^{(a_j)}$ désigne l'ensemble des sites de δs ayant l'étiquette a_j dans r_{-k} . Le processus ligne concerné, $l_{k, \delta s}$, produit une partition de chaque $\delta s^{(a_j)}$, $j=1, \dots, m$, en deux sous ensembles $\{\delta s_{int(k)}^{(a_j)}, \delta s_{ext(k)}^{(a_j)}\}$ désignant respectivement la partie interne et la partie externe de δs .

La deuxième fonction $I_2(l_{k,s}, r_{\delta s})$ représente la mesure d'interaction entre la solution et le modèle relationnel illustré par le graphe $G_{2,\theta}$. I_2 peut être obtenu directement des histogrammes ajoutés aux arcs de $G_{2,\theta}$. Pour cela, on remplace la paire $(l_{k,s}, r_{\delta s})$ par la paire $((l_k, a_j), \phi_s)$, où (l_k, a_j) est associé à l'arc correspondant b_{ij} de $G_{2,\theta}$, et ϕ_s est l'argument de la tangente de $l_{k,s}$. Cependant, l'échantillonnage direct de l'histogramme filtré H_{b_{ij}, ϕ_s} peut biaiser la solution vers ces connaissances *a priori* et compromettre ainsi la précision de la détection des frontières principalement dans les régions à faible rapport S/B. Une façon d'y remédier est de seuiller H_{b_{ij}, ϕ_s} en le transformant en une fenêtre trapézoïdale H_{b_{ij}, ϕ_s}^w . Ainsi, tous les α_s recevront un appui égal s'ils sont à l'intérieur de la fenêtre et une même pénalité dans le cas contraire. Ainsi, nous écrivons:

$$I_2(l_{k,s}, r_{\delta s}) = H_{b_{ij}, \phi_s}^w(\phi_s) \quad (3)$$

Ce terme mesure l'adéquation de la solution par rapport au modèle. Il sert à contraindre et à réduire l'espace des solutions possibles.

La troisième fonction $I_3(l_{k,s}, l_{k, \delta s})$ permet de régulariser l'espace de solutions en pénalisant les variations curvilignes des contours [14]. On choisit la somme pondérée de la longueur élémentaire $|l_{k,s}|$ du contour et des variations quadratiques de ϕ_s et de sa première dérivée. Cela donne:

$$I_3(l_{k,s}, l_{k, \delta s}) = |l_{k,s}| + \left| \frac{\partial \phi_{k,s}}{\partial s} \right|^2 + \left| \frac{\partial^2 \phi_{k,s}}{\partial s^2} \right|^2 \quad (4)$$

Par conséquent, I_3 agit comme une contrainte de lissage. Le terme à droite de l'équation (4) peut aussi être pondéré par un vecteur de coefficients

$\gamma_k = (\gamma_{k,0}, \gamma_{k,1}, \gamma_{k,2})$ spécifique à chaque contour k . L'ensemble des coefficients γ_k , $k=1, \dots, K$, fait partie du modèle. Les fonctions I_1 et I_2 sont à leur tour pondérées respectivement par les coefficients α_k et β_k .

5. Implantation

L'approche a été implantée en utilisant une structure de données semblable à la forme discrète du modèle introduite à la section 4. Chaque contour dynamique comprend deux structures duales: (i) la chaîne codée du processus ligne, et (ii) le processus binaire de région. Les contours internes sont contraints à rester à l'intérieur de leur contour extrémal correspondant. L'interaction avec le modèle se fait à l'aide de l'étiquetage des segments de contours et le processus r qui code la superposition des régions des K contours.

Nous avons choisi une méthode d'optimisation descendante à échantillonnage aléatoire. Les méthodes descendantes sont moins coûteuses que les méthodes stochastiques. De plus, elles fonctionnent généralement bien quand on dispose d'un modèle *a priori* et d'une estimation initiale de la solution. Notre méthode consiste à choisir puis à déformer, d'une façon itérative et aléatoire, un contour parmi les K contours formant le modèle. Le segment de contour à déformer est aussi choisi aléatoirement. La distance maximale de déformation est bornée par une limite supérieure qui est réduite progressivement. Cette étape d'échantillonnage est suivie d'une étape de décision. Une déformation est conservée quand elle fait décroître l'énergie globale E .

6. Résultats expérimentaux

Nous avons effectué des expériences sur deux séries d'images synthétiques. Ces séries d'images ont été créées à partir de deux objets, un objet pseudo-cylindrique et un objet pseudo-conique. La première (deuxième) série est formée de trois

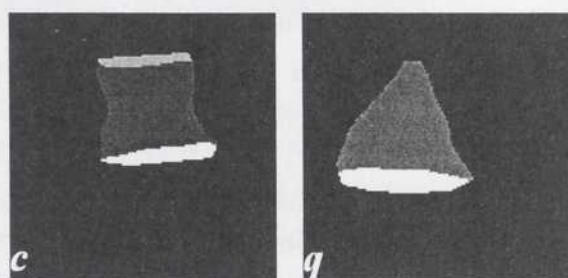


Figure 4. c est le modèle (solution initiale) utilisé pour la première série d'expérience, alors que q est utilisé pour la deuxième série.

images c_0 , c_1 et c_2 (q_0 , q_1 et q_2). Les images c_1 et c_2 (q_1 et q_2) ont été générées respectivement en ajoutant à c_0 (q_0) un bruit gaussien, indépendant et identiquement distribué (i.i.d.) d'une moyenne nulle et d'un écart type égal à 0.3 et 0.5. Le modèle et la solution initiale ont été les mêmes pour chacune des expériences. La figure 4 montre les deux modèles (solutions initiales) c et

q utilisés respectivement pour la première série et la deuxième série d'expériences. Le vecteur des paramètres $(\alpha_k, \beta_k, \gamma_{k,0}, \gamma_{k,1}, \gamma_{k,2})$ a été fixé pour toutes les expériences à $(1.0, 0.2, 0.01, 0, 0.01)$. Les paramètres (γ_k) ont été augmentés linéairement pour atteindre $(0.05, 0, 0.05)$. La figure 5 présente les deux séries d'images synthétiques et leurs résultats respectifs après 5 itérations. Chaque itération est équivalent à $K \times 10^{12}$ tentatives de déformations. De plus, nous avons voulu évaluer l'importance du terme I_2 dans le processus de recherche. Pour cela, nous avons répété les expériences avec les images $c2$ et $q2$ sous les mêmes conditions qu'auparavant, à l'exception du coefficient β_k de l'interaction solution-modèle, qui a été fixé à zéro. La figure 6 présente ces derniers résultats.



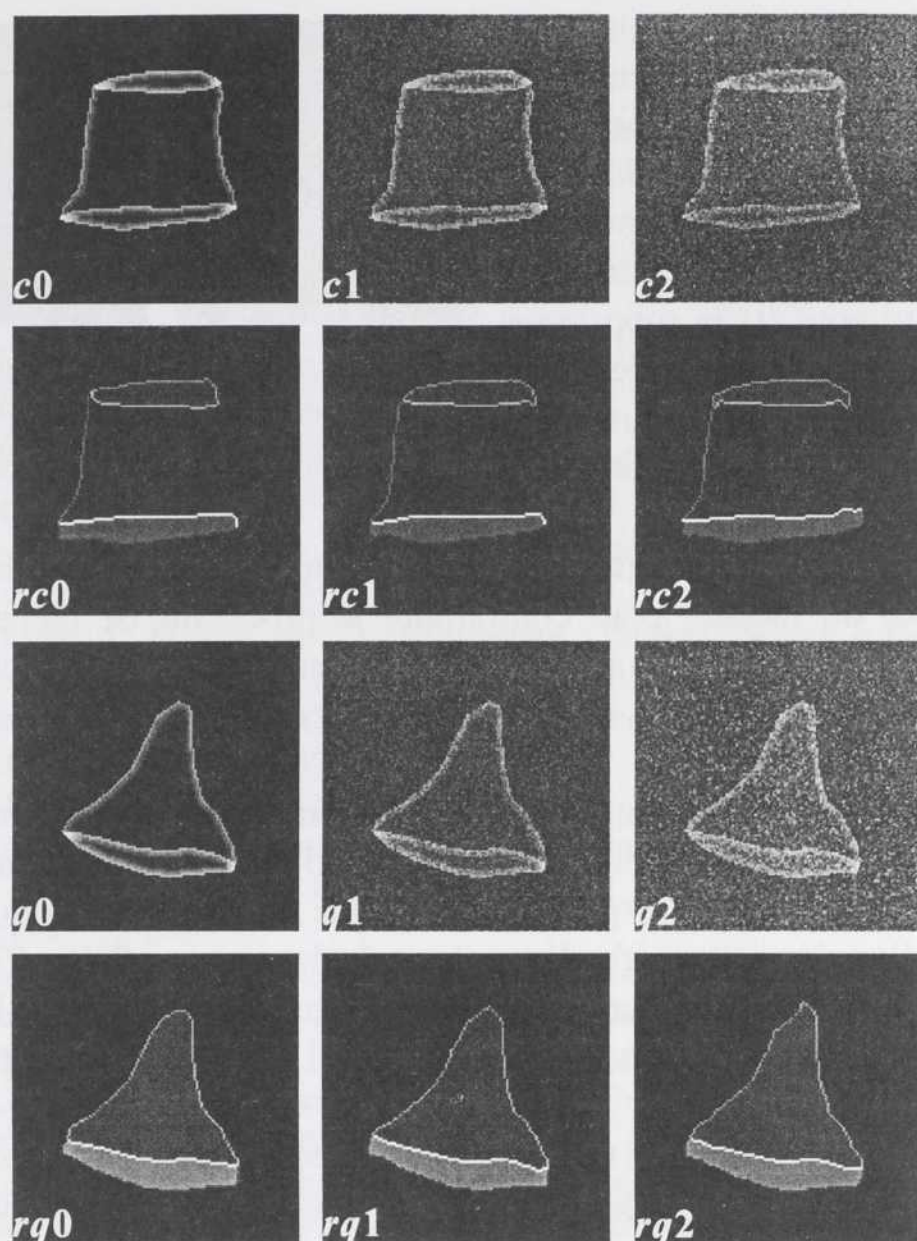


Figure 5. La première et la troisième lignes montrent les images synthétiques des deux séries. $c1$ et $c2$ ($q1$ et $q2$) sont générées de $c0$ ($q0$) en ajoutant respectivement un bruit gaussien i.i.d. $N(0,0.3)$ et $N(0,0.5)$. $rc0$, $rc1$ et $rc2$ ($rq0$, $rq1$ et $rq2$) sont les résultats correspondants à $c0$, $c1$ et $c2$ ($q0$, $q1$ et $q2$). Les étiquettes des régions et des lignes sont codées en niveau de gris.

7. Discussion

Les résultats des expériences d'appariement du modèle déformable avec les images synthétiques utilisant l'approche proposée sont encourageants. Nous soulignons que même avec des images à faible rapport S/N, nous avons obtenu une excellente convergence. Des expériences additionnelles utilisant différents

modèles et différentes solutions de départ ont aussi confirmé la robustesse et la bonne performance de la méthode d'optimisation utilisée. Les résultats ont confirmé deux importantes propriétés de notre approche qui sont (i) la robustesse par rapport au bruit, au modèle et à la solution de départ, et (ii) la précision de la détection des frontières d'intensité.

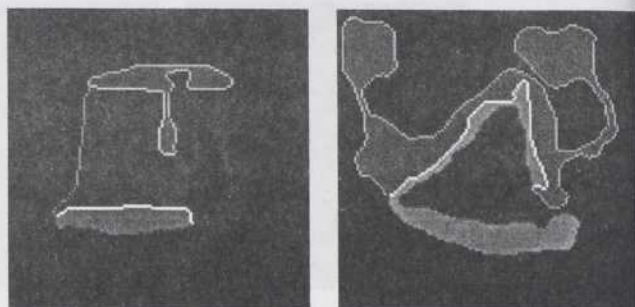


Figure 6. Les résultats des expériences répétées avec les images *c2* et *q2* de la figure 5. Les mêmes paramètres ont été utilisés à l'exception de β_k qui a été fixé à zéro.

Les résultats de la figure 6 montrent l'importance des connaissances relationnelles du modèle pour contraindre le processus de recherche. Ces connaissances relationnelles participent à préserver l'aspect structurel de la solution et par conséquent à restreindre l'espace de recherche des solutions possibles.

L'approche apporte plusieurs propriétés intéressantes. Le processus de recherche est formulé par l'optimisation globale d'une seule fonction d'objectif [10][15]. Cette dernière est exprimée comme une somme d'interactions locales permettant d'implanter la méthode sur des calculateurs parallèles. Finalement, la solution résultant de l'appariement est non biaisée par rapport à la solution de départ ou aux informations *a priori* [1][11][13], ce qui ne compromet pas la précision de la détection.

Nous prévoyons tester notre méthode avec des images plus complexes afin d'évaluer davantage sa capacité de fonctionner avec des images radiographiques réelles. Cependant, la robustesse de la méthode par rapport à l'estimation de la direction de projection θ peut être améliorée en utilisant un modèle stochastique. Ce modèle permettra la modification du graphe 2-D en échantillonnant dans la distribution des graphes possibles. Cela nous entraîne naturellement à choisir une méthode stochastique d'optimisation.

Remerciements

Ce travail a été supporté en partie par le Conseil de recherche en sciences naturelles et en génie du Canada (CRSNG), le Fonds pour la formation de chercheurs et l'aide à la recherche du Québec (FCAR), et une bourse du Conseil canadien des ingénieurs et de la Nord-Américaine vie.

8. Références

- [1] R. Bajscy and S. Kovacic, "Multiresolution elastic matching," *Computer Vision, Graphics, and Image Processing*, vol. 46, pp. 1-21, 1989.
- [2] R.A., Close, "Anatomic 3-D Reconstruction of Vascular Geometry from Two Orthogonal Angiographic Image Sequences," *SPIE*, vol. 1445, *Medical Imaging V*, pp. 513-522.
- [3] J.A. de Guise, H. Mallouche, J. Dansereau, et H. Labelle, "Techniques d'imagerie appliquées à la biomécanique rachidienne," *Rachis*, vol. 7, no. 3, pp. 135-144, 1995.
- [4] J.A. Feldman and Y. Yakimovsky, "Decision Theory and Artificial Intelligence: I. A Semantics-Based Region Analyser," *Artificial Intelligence*, vol. 5, pp. 349-371, 1974.
- [5] W.E.L. Grimson, *Object Recognition by Computer: The Role of Geometric Constraints*, Cambridge: MIT Press, 1990.
- [6] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour Models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321-331, 1988.
- [7] D.G. Lowe, *Perceptual Organization and Visual Recognition*, Boston: Kluwer Academic Publishers, 1985.
- [8] H. Mallouche, J.A. de Guise, and Y. Goussard, "Probabilistic Model of Multiple Dynamic Curve Matching for A Semitransparent Scene," in *Vision Geometry IV*, R.A. Metler, A.Y. Wu, F.L. Bookstein, W.D.K. Green, Editors, Proc. SPIE 2573, July 1995, pp.148-157.
- [9] J. Marroquin, S. Mitter, and T. Poggio, "Probabilistic solution of ill-posed problems in computational vision," *Journal of the American Statistical Association*, vol. 82, no. 397, pp. 76-89, 1987.
- [10] J.D. McCafferty, *Human and Machine Vision: Computing Perceptual Organisation*, New York: Ellis Horwood, 1990.
- [11] L.H. Staib and J.S. Duncan, "Boundary finding with parametrically deformable models," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 14, no. 11, pp. 1061-1075, 1992.
- [12] J.M. Tenenbaum and H.G. Barrow, "Experiments in Interpretation-Guided Segmentation," *Artificial Intelligence*, vol. 8, pp. 241-274, 1977.
- [13] D. Terzopoulos, J. Platt, A. Barr, and K. Fleischer, "Elastically Deformable Models," *Computer Graphics*, vol. 21, no. 4, pp. 205-214, 1987.
- [14] D. Terzopoulos, "Regularization of Inverse Visual Problems Involving discontinuities," *IEEE Trans. On Pattern Analysis and Machine Intelligence*, vol. 8, no. 2, pp. 129-139, 1986.
- [15] A. Witkin, J.M. Tenenbaum, "On the Role of Structure in Vision," in *Human and Machine Vision*, ed. J. Beck, B. Hope, and A. Rosenfeld, New York: Academic Press, pp. 481-543, 1983.

Recalage 3D automatique d'images SPECT du cerveau

J. Meunier^{a,c}, B. Imbert^a, C. Janicki^b, A. Guimond^a, J-P Soucy^b

^aD.I.R.O., Université de Montréal, C.P. 6128, Succ. Centre-ville, Montréal, H3C 3J7, CANADA. e-mail: meunier@iro.umontreal.ca

^bHôpital Notre-Dame, 1560 rue Sherbrooke Est, Montréal, H2L 4M1

^cLTSI (INSERM), Université de Rennes 1, 35042 Rennes Cedex, FRANCE*

Résumé

Nous présentons ici une méthode de recalage de volumes SPECT (Single Photon Emission Computed Tomography) basée sur le calcul du flux optique entre les images. La méthode ne nécessite aucune autre modalité d'imagerie et aucun marqueur externe pour effectuer l'alignement, cela afin d'en faciliter l'application clinique. Pour prendre en compte les difficultés typiques associées au calcul du flux optique (problème d'ouverture et d'estimation des dérivées), nous procédons en plusieurs étapes. Premièrement, le flux optique est contraint à un champ linéaire et est appliqué plusieurs fois afin de recalibrer approximativement les volumes SPECT. Lorsque le champ linéaire ainsi cumulé demeure stable, la contrainte de linéarité est relaxée et remplacée par une contrainte de lissage du champ non-linéaire pour obtenir un ajustement fin du recalage. Enfin, dans le but de démontrer le potentiel de la méthode, nous présentons des résultats obtenus sur un modèle d'ellipsoïdes et ainsi que sur des images SPECT obtenues en clinique.

1. Introduction

La comparaison d'images SPECT (ou autres modalités d'imagerie fonctionnelle) est nécessaire afin d'évaluer les données entre deux états chez un même patient (par ex. avant et après médication et/ou tâche cognitive) ou entre groupes de patients (normaux vs. anormaux) et nécessite donc un recalage précis des deux ensembles d'images. À cette fin, plusieurs techniques ont été proposées; malheureusement, elles souffrent toutes d'un ou plusieurs inconvénients: limitation à 2D [1], non entièrement automatique [2], radiographie latérale du crâne, IRM ou tomographie (CT) requise [3,4,5,6], utilisation d'un cadre stéréotaxique ou de marqueurs [6,7,8], identification de points de repères anatomiques [9,10]. Dans cet article, nous présentons une approche qui élimine ces problèmes afin d'obtenir une technique plus fiable et par-dessus tout plus pratique dans son implantation clinique.

* Jean Meunier est en visite au LTSI.

2. Méthode

Le recalage de deux volumes SPECT du cerveau est effectué en deux étapes principales. Premièrement, les volumes sont approximativement alignés en utilisant une transformation linéaire (affine) obtenue par calcul du flux optique 3D entre eux. Cependant, selon les travaux de Evans et coll. [11] en IRM (imagerie par résonance magnétique), l'utilisation d'un modèle linéaire va laisser une erreur RMS résiduelle de 6 à 7 mm; il est donc impératif d'accomplir une étape supplémentaire d'ajustement fin des deux cerveaux pour prendre en compte les différences morphologiques propres à chaque individu. Cette deuxième étape consistera donc à une correction non-linéaire obtenue encore une fois par calcul de flux optique.

2.1. Étape 1: Recalage linéaire

Une séquence 3D (de deux ou plusieurs images) peut être décrite mathématiquement par une fonction $I(x,y,z,t)$ où I représente l'intensité de l'image (ou le nombre de comptes en imagerie SPECT) au temps t et à la position (x,y,z) . En utilisant la règle de chaîne pour les dérivées, on obtient une première contrainte pour le calcul du flux optique:

$$\frac{dI}{dt} = \frac{\partial I}{\partial x} \cdot \frac{dx}{dt} + \frac{\partial I}{\partial y} \cdot \frac{dy}{dt} + \frac{\partial I}{\partial z} \cdot \frac{dz}{dt} + \frac{\partial I}{\partial t} \quad (1)$$

Les dérivées partielles dans (1) peuvent être estimées directement à partir des deux images SPECT à recaler [12]. Les quatre autres variables $U=dx/dt$, $V=dy/dt$, $W=dz/dt$ et dI/dt représentent respectivement la vitesse (déplacement) selon les axes x , y and z et les changements d'intensité de l'objet (cerveau). En pratique, on pose souvent $dI/dt = 0$, ce qui signifie que les cerveaux ont la même intensité et que seul un déplacement les différencie [12]. En moyenne, cette nouvelle contrainte est correcte si les cerveaux ont été préalablement normalisés par rapport à l'activité totale de chacun des cerveaux (on peut aussi envisager de normaliser en tenant compte du poids et de la taille du patient de même que de la dose injectée etc.). Ces deux contraintes (éq. (1) et $dI/dt = 0$) ne suffisent pas pour obtenir les déplacements U, V, W nécessaires au recalage (en fait, seul la composante parallèle au gradient de l'intensité de l'image peut être obtenue [12]). On impose donc la linéarité du champ (U, V, W) afin d'obtenir une première solution:

$$\begin{bmatrix} U \\ V \\ W \end{bmatrix} = \begin{bmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \\ \frac{dz}{dt} \end{bmatrix} = \bar{T} + M \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad (2)$$

où $\bar{T}$ et M sont, respectivement, un vecteur translation et une matrice 3×3 . En écrivant pour chaque voxel l'équation (1) et en substituant pour U, V et W la contrainte de linéarité (eq. 2) et $dI/dt=0$, on obtient un système surdéterminé qui peut être solutionné par moindres carrés. En pratique, on doit recourir à un filtrage passe-bas des volumes SPECT afin de réduire le bruit et obtenir une estimation plus fiable des dérivées partielles (spatiales et temporelles); cela a pour effet d'améliorer la vitesse et la stabilité du recalage. Ainsi, on obtient la transformation linéaire nécessaire à l'alignement des deux volumes cérébraux. Cependant, ce premier estimé est souvent imparfait si la transformation à effectuer entre les deux images est importante; en effet, dans pareil cas, le calcul numérique des dérivées partielles est très approximatif. Barber et coll. [13] ont montré qu'on pouvait appliquer plusieurs fois le flux optique linéaire afin d'améliorer les résultats en adoptant le schéma suivant: on recalcule les deux volumes à partir d'une première estimation du champ linéaire, puis on calcule à nouveau une estimation par flux optique du champ linéaire mais cette fois *entre les volumes préalablement recalés* (en utilisant une interpolation trilineaire), en combinant les deux champs linéaires on obtient un meilleur estimé du champ linéaire total nécessaire au recalage. On peut répéter jusqu'à ce que le champ linéaire demeure stable. Bien que Barber et coll. [13] utilisent cette approche pour recalculer les cerveaux d'un *même* patient, nous l'utilisons ici avec succès pour obtenir une transformation linéaire entre des cerveaux différents. Cela nous donne une première estimation du recalage. Généralement quelques itérations (2 ou 3) suffisent pour obtenir la meilleure solution linéaire.

2.2. Étape 2: Recalage non-linéaire

La transformation linéaire précédente est nécessaire pour avoir un premier estimé du recalage et sera utilisée comme solution initiale pour calculer une transformation résiduelle non-linéaire pour parfaire le recalage. Sans cette étape, la transformation non-linéaire qui suit risque d'être erronée à cause des estimations médiocres des dérivées partielles lorsque les cerveaux ne sont pas déjà quasi-recalés. L'algorithme non-linéaire s'inspire fortement du

(2) calcul de flux optique présenté par Horn et Schunk [12], mais diffère dans son application et son interprétation (voir fig. 1).

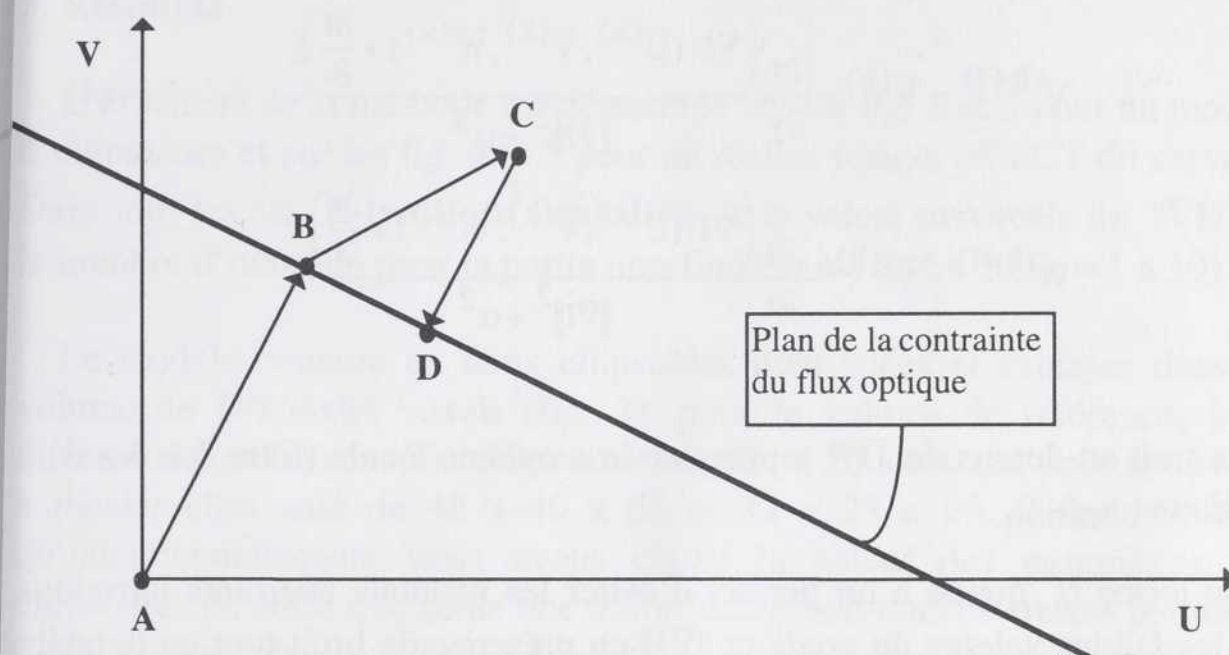


Fig. 1: Représentation 2D de la procédure itérative de calcul du champ non-linéaire par une approche de flux optique.

Initialement les déplacements en x , y et z , c.-à-d. $U=dx/dt$, $V=dy/dt$ et $W=dz/dt$, sont posés nuls (point A sur la représentation 2D de la fig. 1); cette hypothèse serait juste si une transformation linéaire suffisait à faire un recalage parfait. Puis on impose la solution qui satisfait au mieux (distance minimum) l'équation (1) du flux optique pour chaque voxel (point B sur la fig. 1). Cette solution n'est généralement pas satisfaisante puisqu'elle correspond uniquement à la composante du déplacement parallèle au gradient d'intensité alors que pour un voxel donné tous les points appartenant au plan de la contrainte de flux optique (équ. 1) sont valides. On impose donc une condition qui impose itérativement un lissage du champ pour privilégier ce type de solution. Pour cela on calcule localement pour chaque voxel la moyenne du champ (point C sur la fig. 1) dans un voisinage $3 \times 3 \times 3$ que l'on reprojète sur le plan (point D sur la fig. 1). Cette solution respecte donc l'équation (1) puisqu'elle est sur le plan et donne aussi un champ plus lisse. En répétant plusieurs fois cette procédure on obtient un champ encore plus lisse. L'algorithme peut s'écrire comme suit pour chaque voxel:

$$\begin{aligned}
 U^{(k+1)} &= \bar{U}^{(k)} - \frac{\partial I \left(\nabla I \cdot (\bar{U}^{(k)}, \bar{V}^{(k)}, \bar{W}^{(k)}) + \frac{\partial I}{\partial t} \right)}{\partial x \left(\|\nabla I\|^2 + \alpha^2 \right)} \\
 V^{(k+1)} &= \bar{V}^{(k)} - \frac{\partial I \left(\nabla I \cdot (\bar{U}^{(k)}, \bar{V}^{(k)}, \bar{W}^{(k)}) + \frac{\partial I}{\partial t} \right)}{\partial y \left(\|\nabla I\|^2 + \alpha^2 \right)} \\
 W^{(k+1)} &= \bar{W}^{(k)} - \frac{\partial I \left(\nabla I \cdot (\bar{U}^{(k)}, \bar{V}^{(k)}, \bar{W}^{(k)}) + \frac{\partial I}{\partial t} \right)}{\partial z \left(\|\nabla I\|^2 + \alpha^2 \right)}
 \end{aligned} \tag{4}$$

où le trait au-dessus de $\bar{U}^{(k)}$ représente la moyenne locale (filtre 3 x 3 x 3) à la k ème itération.

Le terme α^2 quand à lui permet d'éviter les résultats aberrants introduits par les faibles valeurs du gradient $\|\nabla I\|$ en présence de bruit tout en donnant préséance aux valeurs élevées de $\|\nabla I\|$ dans le processus itératif; cela signifie que les contours (contraste élevé) vont mener la déformation non-linéaire, alors que les régions de faible contraste vont subir le mouvement moyen de leur voisinage. Cela est important afin que le bruit dans les images et les petits changements d'intensités inter-patients n'affectent pas (incorrectement) le calcul du recalage. Sur la fig. 1, la procédure itérative est illustrée pour $\alpha^2 = 0$, en réalité, les points B et D s'approchent du plan (droite sur la fig. 1) sans l'atteindre lorsque $\alpha^2 \neq 0$.

La méthode offre une caractéristique intéressante qui consiste à pouvoir ajouter des contraintes associées, par exemple, à des repères anatomiques préalablement identifiés dans les deux volumes à recalcr. Ainsi, si nous savons que le déplacement du voxel au centre du thalamus est (a,b,c), nous avons simplement à imposer ce mouvement à ce voxel à chaque itération: $U^{(k)} = a$, $V^{(k)} = b$, $W^{(k)} = c$. Nous forcerons ainsi les voxels voisins du centre du thalamus à adopter un mouvement similaire tout en gardant le champ non-linéaire lisse.

Comme pour l'étape linéaire précédente, cette étape non-linéaire peut être appliquée plusieurs fois si nécessaire en combinant les résultats intermédiaires obtenus en recalant les volumes à chaque fois. Nous obtenons donc, suite à cette étape, un champ non-linéaire lisse qui ajouté au champ

linéaire calculé à la première étape donne un recalage précis des deux volumes SPECT.

3. Résultats

L'efficacité de la méthode est démontrée sur les fig. 2 et 3 pour un modèle d'ellipsoïdes et sur les fig. 4 et 5 pour de réelles images SPECT du cerveau. Dans tous les cas, le terme α^2 vaut 20% de la valeur maximale de $\|\nabla I\|^2$, et le nombre d'itération pour la partie non-linéaire est fixé à 10 ($k = 1$ à 10).

Le modèle consiste en deux ellipsoïdes imbriquées et centrées dans un volume de $64 \times 64 \times 64$ voxels (fig. 2); pour le volume de référence, leurs dimensions (axes) sont $56 \times 48 \times 40$ et $28 \times 24 \times 20$ alors que pour le volume à recaler elles sont de $48 \times 40 \times 56$ et $32 \times 28 \times 20$. Il faut noter ici qu'intentionnellement, nous avons choisi la valeur des paramètres des ellipsoïdes de sorte que seule une déformation non-linéaire puisse permettre de faire un recalage adéquat. La première étape ne peut donc pas suffire à la tâche.

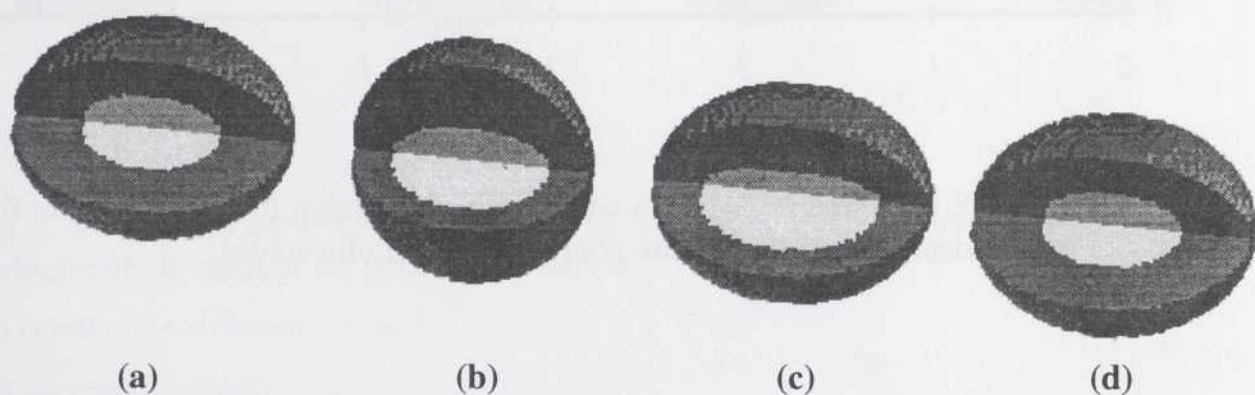


Fig. 2: Modèle d'ellipsoïdes avant et après recalage. Une coupe a été réalisée afin de montrer l'ellipsoïde interne.

(a) volume de référence (la grande et la petite ellipsoïde ont respectivement comme dimensions: $56 \times 48 \times 40$ et $28 \times 24 \times 20$).

(b) volume à recaler (la grande et la petite ellipsoïde ont respectivement comme dimensions: $48 \times 40 \times 56$ et $32 \times 28 \times 20$).

(c) volume en (b) après recalage linéaire (3 applications du flux optique) vers le volume de référence en (a).

(d) volume (b) après recalage linéaire et non-linéaire (3 applications du flux optique) vers le volume de référence en (a).

L'erreur RMS normalisée par rapport à la moyenne des voxels dans l'objet de référence a été calculée à chaque étape du recalage (fig. 3). Les volumes sont présentés sur la fig. 2. Initialement l'erreur est 0.415, (fig. 2a vs. b) et décroît à 0.222 après l'étape linéaire du recalage (fig. 2c). Cette étape a été répétée successivement 3 fois pour obtenir un champ linéaire stable. Tel que prévu, le champ linéaire ne suffit pas afin de recalculer correctement les deux volumes (comparez fig.2(a) et fig.2(c) en particulier le petit ellipsoïde). Suite à l'étape non-linéaire (fig. 2d), l'erreur décroît jusqu'à 0.053 après 3 applications successives du flux optique. L'amélioration du recalage est évidente sur la fig. 2 (a vs. d).

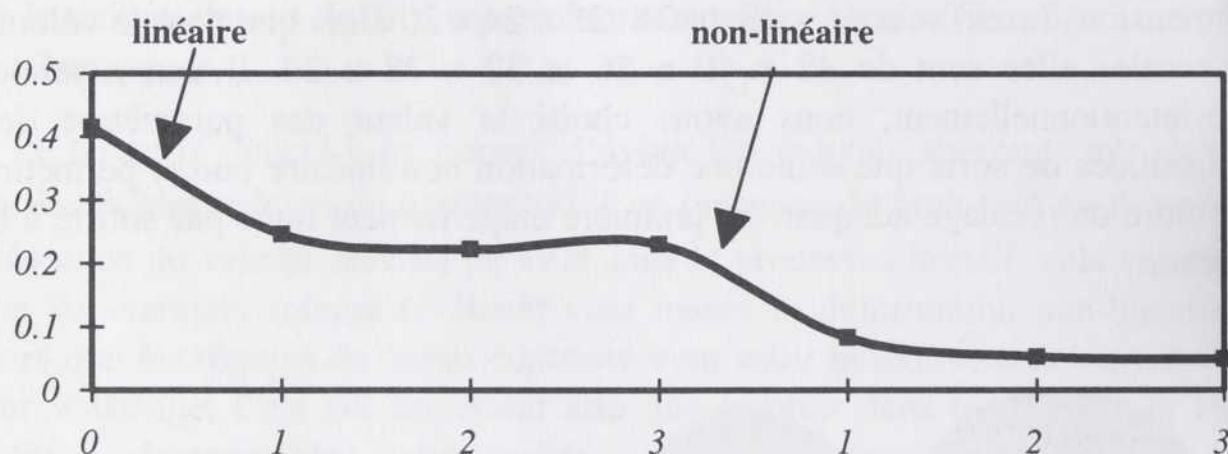


Fig. 3: Erreur RMS pondérée (voir texte) en fonction du nombre d'applications du flux optique à l'étape linéaire puis non-linéaire pour le modèle d'ellipsoïdes.

Les données SPECT ont aussi comme dimensions 64 x 64 x 64 voxels; l'erreur RMS normalisée au nombre de comptes moyen dans le cerveau de référence (fig. 5a) est initialement de 0.396 avec le cerveau à recalculer (fig. 5b). Après recalage linéaire (2 applications successives de l'algorithme de flux optique), nous réduisons cette erreur à 0.215 (fig. 5c) puis 0.129 suite au recalage non-linéaire (3 applications). Encore une fois, nous pouvons noter visuellement la qualité du recalage (fig. 5d).

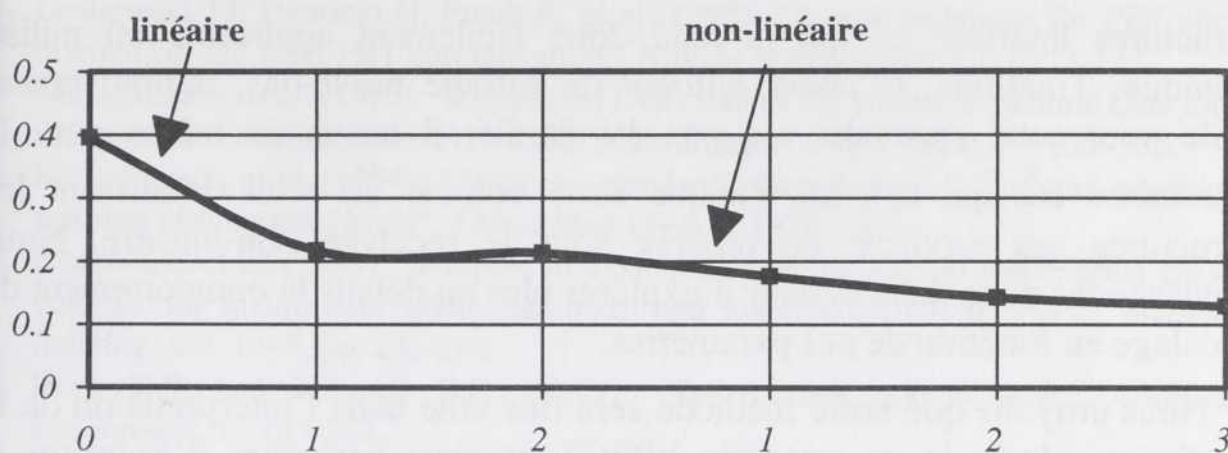


Fig. 4: Erreur RMS pondérée (voir texte) en fonction du nombre d'itérations à l'étape linéaire puis non-linéaire pour des données SPECT réelles.

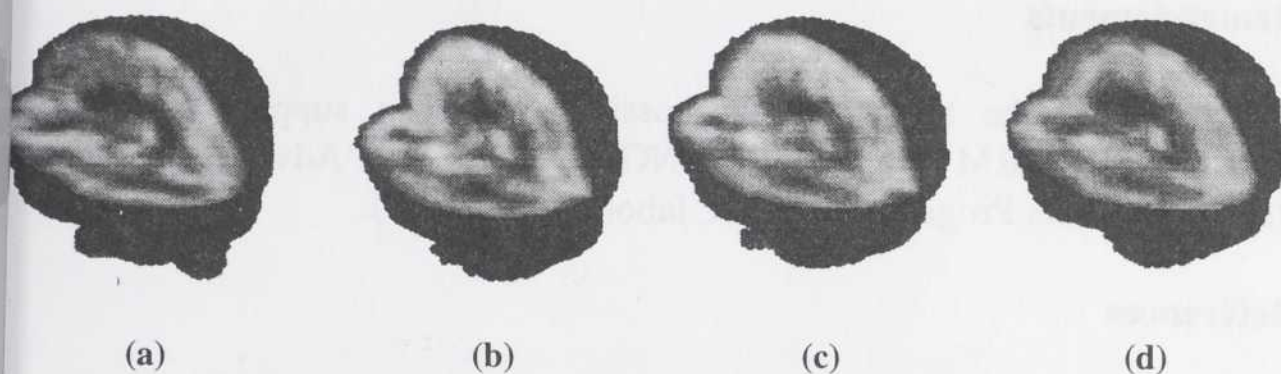


Fig. 5: Vue 3D des cerveaux obtenus en SPECT avant et après recalage. Une coupe a été réalisée afin de montrer les structures internes.

(a) cerveau de référence.

(b) cerveau à recaler.

(c) cerveau en (b) après recalage linéaire (2 applications successives du flux optique).

(d) cerveau en (b) après recalage linéaire et non-linéaire (3 applications du flux optique).

4. CONCLUSIONS

La méthode de recalage proposée élimine la plupart des inconvénients rencontrés dans les autres approches rapportées dans la littérature. En plus d'être automatique, elle est indépendante de toutes autres modalités d'imagerie, de l'utilisation de marqueurs externes ou du repérage de

structures internes; ce qui la rend donc facilement applicable en milieu clinique. Toutefois, le choix *optimal* du filtrage passe-bas, définitivement utile pour cette approche n'a pas été étudié; il en va de même pour le paramètre α^2 , qui agit en quelque sorte comme un seuil définissant les structures qui serviront de repères pour le recalage non-linéaire. Nous envisageons donc dans le futur d'explorer plus en détails le comportement du recalage en fonction de ces paramètres.

Nous croyons que notre méthode sera très utile dans l'interprétation de la perfusion cérébrale en imagerie SPECT et ainsi permettra d'accroître la fiabilité des examens cliniques. Enfin, l'application de notre technique peut être aussi considérée avec d'autres modalités d'imagerie (par ex. IRM ou PET).

Remerciements

Cette recherche a été rendue possible grâce au support du FRSQ (Québec), INSERM (France), CRSNG (Canada) et l'Advanced Clinical Research Grants Program de ADAC laboratories (USA).

Références

1. Barber D.C. (1992): "Registration of low resolution medical images", *Phys. Med. Biol.*, pp. 1485-1498, 1992.
2. Pietrzyk U, Herholz K, Fink G, et al (1994): "An interactive technique for three-dimensional image registration: validation for PET, SPECT, MRI and CT brain studies". *Journal of Nuclear Medicine* 35, p 2011-2018.
3. Fox PT, Perlmuter JS, Raichle ME (1985): "A stereotactic method of anatomical localization for positron emission tomography". *Journal of computer assisted tomography* 9, p 141-153.
4. Pelizzari CA, Chen GTY, Spelbring DR, Weichselbaum RR, Chen CT (1989): "Accurate three-dimensional registration of CT, PET, and/or MR images of the brain". *Journal of computer assisted tomography* 13, p 20-26.
5. Evans AC, Marret S, Torrescorzo J, Ku S, Collins L (1991): "MRI-PET correlative analysis using a volume of interest (VOI) atlas". *Journal of cerebral blood flow and metabolism* 11, p A-69-A78.
6. West, J. M. Fitzpatrick et al. (1996) "Comparison and evaluation of retrospective intermodality registration techniques, *SPIE Medical Imaging*, vol. 2710.
7. Mountz JM, Modell JG, Deutsch G, Harris JM (1992): "Within-subject test:retest variability using Tc-99m HM-PAO brain SPECT". *Journal of Nuclear Medicine* 33, p 826 (Abstract).

3. Grabowski TJ, Damasio H, Frank R, et al (1995): "A new technique for PET slice orientation and MRI-PET coregistration". *Human brain mapping* 2, p123-133.
9. Minoshima S, et al. (1993): "Automated Detection of the Intercommissural Line for Stereotactic Localization of Functional Brain Images". *J Nucl Med*, 34: 322-329.
10. Minoshima S, et al. (1994): "Anatomic standardization: Linear scaling and nonlinear warping of functional brain". *J Nucl Med* 1994;35:1528-1537.
11. Evans A.C. et al. (1991): "Warping of a computerized 3D atlas to match brain image volumes for quantitative neuroanatomical and functional analysis," *SPIE Medical Imaging*, vol. 1445, pp. 236-246.
12. Horn K.P. and Schunck B.G. (1981): "Determining optical flow," *Artificial intelligence*, 17:185-203.
13. Barber et al. (1995): "Automatic registration of SPECT images as an alternative to immobilization of neuroactivation studies," *Phys. Med. Biol.*, pp. 449-463.

Reconstruction du mouvement d'une ellipsoïde à partir de ses projections

Rita Noumeir

*Ecole de technologie Supérieure, département de génie électrique, 4750, avenue Henri-Julien, Montréal (Québec), H2T 2C8, CANADA.
e-mail: rnoumeir@ele.etsmtl.ca*

Résumé

Nous présentons une méthode pour détecter et quantifier le mouvement d'une ellipsoïde tridimensionnelle à partir de ses projections orthographiques. L'ellipsoïde est représentée par une matrice symétrique. La relation entre l'ellipsoïde et sa projection orthographique est alors linéaire. Cette représentation est utilisée pour reconstruire l'ellipsoïde à partir d'une série de projections orthographiques. L'orientation et la longueur des axes des ellipses de projection sont utilisées pour détecter la rotation tridimensionnelle de l'ellipsoïde entre deux projections. Cette rotation est quantifiée en déterminant l'angle de rotation et la direction de l'axe de rotation. Cette méthode peut être utilisée en tomographie cardiaque pour détecter et quantifier, à partir des données de projection, le mouvement du patient durant une acquisition tomographique.

1. Introduction

La reconstruction tomographique consiste à déterminer le volume du patient à partir d'un nombre limité d'images de projection acquises par une caméra qui tourne autour du patient. Cependant, le mouvement du patient durant l'acquisition tomographique introduit des artefacts importants et détériore d'une façon qualitative et quantitative le volume reconstruit. En médecine nucléaire, la durée de l'acquisition tomographique est souvent supérieure à quinze minutes et le mouvement du patient durant l'acquisition est assez fréquent. Ce mouvement diminue la résolution et la sensibilité de la tomographie [1]. De plus, il introduit des artefacts dans le volume reconstruit comme des rayures, des séparations de frontières et des distorsions de forme [2]. En tomographie cardiaque, l'artefact de mouvement ressemble à une ischémie. La détection du mouvement durant l'acquisition tomographique permet de contrôler la qualité de la reconstruction. Pour cette fin, le médecin examine visuellement les images de projections qui sont affichées sur un écran, l'une à la suite de l'autre (en mode d'affichage ciné). Cette méthode de détection est subjective et inadéquate. En effet, seul un mouvement de translation parallèle à l'axe de rotation de la caméra peut être détecté. Par

contre, aucune information n'est obtenue concernant sa grandeur ou l'angle de la caméra où il a lieu. Une méthode quantitative est nécessaire pour déterminer la grandeur du mouvement, sa direction et l'angle de la projection où il a lieu. Nous avons déjà proposé l'utilisation de la méthode du flux optique pour détecter et quantifier le mouvement durant la tomographie [3,4]. La méthode du flux optique, utilisée dans une région d'intérêt, permet de calculer le déplacement horizontal et vertical entre deux images de projections successives. Nous avons montré que cette méthode réussit à détecter et quantifier le mouvement entre les images de projections. Cependant, tout comme les autres méthodes quantitatives [5], elle souffre de deux inconvénients:

1- Considérons les images de projection d'un point qui n'est pas situé sur l'axe de rotation. Son déplacement horizontal entre les images de projection est sinusoidal tandis que son déplacement vertical est nul. Si ce point subit une translation, alors son déplacement entre les images de projection subit une discontinuité. Cette discontinuité est facilement identifiable car le déplacement de base est connu. Par contre, le déplacement horizontal de base d'un objet tridimensionnel est inconnu. Ceci rend difficile la quantification de ce déplacement et limite l'utilisation de ces méthodes à la quantification du mouvement parallèle à l'axe de rotation uniquement.

2- Les images acquises en tomographie sont des images de projection. Le mouvement bidimensionnel analysé à partir de ces images est une projection du mouvement tridimensionnel du patient. La détection et la quantification du mouvement bidimensionnel entre deux images de projection successives ne permet pas de corriger un mouvement tridimensionnel du patient.

Pour remédier à ces problèmes, nous proposons de modéliser le myocarde à l'aide d'une ellipsoïde tridimensionnelle et d'étudier le mouvement de cette ellipsoïde à partir de ces projections. La modélisation du myocarde à l'aide d'une ellipsoïde est fréquente en imagerie médicale (pour mesurer la fraction d'éjection par exemple). La projection orthographique d'une ellipsoïde est une ellipse bidimensionnelle. Une translation de l'ellipsoïde se traduit par une translation du centre de l'ellipse dans les images de projection. Cependant une rotation de l'ellipsoïde se traduit par des changements de l'orientation et des longueurs des axes des ellipses dans les images de projection. Nous proposons d'étudier la relation entre une rotation tridimensionnelle d'une ellipsoïde et

l'orientation et les longueurs des axes de ses projections orthographiques. Vu que nous considérons le cas d'une projection orthographique, les informations sur la forme de l'ellipse (orientation et longueurs des axes) sont utilisées plutôt que son intensité. Les ellipses dans les images de projection sont supposées connues et peuvent être obtenues, en pratique, à l'aide de différents algorithmes de détection d'ellipse dans une image [6].

2. Méthode

Une ellipsoïde est souvent représentée à l'aide des demi-longueurs de ses axes et de leur orientation par rapport à un système d'axes donné. A l'aide de cette représentation, les propriétés géométriques de l'ellipsoïde sont directement identifiables. Cependant, cette représentation génère une relation non linéaire entre l'ellipsoïde et sa projection orthographique. Par contre, la représentation d'une ellipsoïde à l'aide d'une matrice symétrique permet d'avoir une relation linéaire entre l'ellipsoïde et sa projection orthographique. En effet, toute ellipsoïde ζ de degré n , centrée à l'origine, peut être représentée par une matrice symétrique positive semi-définie X de dimensions $n \times n$ [7,8]. L'équation de l'ellipsoïde est alors:

$$u^t X u = 1 \quad (1)$$

où u est un vecteur de position de dimension n et l'exposant t indique la transposée. Les propriétés géométriques de l'ellipsoïde sont cachées dans les propriétés algébriques de la matrice symétrique correspondante. Les directions des axes de l'ellipsoïde sont celles des vecteurs propres v_i de la matrice X et les demi-longueurs des axes sont données par $\sqrt{\lambda_i}$, où λ_i sont les valeurs propres de la matrice X . De plus, toute transformation de coordonnées de la forme $\tilde{u} = Lu$ se traduit par un changement de la matrice X . En effet, l'ellipsoïde devient représentée par:

$$X_L = L X L^t \quad (2)$$

2.1 Projection orthographique

Nous considérons le cas d'une ellipsoïde tridimensionnelle. Une projection

orthographique est décrite par la transformation de coordonnées $\tilde{u} = C^t u$, où la matrice C est donnée par:

$$C = \begin{bmatrix} \cos \varphi & 0 \\ \sin \varphi & 0 \\ 0 & 1 \end{bmatrix} \quad (3)$$

L'angle de projection φ est l'angle de rotation autour de l'axe de rotation $(0,0,1)$. L'origine des angles, $\varphi = 0$, est le plan de projection dont la normale est $(0,1,0)$. La projection orthographique de l'ellipsoïde tridimensionnelle X est une ellipse bidimensionnelle Y . D'après les équations 2 et 3, Y est donnée par:

$$Y = C^t X C \quad (4)$$

2.2 Reconstruction tomographique

La reconstruction tomographique consiste à trouver l'ellipsoïde X à partir d'un nombre q de projections orthographiques Y_i de la forme:

$$Y_i = C_i^t X C_i \quad (5)$$

où

$$C_i = \begin{bmatrix} \cos \varphi_i & 0 \\ \sin \varphi_i & 0 \\ 0 & 1 \end{bmatrix} \quad (6)$$

pour $1 \leq i \leq q$; En tomographie cardiaque nucléaire, 32 projections sont généralement acquises; les angles de projection sont uniformément espacés sur 180 degrés, tel que: $-45^\circ \leq \varphi_i < 180^\circ - 45^\circ$. L'équation (5) peut être écrite sous la forme:

$$y_i = \tilde{C}_i x \quad (7)$$

où y_i et x sont des vecteurs obtenus en arrangeant les colonnes des matrices Y_i et X respectivement, l'une à la suite de l'autre. $\tilde{C}_i$ est le produit de Kronecker de la matrice C_i par elle-même, obtenu en calculant le produit de tous les éléments de C_i par la matrice C_i . En arrangeant tous les vecteurs observés y_i en un même vecteur y , l'acquisition tomographique s'écrit $y = Cx$ ou:

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_q \end{bmatrix} = \begin{bmatrix} \tilde{C}_1 \\ \tilde{C}_2 \\ \dots \\ \tilde{C}_q \end{bmatrix} x \quad (8)$$

La solution x (au sens des moindres carrées) est donnée par $x = y \backslash C$. Cependant, les éléments de x ne sont pas tous indépendants car la matrice X est symétrique. Cette propriété est utilisée pour supprimer la redondance et réduire la dimension de x . La matrice C est aussi réduite en éliminant les colonnes 2, 3 et 6 et en multipliant la colonne 4 est par 2. La solution x existe si le rang de C est 6 ou, en d'autres termes, si C contient au moins 6 lignes indépendantes. Ceci est obtenu si au minimum 3 projections sont utilisées.

2.3 Rotation tridimensionnelle

Une rotation tridimensionnelle de l'ellipsoïde est une transformation des coordonnées $\tilde{u} = Tu$, où T est donné par:

$$T = \begin{bmatrix} n_x^2 v \theta + c \theta & n_x n_y v \theta - n_z s \theta & n_x n_z v \theta + n_y s \theta \\ n_x n_y v \theta + n_z s \theta & n_y^2 v \theta + c \theta & n_y n_z v \theta - n_x s \theta \\ n_x n_z v \theta - n_y s \theta & n_y n_z v \theta + n_x s \theta & n_z^2 v \theta + c \theta \end{bmatrix} \quad (9)$$

n est un vecteur unitaire dont les composantes sont (n_x, n_y, n_z) et dont la direction est celle de l'axe de rotation; $c\theta = \cos\theta$, $s\theta = \sin\theta$ et $v\theta = 1 - \cos\theta$, θ étant l'angle de rotation autour de n [9]. Selon l'équation (2), l'ellipsoïde X devient:

$$X_T = TXT^t \quad (10)$$

Si l'ellipsoïde X subit une rotation juste avant la projection d'indice p , ses projections sont données par l'équation (5) pour $1 \leq i < p$ et par

$$Y_i = C_i^t (TXT^t) C_i \quad (11)$$

pour $p \leq i \leq q$.

2.4 Détection de mouvement

Si la matrice Y_i d'une ellipse de projection, est de la forme:

$$Y_i = \begin{bmatrix} a & h \\ h & b \end{bmatrix} \quad (12)$$

Les demi-longueurs des axes sont données par:

$$a', b' = \left(\frac{a + b \pm \sqrt{(a - b)^2 + 4h^2}}{2} \right)^{1/2} \quad (13)$$

et l'orientation de l'ellipse par rapport à l'horizontale est donné par l'angle [10]:

$$\alpha = \frac{1}{2} \text{atan} \frac{2h}{b - a} \quad (14)$$

Les variations de a' , b' et α sont tracées en fonction de l'indice de projection i . Toute discontinuité indique la présence de mouvement; l'endroit de cette discontinuité indique l'indice p de la projection où le mouvement a lieu.

2.5 Quantification du mouvement

Si $p > 3$ ou $p \leq q - 3$, deux ellipsoïdes peuvent être reconstruites: X_1 avec les données Y_i pour $1 \leq i < p$ et X_2 avec les données Y_i pour $p \leq i \leq q$. Soient V_1 et V_2 deux matrices orthogonales dont les colonnes sont les vecteurs propres de X_1 et X_2 respectivement. Les vecteurs propres de X_1 ou de X_2 sont trois vecteurs unitaires orthonormés qui définissent un système d'axes cartésiens. Les deux systèmes d'axes définis par les matrices V_1 et V_2 sont liés par la relation $V_2 = TV_1$. La matrice de rotation T peut être obtenue par $T = V_2 V_1^{-1}$, où l'exposant -1 représente l'inverse d'une matrice. L'angle de rotation est donné par [9]:

$$\theta = \arccos \frac{t_{11} + t_{22} + t_{33} - 1}{2} \quad (15)$$

et la direction de l'axe de rotation est donnée par:

$$n = \frac{1}{2 \sin \theta} \begin{bmatrix} t_{32} - t_{23} \\ t_{13} - t_{31} \\ t_{21} - t_{12} \end{bmatrix} \quad (16)$$

3. Résultats

Toute ellipsoïde ζ peut être générée en deux étapes: 1-une matrice diagonale est d'abord générée; les éléments de la diagonale sont le carré de la demi-longueur des axes de l'ellipsoïde. 2- La matrice diagonale est tournée selon l'équation (10) pour générer une ellipsoïde dont les axes sont tournés par rapport au système d'axes fixe de référence. Nous illustrons la méthode en utilisant une ellipsoïde dont les demi-longueurs des axes sont 10, 14 et 19 pixels. Ses axes sont tournés autour de la direction (1,1,1), d'un angle de 40 degrés. 64 projections de cette ellipsoïde sont mesurées sur 360 degrés; l'angle de départ est -45 degrés. Juste avant la projection numéro 17, l'ellipsoïde subit une rotation de 30 degrés autour de la direction (1,0,0); c'est ce mouvement qu'on aimerait détecter et quantifier. Un bruit gaussien d'écart-type 2 a été ajouté aux données de projection y . Les données de projection consistent en 64 ellipses. L'orientation de ces ellipses par rapport à l'horizontale, l'angle α

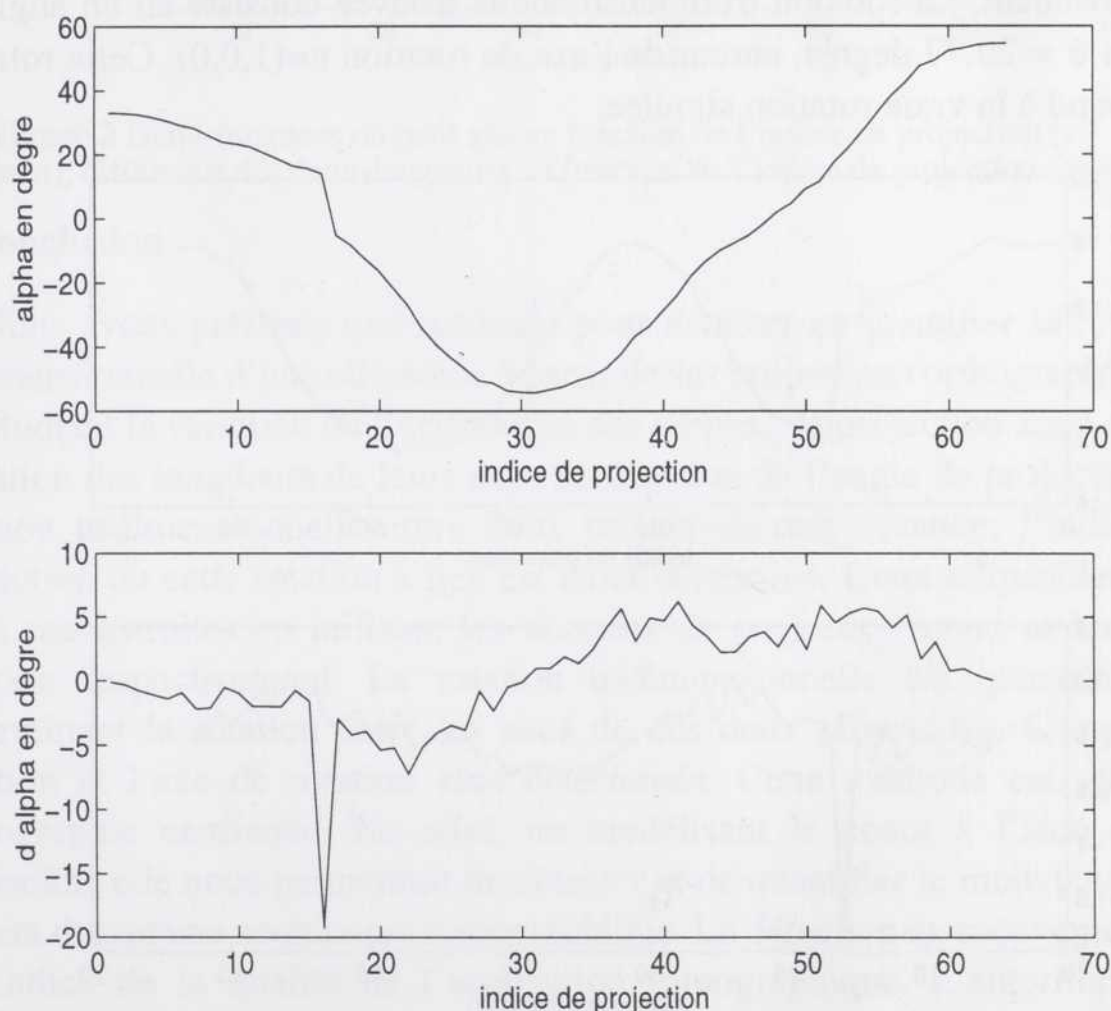


Figure 1 Orientation des ellipses en fonction de l'indice de projection (en haut)
Différence des orientations en fonction de l'indice de projection (en bas).

selon l'équation (14), est affichée à la figure 1 (en haut) en fonction de l'indice de projection. Le mouvement de l'ellipsoïde provoque une discontinuité de l'orientation de l'ellipse à la projection numéro 17. Pour mieux illustrer cette discontinuité, la différence de l'orientation entre des projections successives est montrée à la figure 1, en bas. Les demi-longueurs du grand axe et du petit axe des 64 ellipses de projection sont affichées aux figures 2 et 3 respectivement, en fonction de l'indice de projection. Le mouvement de l'ellipsoïde provoque aussi une discontinuité dans la longueur du grand axe, visible à la projection numéro 17 (figure 2, en haut). Cette discontinuité est mise en évidence à la figure 2 en bas, où la variation de la demi-longueur du grand axe, entre deux projections successives, est affichée. D'après la figure 3 (en bas), la variation de la demi-longueur du petit axe, entre deux projections successives, est plus petite qu'un pixel et la discontinuité causée par le mouvement est noyée dans le bruit. Pour quantifier le mouvement après l'avoir détecté, deux ellipsoïdes X_1 et X_2 sont reconstruites en utilisant les données Y_i pour $1 \leq i < 17$ et pour $17 \leq i \leq 64$ respectivement. La rotation tridimensionnelle trouvée consiste en un angle de rotation $\theta = 29.77$ degrés, autour de l'axe de rotation $n=(1,0,0)$. Cette rotation correspond à la vraie rotation simulée.

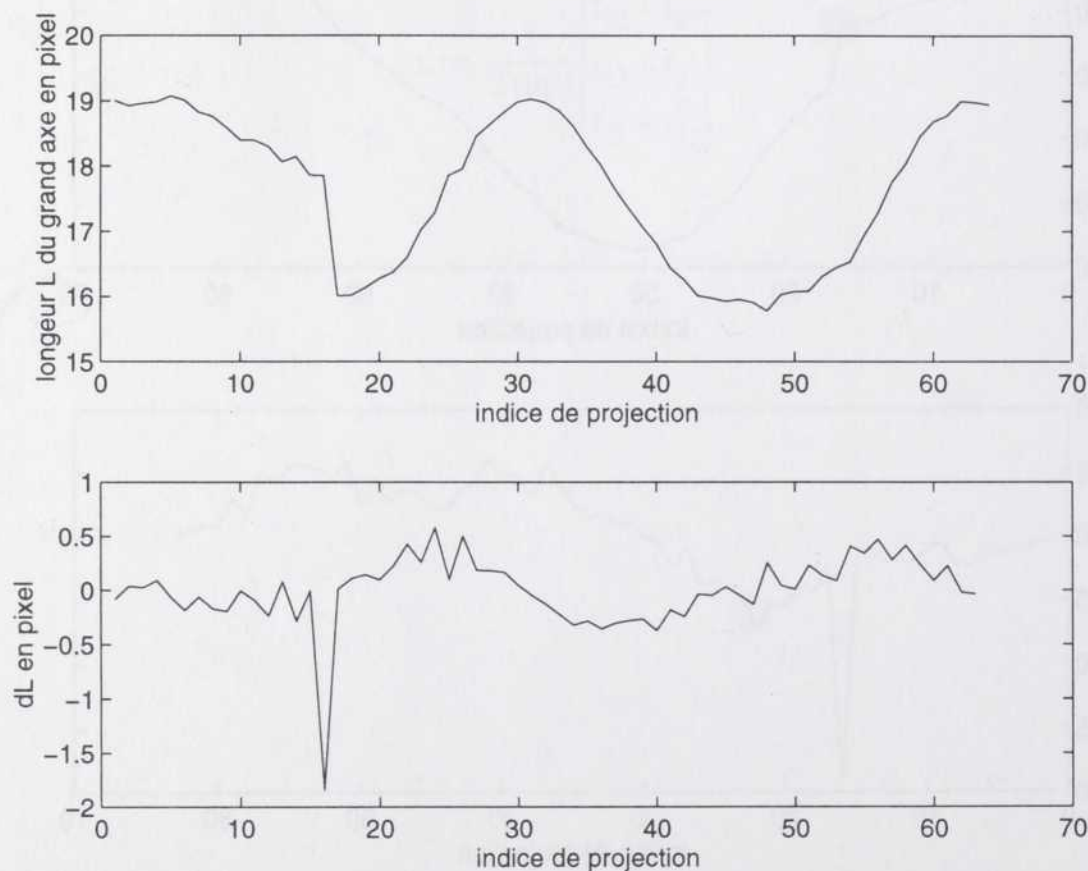


Figure 2 Demi-longueur du grand axe en fonction de l'indice de projection (en haut); différence des demi-longueurs en fonction de l'indice de projection (en bas).

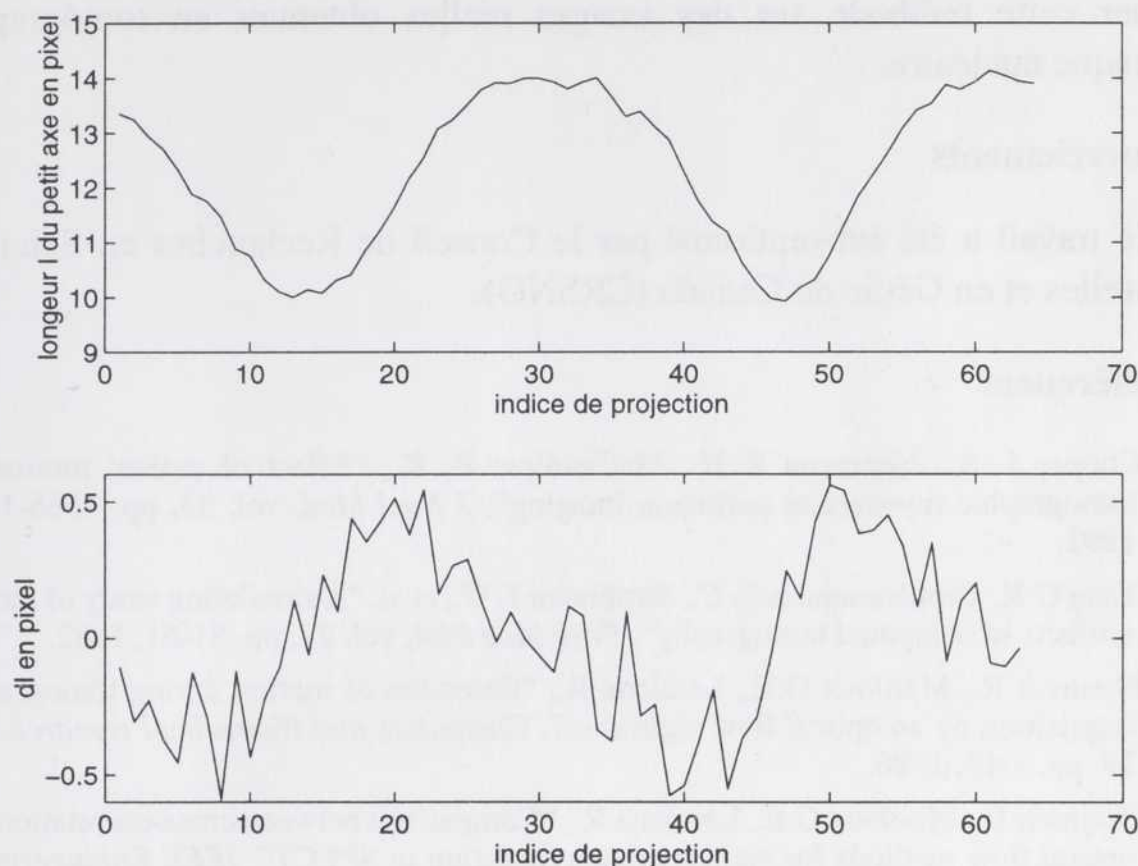


Figure 3 Demi-longueur du petit axe en fonction de l'indice de projection (en haut); différence des demi-longueurs en fonction de l'indice de projection (en bas).

4. Conclusion

Nous avons présenté une méthode pour détecter et quantifier la rotation tridimensionnelle d'une ellipsoïde à partir de ses projections orthographiques. En étudiant la variation de l'orientation des ellipses de projection ainsi que la variation des longueurs de leurs axes en fonction de l'angle de projection, la rotation tridimensionnelle entre deux projections est détectée; l'indice de projection où cette rotation a lieu est aussi déterminé. Deux ellipsoïdes sont alors reconstruites en utilisant les données de projection avant et après la rotation respectivement. La rotation tridimensionnelle est quantifiée en déterminant la rotation entre les axes de ces deux ellipsoïdes. L'angle de rotation et l'axe de rotation sont déterminés. Cette méthode est utile en tomographie cardiaque. En effet, en modélisant le coeur à l'aide d'une ellipsoïde, elle nous permettrait de détecter et de quantifier le mouvement du patient durant une acquisition tomographique. La détection du mouvement est un indice de la qualité de l'acquisition tomographique. L'algorithme de reconstruction tomographique peut tenir compte du mouvement quantifié pour améliorer la reconstruction. Nous poursuivons actuellement les travaux pour

utiliser cette méthode sur des images réelles obtenues en tomographie cardiaque nucléaire.

Remerciements

Ce travail a été subventionné par le Conseil de Recherches en Sciences Naturelles et en Génie du Canada (CRSNG).

5. Références

- [1] Cooper J. A., Neumann P. H., McCandless B. K., "Effect of patient motion on tomographic myocardial perfusion imaging", *J Nucl Med*, vol. 33, pp. 1566-1571, 1992.
- [2] Yang C-K, Orphanoudakis S. C., Strohbehn J. W., et al. "A simulation study of motion artifacts in computed tomography", *Phys Med Biol*, vol. 27, pp. 51-61, 1982
- [3] Noumeir R., Mailloux G.E., Lemieux R., "Detection of motion during tomographic acquisition by an optical flow algorithm", *Computers and Biomedical research*, vol. 29, pp. 1-15, 1996.
- [4] Noumeir R., Mailloux G.E., Lemieux R., "Comparison between cross-correlation and optical flow methods for patient motion detection in SPECT", *IEEE Engineering in Medicine & Biology, 17th annual conference*, 1995.
- [5] Cooper J. A., Neuman P.H., McCandless B.k. "Detection of patient motion during tomographic myocardial perfusion imaging", *J Nucl Med*, vol. 34, pp. 1341-1348, 1993.
- [6] Yuen H.K, Illingworth J., Kittler J., "Detecting partially occluded ellipses using Hough transform", *Image Vision Computing*, vol. 7, pp. 31-37, 1989.
- [7] Karl W. C., Verghese G. C., Willsky A. S., "Reconstructing ellipsoids from projections", *CVGIP: Graphical Models and Image Processing*, vol. 56, pp. 124-139, 1994.
- [8] Strang G., *Linear Algebra and its Applications*, Saunders HBJ, 1988.
- [9] Zeid I., *CAD/CAM Theory and Practice*. McGraw-Hill, 1991.
- [10] Stewart J., Davison T.M.K., Hamilton O.M.G., Laxton J., Lenz M.P., *Algebra and Geometry*, McGraw-Hill 1989.

Table ronde

Président de séance : M. J. L. J. J. J.

Président de séance : M. J. L. J. J. J.

Président de séance : M. J. L. J. J. J.

Table ronde

... les ...

Remerciements

Ce travail a été soutenu par le Conseil de Recherche en Santé Humaine et en Études du Corps (CHRC) ...

Références

- [1] ...
- [2] ...
- [3] ...
- [4] ...
- [5] ...
- [6] ...
- [7] ...
- [8] ...
- [9] ...
- [10] ...
- [11] ...
- [12] ...
- [13] ...
- [14] ...
- [15] ...
- [16] ...
- [17] ...
- [18] ...
- [19] ...
- [20] ...

Résumé

Le colloque a permis d'appliquer de dégager la nature du ressort du connu lorsqu'il

Les interventions

Discussions

Jean des concepts artificiels consistant d'intelligence

Haute

nous son en me humain modèle d'association règles gramm

Table ronde

Participants du colloque

Présidée par Jean-Pierre Tubach, ENST, Paris

Compilé par Richard Lepage, ETS

Résumé

Le colloque s'est terminé par une table ronde réunissant la majorité des participants au colloque. Le Dr. Jean-Pierre Tubach, notre conférencier invité, animait cette table ronde qui a permis aux participants de débattre des grands enjeux de l'intelligence artificielle appliquée aux technologies de l'information. Le débat fut des plus profitables et a permis de dégager quelques axes majeurs de recherche en intelligence artificielle. Le débat s'est naturellement engagé très tôt sur l'approche connexionniste vs. l'approche symbolique. Il en ressort de la discussion que nous devrions orienter nos recherches vers une collaboration du connexionnisme et du symbolique, notamment à l'aide de systèmes multi-agents, lorsqu'il apparaît approprié de le faire.

Les propos rapportés ci-après ne sont qu'un échantillonnage bien incomplet des interventions de quelques intervenants. J'ai tenté de faire ressortir les idées maîtresses des interventions.

Discussions

Jean-Pierre Tubach lance le débat en faisant remarquer que la majorité des communications présentées lors du colloque essaient de modéliser la perception humaine à l'aide de différentes techniques d'intelligence artificielle. Une avenue très intéressante pour la poursuite de nos recherches consisterait en l'exploration de la fusion de différentes méthodologies d'intelligence artificielle.

Hail Mallouche s'interroge sur la nature de l'intelligence humaine que nous tentons d'utiliser comme modèle. Une première piste d'explication est son fonctionnement apparent par association. **Jean-Guy Meunier** enchaîne en mentionnant plusieurs modèles de fonctionnement de l'intelligence humaine. Ces modèles sont pour la plupart multidisciplinaires. Un premier modèle, numérique, consiste à trouver des classes de représentation par cumul d'associations. Un second modèle, logico-catégoriel, fonctionne selon des règles rigoureuses et est à l'origine du débat entre cette approche symbolique, grammaticale, et les modèles associatifs. Les bas niveaux de l'intelligence

humaine sont souvent associés aux modèles associatifs tandis que les hauts niveaux sont associés aux modèles symboliques, à base de règles de production. **Daniel Crevier** et **Robert Sabourin** proposent l'exploration des systèmes multi-agents pour la fusion des approches, afin de faire coopérer les classificateurs.

Robert Proulx avance l'hypothèse que tout, à la limite, est basé sur le modèle associatif. Les difficultés actuelles des modèles connexionnistes sont en grande partie causées par le fait qu'ils sont utilisés pour analyser et apprendre des patrons d'entrée. Les modes de représentation de plus haut niveau pourraient être modélisés en s'appuyant sur la notion de vecteurs et de vecteurs propres, au lieu de s'appuyer sur la notion de patron.

Robert Proulx relance le débat en posant deux catégories de recherche en intelligence artificielle: la recherche *dure* qui a comme principal objectif la modélisation de l'intelligence humaine et la recherche *faible* qui vise à utiliser les techniques de l'intelligence artificielle dans un but d'application aux technologies d'information. **Jean-Pierre Tubach** signale que la distinction n'est pas si claire. Par exemple, la théorie des champs de Markov, un domaine de recherche dit *faible*, n'est pas suffisant pour comprendre le mécanisme du langage. **Jean-Guy Meunier** rappelle qu'on est en présence de deux courants de pensée dominants: tout est associatif, selon un courant de pensée, et tout est inférentiel, selon le second. **Daniel Crevier** mentionne un troisième courant de pensée: une technique d'intelligence artificielle doit être programmée avec un potentiel explicatif si l'on veut garder le contrôle de la machine construite. **Jean-Jules Brault** souligne pour sa part que l'on associe, depuis le début du débat, les modèles connexionnistes aux modèles associatifs, mais que les réseaux neuroniques artificiels ne sont pas que seulement associatifs. Il existe bien d'autres modèles non-associatifs. **Denis Gingras**, pour sa part, souligne l'approche de l'ingénieur pour la réalisation de systèmes d'intelligence artificielle. La réalisation de tels systèmes doit forcément être fondée sur la modularité et se heurte à de nombreuses difficultés de réalisation matérielle. En effet, il faut tenir compte des contraintes et des limites des technologies disponibles dans le processus de conception d'un système d'intelligence artificielle. De plus, l'aspect temporel de l'apprentissage constitue un défi majeur. Le bagage génétique optimisé de l'être humain a nécessité des millions d'années à se bâtir. Pour un enfant, l'apprentissage du langage et de la reconnaissance visuelle d'objets tridimensionnels nécessite de nombreuses

années. Comment peut-on par exemple matérialiser un tel bagage de connaissances dans les poids de réseaux de neurones artificiels forcément limités en étendue physique et en capacité? Comment peut-on réaliser un tel apprentissage à l'aide d'un système d'intelligence artificielle lorsque nous ne disposons en pratique que d'un laps de temps très court de quelques heures ou de quelques jours tout au plus pour «enseigner» au système? Ce défi demeure loin d'être résolu. **Mohamed Cheriet** suggère d'aborder ces problèmes complexes à plusieurs niveaux d'abstraction. Des systèmes multi-agents, composés d'agents neuroniques, à base de grammaires, syntaxiques, à base de règles, etc., pourraient coopérer dans la recherche de solutions. Il faut aussi distinguer les systèmes multi-agents, capables de résoudre plusieurs problèmes en collaboration, des systèmes multi-experts, qui contribuent à résoudre un seul problème, mais avec la contribution de plusieurs avis experts. **Jean-Pierre Tubach** rappelle que cette problématique de coopération multiple fut rencontrée en premier dans le traitement de la parole, au début des années 70. **Robert Proulx** soumet qu'on n'a pas à privilégier une approche plutôt qu'une autre. L'approche à favoriser dépend beaucoup de l'application.

Conclusion

Les discussions furent, de l'avis de tous, très enrichissantes. Tous repartent avec des projets de collaboration ou simplement avec de nouvelles pistes de recherche à explorer. Plusieurs suggèrent qu'un tel colloque devrait être annuel, pour nous permettre de se ressourcer et échanger avec les collègues. L'idée de la table ronde a plu aux participants et il est suggéré, pour les prochains colloques, d'y poser une question à débattre, avec une ébauche de réponse écrite par les participants et qui servirait de point de départ de la discussion.







Les technologies de l'information constituent un enjeu majeur pour la société moderne. Elles touchent autant le développement explosif de l'informatique que celui des télécommunications. Ce recueil d'articles examine quatre domaines clés des technologies de l'information : la vision par ordinateur, l'analyse de l'écrit, le traitement de la parole et les applications dans l'imagerie biomédicale. Nous y examinons l'état de la recherche francophone selon trois approches inspirées de l'intelligence artificielle : les réseaux neuroniques artificiels, les systèmes symboliques à base de connaissance et les algorithmes numériques.

Ont collaboré à cet ouvrage :

Alain Albert, Abdenour Bouzouane, Jean-Jules Brault, Jean Bégin, Mohamed Cheriet, Daniel Crevier, Myriam Côté, Jean Dansereau, Jacques A. de Guise, Robert Desjardins, Pierre Dumouchel, Denis Gingras, Yves Goussard, Alexandre Guimond, Bernard Imbert, Steve Iris, Christian Janicki, Dave Aimé-Désiré Kodjo, Eric Lecolinet, Nicolas Ledoux, Richard Lepage, Ghislain Lévesque, Hail Mallouche, Guy Massicotte, Jean Meunier, Jean-Guy Meunier, Georges Nault, Rita Noumeir, Douglas O'Shaughnessy, Guturu Parthasarathy, Julie Potvin, Robert Proulx, Yvan Ross, Jean Rouat, Denis W. Roy, Robert Sabourin, Joseph N. Said, Roland Schaefer, Jean-Paul Soucy, Benoît A. St-Onge, Ching Y. Suen, Marie-Hélène Talon, Jean-Pierre Tubach, Rivarol Vergin.

ISBN 2-89245-145-0



9 782892 451450

BNQ



000 251 755