Quantification de la qualité d'un geste chirurgical à partir de connaissances a priori

Rémi Wolf

▶ **To cite this version:**

Rémi Wolf. Quantification de la qualité d'un geste chirurgical à partir de connaissances a priori. Modélisation et simulation. Université de Grenoble, 2013. Français. NNT : . tel-00965163v1

HAL Id: tel-00965163

https://theses.hal.science/tel-00965163v1

Submitted on 24 Mar 2014 (v1), last revised 5 Jul 2017 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Modèles, Méthodes et Algorithmes pour la Biologie, la Santé et l'Environnement**

Arrêté ministériel : du 7 août 2006

Présentée par

Rémi WOLF

Thèse dirigée par **Philippe CINQUIN**
et codirigée par **Sandrine VOROS**

préparée au sein **du laboratoire Techniques de l'Ingénierie Médicale et de la Complexité - Informatique, Mathématiques et Applications de Grenoble (TIMC-IMAG)**
et de **l'école doctorale Ingénierie pour la Santé, la Cognition et l'Environnement (EDISCE)**

Quantification de la qualité d'un geste chirurgical à partir de connaissances a priori

Thèse soutenue publiquement le **17 juin 2013**,
devant le jury composé de :

Mme. Jocelyne TROCCAZ
DR CNRS, TIMC-IMAG, Grenoble, Présidente
M. Christophe DOIGNON
PU, ICube, Université de Strasbourg - CNRS, Rapporteur
M. Pierre JANNIN
CR INSERM, LTSI, Université de Rennes - INSERM, Rapporteur
M. Aurel MESSAS
MD, Hôpital Max Fourestier, Nanterre, Examinateur
M. Nassir NAVAB
PU, Fakultät für Informatik, Technische Universität München, Allemagne, Examinateur
M. Philippe CINQUIN
PU-PH, TIMC-IMAG, UJF-CNRS, Grenoble, Directeur de thèse
Mme. Sandrine VOROS
CR INSERM, TIMC-IMAG, UJF-CNRS, Grenoble, Co-Directrice de thèse

Remerciements

Je voudrais tout d'abord remercier Philippe Cinquin et Sandrine Voros pour m'avoir proposé ce sujet de thèse et m'avoir accordé leur confiance durant ces quatre années, remontant parfois la mienne lorsque c'était nécessaire, ainsi que pour le bagage scientifique qu'ils m'ont apporté. Je remercie en particulier Philippe de m'avoir incité à mener ce travail de thèse en parallèle de l'externat, et d'avoir rendu ce double-cursus réalisable.

Je remercie également Christophe Doignon et Pierre Jannin d'avoir accepté de rapporter ce travail, ainsi que Aurel Messas, Nassir Navab et Jocelyne Troccaz d'avoir accepté de l'évaluer en faisant partie du jury.

Je remercie vivement tous les médecins avec lesquels j'ai pu collaborer au cours de ce travail, en particulier Jean-Alexandre Long et Gaëlle Fiard pour les longues heures passées à visionner les enregistrements des volontaires, Yohann Robert pour les manipulations au laboratoire d'anatomie, ainsi que tous les volontaires qui ont accepté de se prêter au jeu du banc d'entraînement. Je remercie par ailleurs les médecins et chefs de service du CHU qui m'ont soutenu dans ce projet et aidé à acquérir les bases du savoir médical, en particulier François Berger, Daniel Fagret et Olivier Palombi au moment de la passerelle, ainsi que Raphaël Briot qui m'a permis de garder un pied dans la médecine au cours de la finalisation de cette thèse.

Merci également à tous les membres de l'équipe GMCAO que j'ai côtoyés au cours de cette thèse, qui ont rendus ces années particulièrement agréables et ont enrichi ce travail de leur vision : mes co-bureaux actuels, Mathieu et Vincent, les passés, Marek et Olivier, ainsi que Adeline, Anselme, Anthony, Aurélien, Bérangère, Brahim, Céline, Christophe, Gabriel, Gaël, Ivan, Johan, Laurent, Manhu, Matthieu, Micha, Nassos, Nicolas, Nikolaï, Sonia, Yannick, Yohan. Un merci particulier à Corinne pour son soutien indispensable, ainsi qu'à Jocelyne pour permettre à cette équipe de si bien vivre. Merci aussi aux autres membres du laboratoire qui m'ont épaulé, en particulier l'équipe système ainsi que Maud et Alexandre qui ont apporté une contribution scientifique majeure à ce travail grâce à leurs talents en statistiques.

Enfin, je remercie ma famille pour son soutien et pour m'avoir permis d'arriver où je suis aujourd'hui, ainsi que mes amis pour les moments d'oxygénation en montagne ou ailleurs pendant ces quatre années. Merci à Jean-Luc pour m'avoir guidé depuis le jour où j'ai choisi la médecine.

Juliette, merci pour tout, ton soutien inconditionnel, tes encouragements perpétuels et ta présence passée, présente et future.

TABLE DES MATIÈRES

Table des figures **11**

Liste des tableaux **15**

Introduction **17**

I Évaluation de la qualité des gestes en chirurgie : un état de l'art **21**

1 Contexte clinique **23**
1.1 Justification de l'évaluation de la qualité des gestes en chirurgie . 23
1.1.1 Une demande de réduction des aléas chirurgicaux 24
1.1.2 Une évolution du modèle de compagnonnage 25
1.1.3 Au delà de l'erreur 27
1.2 Spécificités liées à la laparoscopie 28

2 Contexte scientifique **33**
2.1 Méthodes de description d'une procédure chirurgicale 34
2.1.1 Notion de granularité et modélisation de procédures chirurgicales 34
i) Objectifs 34
ii) Formalisme et définitions 34
2.1.2 Reconnaissance de tâches en fonction du contexte 36
2.2 Méthodes de quantification existantes 38
2.2.1 Quelles sont les propriétés d'une méthode d'évaluation idéale ? 39
2.2.2 Systèmes de notation structurée 40
i) Objective Structured Assessment of Technical Skills 40
ii) Global Operative Assessment of Laparoscopic Skills 43
iii) Observational Clinical Human Reliability Assessment 45

2.2.3 Évaluation à l'aide de statistiques descriptives 47
i) Analyse du résultat 47
ii) Analyse du mouvement 49
2.2.4 Modélisations statistiques 56
i) Modèles de Markov et modèles de Markov cachés . 56
ii) *Dynamic Time Warping* 58
2.3 Conclusion et problématique de la thèse 58

II Développement d'une méthode de suivi des instruments chirurgicaux à partir d'images endoscopiques 69

3 Détection d'instruments chirurgicaux 71
3.1 Suivi des instruments chirurgicaux : méthodes existantes 72
3.1.1 Quelles sont les propriétés d'une méthode de mesure idéale ? 72
3.1.2 Méthodes basées sur des informations non-visuelles 72
i) Données de cinématique des robots chirurgicaux . . 73
ii) Données des simulateurs de réalité virtuelle 73
iii) Localisateurs optiques, magnétiques, acoustiques ou mécaniques 74
3.1.3 Méthodes basées sur le traitement d'images endoscopiques . . 78
i) Segmentation de l'instrument 79
ii) Optimisation de la vitesse du suivi 80
3.2 Approche développée 81
3.2.1 Cadre de travail général 81
3.2.2 Calibrage de la caméra 83
3.2.3 Représentation géométrique de l'instrument 85
3.2.4 Traitement de l'image 88
i) Choix de l'espace colorimétrique 88
ii) Filtrage 90
iii) Détection des angles dans l'image par l'algorithme de Canny 93
iv) Définition des seuils du filtre de Canny 96
3.2.5 Détermination de la position du point d'insertion dans le repère lié à la caméra 105
3.2.6 Représentation spatiale de l'orientation de l'instrument . . . 109
i) Représentation discrète de l'orientation sur une sphère 109
ii) Structure des données 111
3.2.7 Algorithme de CONDENSATION 114
i) Principe général 114
ii) Modèle particulaire de l'orientation de l'instrument 115
iii) Mise à jour des poids des particules 117
3.2.8 Détermination profondeur d'insertion 119

i) Partition en classes par l'algorithme d'Otsu 119
ii) Autres méthodes explorées 123
3.2.9 Améliorations apportées à la méthode 124
3.3 Résultats bruts 124

4 Filtrage des trajectoires et évaluation du suivi 133
4.1 Filtrage des trajectoires 134
4.1.1 Nature du bruit enregistré 134
4.1.2 Filtrage des mesures aberrantes 135
4.1.3 Lissage de la trajectoire obtenue 137
4.2 Précision de la mesure et influence des paramètres 139
4.2.1 Dispositif d'enregistrement des trajectoires d'évaluation . . . 139
4.2.2 Définition des trajectoires réelles 140
4.2.3 Mise en évidence d'une erreur systématique dans l'estimation de la position de l'instrument dans l'espace 141
4.2.4 Erreur dans l'estimation de la position de l'instrument dans le plan image 146
4.2.5 Analyse de l'influence du choix des paramètres 146
i) Influence du nombre de particules pour l'orientation de l'instrument 148
ii) Influence du nombre de segments explorés pour identifier la pointe de l'instrument 149
iii) Influence de la résolution de la grille géodésique . . 149
iv) Influence du type de fond 151
v) Influence du choix du filtre 152
vi) Influence du nombre d'instruments suivis 152
4.2.6 Évaluation sur spécimen anatomique 153
4.3 Discussion et conclusion pour le suivi d'instruments 154
4.3.1 Limites actuelles de la méthode proposée 154
4.3.2 Apports principaux de la méthode proposée 156
4.3.3 Optimisations envisagées 157
4.3.4 Conclusion 159

III Analyse de la qualité de gestes chirurgicaux à partir de connaissances a priori 163

5 Acquisition de gestes chirurgicaux 165
5.1 Banc d'entraînement MISTELS et score FLS 166
5.2 Évaluation de la qualité du geste à l'aide du score GOALS 169
5.3 Données acquises 170
5.3.1 Volontaires 170
5.3.2 Chaîne d'acquisition des données 171

5.3.3 Exemple de trajectoires obtenues . . . 172
5.4 Métriques explorées . . . 173

6 Analyse des gestes chirurgicaux **179**
6.1 Reproductibilité intra- et inter-observateurs . . . 180
6.2 Validité du score GOALS pour le banc d'essai MISTELS . . . 182
6.2.1 Validité concurrente . . . 182
i) Outils statistiques utilisés pour la corrélation . . . 182
ii) Moyenne des répétitions et adaptation du score MISTELS . . . 182
iii) Corrélation entre le score MISTELS et le score GOALS 183
6.2.2 Étude des composantes du score GOALS expliquant le score MISTELS . . . 184
6.2.3 Étude de la validité prédictive . . . 186
6.3 Prédiction d'un score à partir des métriques acquises . . . 189
6.3.1 Outils statistiques utilisés . . . 189
6.3.2 Plan d'analyse . . . 192
6.3.3 Métriques significatives dans le modèle prédictif pour chaque exercice . . . 193
6.3.4 Prédiction d'un score à partir de l'ensemble des métriques . . . 195
i) Prédiction du score GOALS pour chaque exercice et du score GOALS moyen . . . 195
ii) Validité prédictive du score obtenu avec la régression PLS . . . 198
6.3.5 Prédiction d'un score à partir des métriques significatives de chaque exercice . . . 200
6.4 Classification en fonction des métriques uniquement . . . 203
6.5 Prédiction de chaque aptitude du score GOALS . . . 204
6.6 Discussion et conclusion pour l'analyse de gestes chirurgicaux . . . 206
6.6.1 Banc d'entraînement MISTELS . . . 207
6.6.2 Définition du score de référence . . . 208
6.6.3 Choix des métriques explorées . . . 208
6.6.4 Méthode de prédiction . . . 209
6.6.5 Limites de la prédiction . . . 209
6.6.6 Apports principaux de la méthode . . . 210
6.6.7 Conclusion . . . 211

Discussion, conclusion et perspectives **213**

7 Discussion générale **215**
7.1 Acceptabilité par les chirurgiens de dispositifs d'assistance par ordinateur . . . 215

7.1.1 Acceptabilité des dispositifs d'évaluation 215
7.1.2 Acceptabilité de dispositifs d'aide per-opératoire 216
7.2 Fusion avec des modèles de procédure chirurgicale 217
7.3 Vers l'optimisation du geste chirurgical 218
7.3.1 Compétences non-techniques du chirurgien 218
7.3.2 Apport d'informations contextualisées au chirurgien 218

8 Conclusion et perspectives 221

Annexes 225

A Résultats de la régression PLS pour les exercices 227

B Résultats de la régression PLS pour les aptitudes 231

TABLE DES FIGURES

1.1 Illustration d'une intervention par laparoscopie 29

2.1 Exemple de système de reconnaissance de tâches, basé sur les instruments utilisés à un instant donné, extrait de [Padoy et al., 2012] 36

3.1 Robot chirurgical daVinci avec API 74
3.2 Interfaces d'acquisition de la position des instruments pour les simulateurs de réalité virtuelle 75
3.3 Dispositifs de suivi d'instruments laparoscopiques 77
3.4 Degrés de liberté d'un instrument de laparoscopie 81
3.5 Diagramme global de la méthode de suivi d'instrument 83
3.6 Modèle du sténopé pour la caméra 84
3.7 Représentation géométrique de l'instrument dans l'espace 86
3.8 Détermination de la position dans l'espace de l'instrument à partir de ses bords dans l'image. 87
3.9 Détermination de la position des bords à partir de l'axe et d'un point de passage de l'instrument. 88
3.10 Représentations graphiques des espaces colorimétriques TSL et TSV . . . 90
3.11 Comparaison entre les deux modes de calcul de la saturation 91
3.12 Illustration du fonctionnement du filtre bilatéral 92
3.13 Variations du filtrage bilatéral en fonction de σ_d et σ_r 93
3.14 Évolution du temps de calcul pour différents filtres de bruit en fonction de leur diamètre 94
3.15 Effet du diamètre du filtre bilatéral sur l'image de saturation 95
3.16 Illustration de l'effet du filtrage bilatéral sur la détection des angles par l'algorithme de Canny 96
3.17 Variations de la détection des angles avec la méthode Canny en fonction des seuils T_h et T_b et du diamètre d du filtre bilatéral 97

3.18 Histogramme cumulé des valeurs du gradient de saturation 98
3.19 Illustration des images PGT^j 100
3.20 Processus de sélection de l'EGT à partir des PGT^j, et du triplet de paramètres optimaux à partir des C^i et de l'EGT. 102
3.21 Mise en évidence de la sensibilité des paramètres étudiés aux changements de phases au cours d'une chirurgie 103
3.22 Robustesse de la méthode d'optimisation des paramètres des filtres utilisés dans différents contextes d'utilisation 104
3.23 Détermination du point d'insertion à l'aide d'un dispositif de suivi Polaris . 106
3.24 Détermination du point d'insertion par analyse d'image et sélection des droites candidates par RANSAC 107
3.25 Orientations possibles de l'instrument autour du point d'insertion 109
3.26 Construction itérative de la sphère géodésique 110
3.27 Pavage de la grille sphérique géodésique avec des cellules de Voronoi . . . 111
3.28 Représentation de la partition de la grille géodésique sphérique en panneaux de cellules 112
3.29 Cas général de l'algorithme de CONDENSATION 115
3.30 Transition entre le pavage hexagonal et le pavage rectangulaire, avec conservation de la densité de probabilité sur la sphère. 117
3.31 Distribution des valeurs des pixels le long de la projection de l'axe central de l'instrument 120
3.35 Suivi simultané de deux instruments en utilisant une géode centrée sur chaque point d'insertion 125
3.36 Étape de calibrage de la caméra avec la procédure de Zhang [2000] 126
3.37 Position des points d'insertion avec plusieurs déterminations itératives . . 126
3.38 Exemples d'image de suivi sur un fond de prostatectomie radicale laparoscopique. 127
3.39 Représentation de la trajectoire de la pointe dans l'espace avant filtrage . . 128

4.1 Erreur dans la détermination de l'orientation de l'instrument liée à un échec de la détection d'un des bords 134
4.2 Erreur de mesure dans l'espace en fonction de l'angle de l'instrument par rapport au plan image 136
4.3 Variation fine de l'orientation de l'instrument entre deux images successives. 136
4.4 Exemples de trajectoires avant et après exclusion des mesures aberrantes . 138
4.5 Effet du lissage par une spline 139
4.6 Suivi de la position de l'instrument à l'aide du dispositif accuTrack 250. . . 141
4.7 Erreur angulaire systématique dans l'estimation de l'orientation de l'instrument 142
4.8 Erreur systématique dans l'estimation de la position dans l'espace de l'instrument 142

4.9 Image correspondant à une détection de bonne qualité visuellement mais associée à une erreur importante . . . 143
4.10 Projection de la position de l'instrument sur les axes X et Z . . . 144
4.11 Erreurs de re-projection liées aux incertitudes des calibrages intrinsèque et extrinsèque . . . 145
4.12 Trajectoires obtenues après recalage rigide. . . . 145
4.13 Projection de la position de l'instrument sur les axes X et Z après recalage rigide . . . 146
4.14 Erreur sur la détermination de l'orientation et de la position de la pointe de l'instrument dans le plan image . . . 147
4.15 Influence du nombre de particules sur la précision de la détermination de l'orientation de l'instrument . . . 148
4.16 Effet sur la vitesse de calcul du nombre de particules . . . 149
4.17 Influence du nombre de segments explorés le long de l'axe sur la précision de la détermination de la profondeur d'insertion. . . . 150
4.18 Influence de la résolution de la grille géodésique sur la précision de la détermination de l'orientation de l'instrument . . . 150
4.19 Effet des filtres utilisés pour réduire le bruit dans l'image sur la précision de la détection de l'instrument. . . . 152
4.20 Acquisition de données en conditions réelles au laboratoire d'anatomie . . . 154

5.1 Illustration des exercices du banc d'entraînement MISTELS . . . 168
5.2 Diagramme à moustaches des auto-évaluations de la compréhension et de l'expertise en chirurgie laparoscopique par les volontaires, en fonction de leur statut . . . 171
5.3 Exemples de trajectoires pour différents exercices . . . 173
5.4 Exemples de volumes et de couplage entre les paramètres vitesse, accélération et jerk pour différents exercices . . . 177

6.1 Comparaison des scores GOALS moyens G entre les deux évaluateurs . . . 181
6.2 Droite de régression linéaire entre le score GOALS et le score MISTELS . 185
6.3 Classification des volontaires en fonction de leur score GOALS et de leur score MISTELS . . . 187
6.4 Illustration de projection dans le plan des composantes des observations avec un modèle PLS . . . 191
6.5 Construction de la régression PLS . . . 192
6.6 Régression linéaire entre le score GOALS moyen réel G et le score GOALS moyen prédit $\tilde{G}$ . . . 196
6.7 Score GOALS prédit $\tilde{G}_j$ en fonction du score GOALS réel G_j pour chaque exercice . . . 197
6.8 Classification des volontaires par un classifieur *k-means* basé sur le score GOALS prédit dans les cinq exercices . . . 199
6.9 Classification *k-means* pour chaque exercice basée sur le score GOALS prédit 201

6.10 Régression linéaire entre le score GOALS moyen réel G et le score GOALS moyen prédit à partir d'un sous-ensemble de métriques $\tilde{g}$ 202

A.1 Représentation des variables sur les deux premières composantes des régressions PLS 230

B.1 Représentation des variables sur les deux premières composantes des régressions PLS pour modéliser les aptitudes 234
B.2 Régressions linéaires entre les scores GOALS prédits et réels pour les aptitudes. 235
B.3 Comparaison entre les classifications *k-means* obtenues avec le score GOALS réel et le score GOALS prédit pour chaque aptitude 236

LISTE DES TABLEAUX

1.1 Les différents types d'erreurs 26

2.1 Échelle d'évaluation globale OSATS 42
2.2 Échelle d'évaluation GOALS 44
2.3 Métriques d'évaluation pour l'analyse du mouvement 53
2.4 Comparaison des caractéristiques des différentes méthodes d'évaluation de la qualité du geste 59

3.1 Distance entre deux points voisins de l'espace correspondant à une profondeur donnée de l'instrument 113
3.2 Précision de la méthode avant traitement des trajectoires 128

4.1 Vitesses de calcul de la méthode de suivi en fonction du filtrage appliqué à l'image de saturation 152
4.2 Vitesses de calcul de la méthode de suivi en fonction du nombre d'instruments suivis 153

6.1 Reproductibilité intra- et inter-observateur du score GOALS 180
6.2 Coefficients de corrélation entre les scores MISTELS et l'auto-évaluation des chirurgiens 183
6.3 Coefficients de corrélation entre les scores MISTELS et le score GOALS . . 184
6.4 Aptitudes du score GOALS les plus significatives pour expliquer le score MISTELS 186
6.5 Exemple de matrice de confusion 188
6.6 Taux de reconnaissance entre groupes basés sur des scores et groupes basés sur le statut 189
6.7 Paramètres les plus significatifs pour les différents exercices dans les modèles prédictifs PLS 194
6.8 Coefficients de corrélation entre les socres GOALS prédits et réels. 195

6.9 Erreurs quadratiques moyennes des prédictions en fonction des groupes de volontaires 198
6.10 Matrice de confusion obtenue avec la classification *k-means* effectuée sur le score GOALS prédit. 200
6.11 Coefficients de corrélation entre les scores GOALS prédits et réels. 200
6.12 Taux de reconnaissance pour les classifications obtenues à partir d'un sous-ensemble de métriques. 203
6.13 Taux de reconnaissance de la classification obtenue à partir des métriques significatives pour chaque exercice 204
6.14 Paramètres les plus significatifs pour les différentes aptitudes élavluées par le score GOALS dans les modèles prédictifs PLS 205

INTRODUCTION

Cette thèse, dont le sujet est le développement d'une méthode de quantification de la qualité de gestes opératoires à partir de connaissances a priori, a été réalisée au sein de l'équipe Gestes Médico-Chirurgicaux Assistés par Ordinateur (GMCAO) du laboratoire Techniques de l'Ingénierie Médicale et de la Complexité - Informatique, Mathématiques et Applications de Grenoble (TIMC-IMAG), sous la direction de Philippe Cinquin et de Sandrine Voros.

Les gestes médico-chirurgicaux assistés par ordinateur ont pour objectif d'aider le médecin ou le chirurgien à offrir des soins de la meilleure qualité possible aux patients, en guidant le praticien dans la réalisation de gestes diagnostiques et thérapeutiques, grâce à l'exploitation :

- d'informations médicales multi-modales visant à améliorer la *perception* du contexte dans lequel le praticien intervient : ces informations doivent être pertinentes par rapport à l'objectif recherché, et peuvent être acquises à différents stades de la prise en charge (pré-, per- et post-opératoire par exemple pour la chirurgie) ;
- de modèles de hiérarchisation et d'interprétation des informations multi-modales, afin d'aider le praticien dans son *raisonnement* pour la planification de la stratégie optimale à adopter pour la prise en charge : ces modèles sont basés sur des connaissances a priori ;
- de dispositifs assistant le praticien dans la réalisation des *actions* nécessaires à cette prise en charge : cette aide peut être apportée par des robots ou des systèmes de navigation et de guidage.

On définit les soins de la meilleure qualité possible comme les soins les plus sûrs, les plus précis et les moins invasifs possibles, assurant des résultats thérapeutiques et fonctionnels optimaux, et avec le moins possible de conséquences négatives, à court, moyen et long terme.

Le contexte médical de ce travail est l'évaluation de la qualité des gestes au cours la chirurgie laparoscopique, cette dernière présentant des difficultés spécifiques pour le chirurgien (modification de la perception visuelle et tactile du site opératoire en particulier). Je m'intéresse donc à la composante *raisonnement* des GMCAO, avec l'objectif de proposer à long terme au chirurgien un système d'analyse de la qualité de son geste, dédié à l'amélioration de celui-ci.

L'amélioration du geste chirurgical peut s'appuyer sur de nombreuses composantes exploitant l'analyse de la qualité de ce dernier :

- la quantification des capacités du chirurgien : il s'agit de l'habileté technique, mais également de facultés intellectuelles telles que la connaissance du sujet, la capacité de reconnaissance d'une situation ou encore la capacité à prendre une décision optimale au moment voulu ;
- l'apport d'informations contextualisées au cours de la chirurgie en fonction de l'avancement de cette dernière afin d'optimiser la qualité du geste. Ces informations supplémentaires peuvent provenir de l'extraction de connaissances à partir de situations antérieures similaires (explicitation d'un savoir-faire implicite), d'acquisitions d'images multi-modales pré- et per-opératoires (recalage d'image permettant d'identifier des structures anatomiques) ou encore de la trajectoire des instruments à un instant donné. En exploitant des informations multi-modales et des connaissances a priori, dont la fusion permettra de modéliser la procédure, on peut envisager d'alerter le chirurgien afin qu'il rectifie ses erreurs au cours de la procédure endoscopique, avant que celles-ci n'aient de conséquences pour l'opération en cours.

L'objectif de ce travail en particulier est de développer une méthode d'évaluation de la qualité technique des gestes en chirurgie à partir de connaissances a priori, dans le contexte per-opératoire et sans modifier l'environnement du chirurgien, afin de permettre à terme une amélioration de la qualité de ce geste.

Dans la première partie de cette thèse, je propose une analyse détaillée des contextes clinique et scientifique par rapport auxquels je me situe, et précise sur quels pré-requis je m'appuie pour définir mon cadre de travail. En particulier, j'argumente sur les connaissances a priori nécessaires à ma méthode et présente les solutions proposées jusqu'à présent pour les extraire. Je réalise également un bref état de l'art des différentes méthodes d'évaluation de la qualité du geste chirurgical, en précisant les intérêts et limites de chacune.

Je présente dans la deuxième partie de ce manuscrit la méthode de suivi d'instruments de chirurgie laparoscopique mise au point au cours de cette thèse, qui permet à partir de modèles géométriques et statistiques d'obtenir la trajectoire de la pointe de l'instrument dans l'espace. Je justifie et détaille les solutions adoptées face aux différents verrous rencontrés, et explicite le filtrage final que j'applique à

la trajectoire brute.

Enfin, j'introduis dans la troisième partie un ensemble de métriques dérivées de ces trajectoires, et analyse leur caractère prédictif quant à la qualité technique du geste chirurgical. Je présente le dispositif ayant permis l'acquisition d'un grand nombre de gestes chirurgicaux, réalisés sur banc d'entraînement par des volontaires de différents niveaux en chirurgie. Je définis l'échelle d'évaluation qui servira d'"étalon-or" pour la mesure de la qualité technique du geste, et propose une méthode permettant d'une part d'identifier les métriques les plus significatives pour évaluer cette qualité, d'autre part de prédire un score reflétant celle-ci.

Dans la quatrième et dernière partie, je discute des apports et des limites de cette thèse dans le domaine des GMCAO, et conclus en présentant son articulation par rapport à d'autres méthodes existantes pour construire un système contextualisé d'analyse de la qualité du geste chirurgical.

Première partie

Évaluation de la qualité des gestes en chirurgie : un état de l'art

CHAPITRE 1

CONTEXTE CLINIQUE

Sommaire

1.1 Justification de l'évaluation de la qualité des gestes en chirurgie 23
1.1.1 Une demande de réduction des aléas chirurgicaux 24
1.1.2 Une évolution du modèle de compagnonnage 25
1.1.3 Au delà de l'erreur 27
1.2 Spécificités liées à la laparoscopie 28

1.1 Justification de l'évaluation de la qualité des gestes en chirurgie

Il convient d'abord de définir ce que je qualifie de qualité chirurgicale dans la suite de ce manuscrit. Celle-ci peut être décrite dans de nombreux domaines [Vallancien, 2006] : techniques (la qualité de réalisation de l'intervention), socio-économiques (la qualité de l'offre de soin et la pertinence de l'indication opératoire, le respect des normes de sécurité) ou encore éthiques (la qualité de la prise en charge globale du patient et de son vécu). Les différentes compétences nécessaires pour proposer une prise en charge chirurgicale de qualité dans ces domaines sont multiples, et sont décrites par Ponsky [2004]. Elles comprennent, entre autres :

- la capacité à établir une relation patient-médecin basée sur la confiance, le respect et l'empathie,
- les connaissances médicales théoriques permettant un raisonnement diagnostique et thérapeutique basé sur des preuves,
- la faculté d'analyse de sa pratique et son amélioration,

– l'aptitude à s'inscrire dans un système de soins et à interagir avec d'autres professionnels.

Je me concentre dans ce travail sur l'aspect technique de la qualité de la chirurgie. On peut reprendre pour définir un geste chirurgical de la meilleure qualité technique possible les termes présentés dans l'introduction : il s'agit de la chirurgie la plus sûre, la plus précise et la moins invasive possible, assurant des résultats thérapeutiques et fonctionnels optimaux, et avec le moins possible de conséquences négatives, à court, moyen et long terme.

Il existe actuellement une demande importante d'évaluation et d'amélioration de la qualité de la chirurgie, dont je présente dans la suite de ce chapitre les motivations.

1.1.1 Une demande de réduction des aléas chirurgicaux

Au cours des dernières années, la pression s'est accrue sur les chirurgiens pour qu'ils soient en mesure de démontrer leur niveau de compétence d'une part, et qu'ils réduisent le risque potentiel d'erreurs chirurgicales d'autre part. Cette pression fait suite à la publication de plusieurs rapports montrant qu'une part non négligeable d'erreurs étaient évitables, aussi bien dans les pays anglo-saxons [Linda T. Kohn et Molla S. Donaldson, 2000] qu'en France [Michel et al., 2005], ces erreurs ayant un coût considérable pour la collectivité, au-delà du préjudice individuel porté au patient.

Les études d'analyse du risque en mileu hospitalier [Donaldson, 2007, Vincent et al., 2004] ont insisté sur le rôle des problèmes organisationnels comme origine des erreurs, ce qui explique que la politique d'évaluation [Undre, 2006] et de réduction du risque chirurgical se soit jusqu'à présent principalement orientée vers l'environnement péri-opératoire, avec le développement de procédures qualité (check-lists pré-opératoires, systèmes de gestion de l'occupation des blocs). Cependant Fabri et Zayas-Castro [2008] montrent que les erreurs organisationnelles n'expliquent qu'une minorité des complications chirurgicales, tandis que des erreurs d'exécution, de jugement, d'inattention ou de compréhension contribuent à la majorité des complications.

Ce concept de risque et de réduction du risque applicable dans le domaine de la santé est décrit de façon exhaustive dans Amalberti et al. [2003], et dérive des travaux théoriques de Reason [Reason et Hoc, 1993]. Ce dernier classe les erreurs potentielles en 3 catégories distinctes :

- les erreurs de routine correspondent au fonctionnement fondé sur les habitudes. Il s'agit de défaillances dans la surveillance de l'exécution. L'action se déroule sans contrôle conscient, dans le cadre d'un travail bien connu et

codifié. Le sujet n'a pas pris conscience qu'il y avait un problème. Ce sont les erreurs les plus fréquentes ;

- les erreurs d'activation de connaissance : le sujet est face à une difficulté qu'il ne peut pas résoudre de façon routinière. Il a conscience d'avoir un problème et cherche une solution. L'erreur va résulter d'une mauvaise solution qui résulte elle-même de l'activation d'une mauvaise règle. Cette erreur n'est pas contradictoire avec l'idée que le sujet possède par ailleurs la connaissance de la bonne solution ; mais il n'a pas su l'activer, la retrouver en mémoire, ou n'a pas pu, faute de temps, s'en servir ; une autre solution - moins valide mais immédiatement disponible - s'est imposée à sa logique d'action ;
- les erreurs de possession de connaissance. Le sujet est ignorant de la solution du problème qu'il a à régler. Il mobilise toute sa cognition, lentement, pas à pas, pour produire une nouvelle solution. L'erreur peut alors revêtir différentes formes : bonne solution hors délais, mauvaise solution, etc...

On se rend compte que chacun de ces types d'erreurs est susceptible de se produire au cours d'une intervention chirurgicale, en fonction de la difficulté de la chirurgie et de l'expérience du chirurgien.

L'enjeu de l'évaluation de la qualité des gestes en chirurgie est de limiter autant que possible chacun de ces types d'erreur. Chaque type d'erreur appelle une solution différente :

- l'erreur de routine peut être évitée en détectant la situation à risque à temps, et en retournant une alerte à l'opérateur ;
- l'erreur d'activation de connaissance peut être évitée en reconnaissant un problème donné et en proposant à l'opérateur une solution habituellement utilisée pour résoudre le problème ;
- l'erreur de possession de connaissance peut être évitée en améliorant les dispositifs d'apprentissage et en contrôlant l'aptitude de l'opérateur avant de le laisser réaliser un geste seul.

Le tableau 1.1 résume les caractéristiques principales des différents types d'erreurs.

La quantification de la qualité du geste peut donc contribuer à une stratégie de réduction du risque en identifiant des situations d'erreur potentielle à tous les stades : découverte, apprentissage et optimisation du geste chirurgical.

1.1.2 Une évolution du modèle de compagnonnage

L'apprentissage de la chirurgie est basé depuis plus d'un siècle sur le dogme "*see one, do one, teach one*" conceptualisé par William Halsted de la faculté de médecine de la John Hopkins University [Hamdorf et Hall, 2000], qui veut que l'apprenti se forme aux côtés de son maître au bloc opératoire, en prenant peu à peu des responsabilités dans la réalisation des gestes et en réduisant parallèlement la

Dimension	Erreurs de routine	Erreurs d'activation de connaissance	Erreurs de possession de connaissance
Type d'activité	Actions routinières	Activités de résolution de problème	
Concentration de l'attention	Sur autre chose que la tâche en cours	Sur des considérations liées au problème	
Mode de contrôle de l'erreur	Schémas	Règles stockées	Processus conscients limités
Caractère prévisible de l'erreur	Largement prévisible	Variable	
Fréquence	Élevée dans l'absolu, mais paradoxalement faible en proportion du grand nombre de routines	Dépendante du contexte et de l'expertise du professionnel	Faible dans l'absolu, mais élevée en proportion du très faible nombre de ces situations d'ignorance quasi-totale
Capacités de détection	Élevées	Très faibles sans intervention	

Tab. 1.1. – Les différents types d'erreurs, d'après Reason et Hoc [1993]

supervision. Cette approche rencontre aujourd'hui plusieurs limites, que je présente brièvement ici.

Limitation du temps d'exposition au bloc opératoire

Une limitation à l'apprentissage selon le paradigme précédent est la réglementation croissante du nombre d'heures de travail par semaine pour les internes, avec en particulier l'obligation de respecter un repos compensateur de garde. Cette évolution est probablement bénéfique pour les patients, car elle contribue à réduire les événements indésirables potentiels, aussi bien lors de la réalisation des gestes que lors des prescriptions pré- et post-opératoires. Cependant, elle limite le temps d'exposition des futurs chirurgiens à la pratique de leur future spécialité, allongeant ainsi la durée de leur courbe d'apprentissage [Fischer, 2004], telle que définie par Subramonian et Muir [2004] par exemple. Plusieurs solutions ont été proposées pour contourner cette limitation : des logiciels de mise en situation [Sarker et al., 2009] permettent d'améliorer les processus de prise de désicion, tandis que la pratique des gestes chirurgicaux sur des simulateurs (bancs d'entraînements ou simulateurs virtuels) ou sur des modèles animaux [Roberts et al., 2006] permet d'améliorer la technique chirurgicale. Il apparaît alors nécessaire d'évaluer la qualité des gestes réalisés dans ces conditions, pour autoriser leur retranscription dans l'environnement du bloc opératoire, mais également pour suivre la progression des internes.

Responsabilité et coût : une demande des autorités et du public

L'apprentissage des techniques chirurgicales par les internes au bloc opératoire a un coût, financier mais également en termes de temps d'occupation du bloc, qui est

loin d'être négligeable (surcoût évalué en 1999 à environ 50000$ pour 4 ans de formation aux USA par Bridges et Diamond [1999]). Dans le contexte actuel de recherche de réduction des coûts des structures hospitalo-universitaires, le modèle de compagnonnage décrit plus haut est remis en question, avec comme solution proposée, entre autres, l'augmentation du temps d'apprentissage sur simulateur [Vallancien, 2006], ce qui renvoie à la conclusion du point précédent.

Accélération de l'évolution des techniques chirurgicales : un compagnonnage à vie

La chirurgie et ses techniques évoluent à une vitesse de plus en plus élevée ; la demi-vie d'une information médicale est ainsi évaluée à 5 ans [Gorman et al., 2000], tandis que la durée d'exercice d'un chirurgien augmente pour dépasser actuellement 30 ans. Une formation continue aux nouveaux traitements, aux nouvelles pratiques et aux nouvelles technologies apparaît donc comme indispensable. Parallèlement à ce constat, la disponibilité en temps des chirurgiens pour se former au cours de leur exercice professionnel est limitée, bien que des réglementations existent pour favoriser cette formation continue.

Un dispositif permettant d'apporter des connaissances expertes ou des alertes au chirurgien au cours de la réalisation d'un geste nouveau, dont il maîtrise la théorie mais pour lequel son temps d'exposition en situation réelle est faible, apparaît donc comme un outil parmi d'autres pour limiter l'impact sur le patient de l'apprentissage du chirurgien.

1.1.3 Au delà de l'erreur

En reprenant la définition de la qualité de la chirurgie présentée dans l'introduction de ce chapitre, on s'aperçoit que l'intérêt de l'évaluation de la qualité des gestes en chirurgie ne se limite pas à la détection d'erreurs et à la notation d'habiletés du chirurgien : elle doit aussi permettre aux meilleurs chirurgiens d'améliorer encore davantage la qualité de leurs interventions. En effet, les aléas chirurgicaux ne sont pas toujours des erreurs au sens dans lequel Reason et Hoc [1993] les envisagent, et peuvent survenir en dehors de toute faute technique du chirurgien [Rougé-Maillart et al., 2008, Haertig et al., 2005]. Ainsi, lors d'une chirurgie carcinologique, il peut subsister après l'intervention des cellules tumorales qui font le lit de la récidive (marges positives sur la pièce d'anatomo-pathologie), malgré une excellente maîtrise technique du chirurgien, car cette information fonctionnelle n'est pas directement accessible lors de la dissection.

Il existe donc de nombreuses autres situations dans lesquelles le chirurgien, sans commettre d'erreur, n'est pas optimal car ses capacités cognitives sont dépassées par la quantité ou la complexité des informations dont il dispose et qu'il doit intégrer en temps réel. Un exemple d'une telle situation est le cas où le chirurgien est

assisté d'un interne, et doit à la fois se concentrer sur la réalisation de son geste et contrôler les actions de son assistant. Parfois c'est au contraire l'absence d'une information qui va limiter l'optimisation du geste, telle que la visualisation des bords d'une tumeur en temps réel, pour minimiser le risque d'obtenir des marges carcinologiques positives. Par exemple, au cours d'une prostatectomie, les bords de la tumeur ne sont pas visibles dans l'image endoscopique. En revanche l'apport d'une information de fluorescence [Voros et al., 2013] permet d'identifier ces bords et de réduire le risque de marges positives.

En absence d'erreur manifeste, il existe donc un continuum entre un "bon" et un "mauvais" résultat, qui dépend des compromis effectués par le chirurgien pour traiter l'information dont il dispose. Un domaine entier de l'évaluation de la qualité des gestes s'attache donc à décrire le processus permettant d'identifier, en temps réel, les situations où les choix du chirurgien ne sont pas optimaux. L'objectif est alors d'apporter des "connaissances expertes" au chirurgien, afin de l'aider lors de ces situations à hiérarchiser les informations dont il dispose, et lui apporter des informations complémentaires pouvant participer à l'optimisation de son geste.

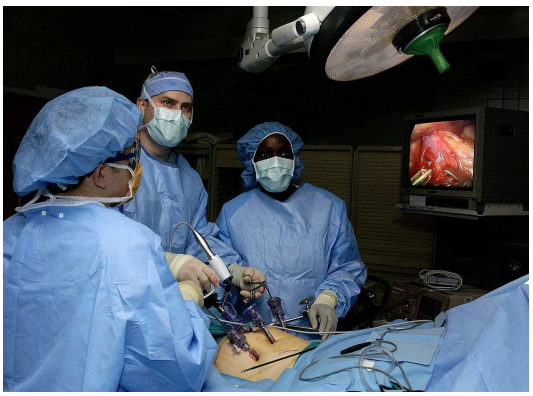
1.2 Spécificités liées à la laparoscopie

Depuis une vingtaine d'années, la chirurgie dite minimalement invasive, et en particulier la laparoscopie, s'est fait une place de choix au sein du bloc opératoire. La laparoscopie consiste à aborder le site de l'intervention à travers de petites incisions pratiquées dans la paroi abdominale et à y introduire des instruments fins permettant de réaliser le geste sous contrôle vidéo (voir fig. 1.1). Cette approche présente de nombreux bénéfices pour le patient (réduction des saignements per-opératoires, réduction des douleurs post-opératoires, durée d'hospitalisation plus courte, bénéfice esthétique,...) mais requiert certaines adaptations de la part du chirurgien :

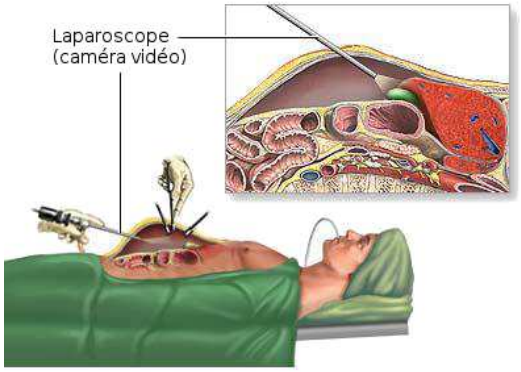
- la coordination oeil-main est fortement altérée par la modification de la localisation de la scène observée (moniteur déporté), par l'inversion de la commande des outils entre la main et l'extrémité effectrice et par l'amplification de l'amplitude des mouvements (pouvoir grossissant de la caméra) [Breedveld et Wentink, 2001] ;
- la perception de la profondeur est diminuée, le champ opératoire devant être reconstruit mentalement par le chirurgien à partir de l'image 2D de la caméra [Breedveld et al., 1999] ;
- le retour haptique permettant au chirurgien de contrôler l'effort qu'il applique aux tissus est réduit par rapport au contact direct entre ses gants et les tissus [Bholat et al., 1999].

Modification de la courbe d'apprentissage

L'apprentissage de ces techniques par les chirurgiens ne s'est pas fait sans incidents, et plusieurs rapports récents ont mis en évidence une sous-évaluation des risques liés à la chirurgie minimalement invasive [Stassen et al., 2010, Gezondheidszorg, 2007]. Les raisons identifiées pour expliquer l'augmentation des erreurs lors de l'appropriation de ces techniques sont l'allongement de la courbe d'apprentissage [Secin et al., 2010, Vickers et al., 2008, Kumar et Gill, 2006] - telle que définie par Feldman et al. [2009] par exemple - et le fait que les compétences requises ne soient pas les mêmes que celles de la chirurgie ouverte [Vickers et al., 2009], imposant à un chirurgien expérimenté en chirurgie ouverte de "ré-apprendre" un geste à partir de nouvelles bases, avec les erreurs associées au parcours du début de la courbe d'apprentissage, lors de ses premières interventions réelles sous endoscopie.



(a) Bloc opératoire de laparoscopie



(b) Illustration du principe de laparoscopie (cholécystectomie)

FIGURE 1.1. – Illustration d'une intervention par laparoscopie

Limites du compagnonnage liées à la laparoscopie

L'apprentissage de la chirurgie se fait habituellement de façon graduelle, avec une participation croissante à la réalisation de gestes au fur et à mesure que l'expérience augmente. Cette approche s'applique difficilement à la laparoscopie, car l'apprenti manipule en général l'endoscope, et ne peut donc pas aisément intervertir son rôle avec le chirurgien senior qui réalise le geste. Cette inversion est difficilement réalisable du fait de l'organisation de l'espace de travail [Villegas et al., 2003]. La gestion de l'endoscope permet de se familiariser avec la représentation mentale de l'espace de travail à partir de la vue 2D, et permet également d'acquérir les notions théoriques concernant le déroulé de la procédure, mais n'autorise pas l'interne à s'entraîner à effectuer les gestes techniques, pourtant différents de ceux utilisés en

chirurgie ouverte. Ce constat a conduit au développement de nombreux simulateurs de laparoscopie, physiques ou virtuels, dont l'apport pour l'apprentissage de la chirurgie laparoscopique a été largement étudié et démontré [Gurusamy et al., 2009]. De même que pour les gestes de chirurgie ouverte étudiés sur simulateurs, il est important de permettre l'évaluation de la qualité des gestes réalisés sur ces simulateurs.

Les solutions apportées jusqu'à présent pour limiter les risques d'une intervention par laparoscopie se sont principalement concentrées sur l'environnement chirurgical, en optimisant l'ergonomie des instruments, et assez peu sur l'évaluation de la qualité du geste. Cette évaluation pourrait cependant permettre d'identifier les habiletés du chirurgien spécifiquement mises en défaut lors de la réalisation de procédures laparoscopiques, afin de focaliser l'apprentissage sur la correction de ces défauts et de réduire la longueur de la courbe d'apprentissage.

Dans le cadre de ce travail, je me suis focalisé sur l'évaluation des aptitudes techniques du chirurgien au cours de la réalisation de gestes en laparoscopie. Cet apport devra ensuite s'intégrer dans un dispositif plus complexe, permettant de dépasser l'évaluation basée sur les erreurs et d'aller vers l'optimisation du geste.

Références

R. Amalberti, X. Cadilhac, H. Garrigue-Guyonnaud, E. Hergon, V. Mounic, ML Pibarot et al. : Principes méthodologiques pour la gestion des risques en établissement de santé. Principes méthodologiques pour la gestion des risques en établissement de santé, 2003. 24

Omar S. Bholat, Randy S. Haluck, Willie B. Murray, Paul J. Gorman et Thomas M. Krummel : Tactile feedback is present during minimally invasive surgery. Journal of the American College of Surgeons, 189 (4):349 – 355, 1999. 28

P. Breedveld, H. G. Stassen, D. W. Meijer et L. P. S. Stassen : Theoretical background and conceptual solution for depth perception and eye-hand coordination problems in laparoscopic surgery. Minimally Invasive Therapy & Allied Technologies, 8(4):227–234, 1999. 28

P. Breedveld et M. Wentink : Eye-hand coordination in laparoscopy-an overview of experiments and supporting aids. Minimally Invasive Therapy and Allied Technologies, 10(3):155–162, 2001. 28

M Bridges et D L Diamond : The financial impact of teaching surgical residents in the operating room. American journal of surgery, 177(1):28–32, janvier 1999. PMID : 10037304. 27

L J Donaldson : The quest for safer surgery. The surgeon : journal of the Royal Colleges of Surgeons of Edinburgh and Ireland, 5(6):324–326, décembre 2007. PMID : 18080604. 24

Peter J. Fabri et José L. Zayas-Castro : Human error, not communication and systems, underlies surgical complications. Surgery, 144(4):557 – 565, 2008. 24

L.S. Feldman, J. Cao, A. Andalib, S. Fraser et G.M. Fried : A method to characterize the learning curve for performance of a fundamental laparoscopic simulator task : Defining. Surgery, 146(2):381–386, 2009. 29

Josef E Fischer : Continuity of care : a casualty of the 80-hour work week. Academic Medicine, 79 (5):381–383, 2004. 26

Inspectie voor de Gezondheidszorg : Risico's kijkoperaties onderschat. http ://www.igz.nl/actueel/nieuws/risicoskijkoperatiesonderschat.aspx, 2007. De patiëntveiligheid bij laparoscopische operaties (kijkoperaties) is onvoldoende gewaarborgd. Dit concludeert de Inspectie voor de Gezondheidszorg (IGZ) in haar rapport 'Risico's minimaal invasieve chirurgie onderschat.' De IGZ ziet reden tot grote zorg. 29

Paul J Gorman, Andreas H Meier, Chantal Rawn et Thomas M Krummel : The future of medical education is no longer blood and guts, it is bits and bytes. The American Journal of Surgery, 180 (5):353–356, novembre 2000. 27

Kurinchi Selvan Gurusamy, Rajesh Aggarwal, Latha Palanivelu et Brian R Davidson : Virtual reality training for surgical trainees in laparoscopic surgery. In Cochrane Database of Systematic Reviews. John Wiley & Sons, Ltd, 2009. 30

A Haertig, S Hansen et F Richard : La responsabilité chirurgicale dans le cadre de la loi d'indemnisation des victimes d'accidents médicaux.(Loi du 4 mars 2002). e-mémoires de l'Académie Nationale de Chirurgie, 4(4):10–14, 2005. 27

J. M. Hamdorf et J. C. Hall : Acquiring surgical skills. British Journal of Surgery, 87(1):28–37, 2000. 25

Udaya Kumar et Inderbir S. Gill : Learning curve in human laparoscopic surgery. Current Urology Reports, 7(2):120–124, avril 2006. 29

Janet M. Corrigan Linda T. Kohn et Editors ; Committee on Quality of Health Care in America Institute of Medicine Molla S. Donaldson : To Err Is Human : Building a Safer Health System. The National Academies Press, 2000. 24

P. Michel, J. L. Quenon, A. Djihoud, S. Tricaud-Vialle, A. M. Sarasqueta, S. Domecq, B. Haury et C. Cases : Les évènements indésirables graves liés aux soins observés dans les établissements de santé : premiers résultats d'une étude nationale. Études et résultats, 398:1–15, 2005. 24

J.L. Ponsky : Addressing the" general competencies" : what is this all about ? Surgery, 135(1):1, 2004. 23

J. Reason et J.M. Hoc : L'erreur humaine. Presses universitaires de France, 1993. 24, 26, 27

Kurt E Roberts, Robert L Bell et Andrew J Duffy : Evolution of surgical skills training. World Journal of Gastroenterology, 12(20):3219, 2006. 26

Clotilde Rougé-Maillart, Nathalie Jousset, Arnaud Gaudin et Michel Penneau : Faut-il reconnaître l'« aléa chirurgical » ? Médecine & Droit, 2008(89):43–47, mars 2008. 27

Sudip K. Sarker, Saif Rehman, Meera Ladwa, Avril Chang et Charles Vincent : A decision-making learning and assessment tool in laparoscopic cholecystectomy. Surgical Endoscopy, 23(1):197–203, janvier 2009. 26

Fernando P. Secin, Caroline Savage, Claude Abbou, Alexandre de La Taille, Laurent Salomon, Jens Rassweiler, Marcel Hruza, Fran\ccois Rozet, Xavier Cathelineau, Gunther Janetschek, Faissal Nassar, Ingolf Turk, Alex J. Vanni, Inderbir S. Gill, Philippe Koenig, Jihad H. Kaouk, Luis Martinez Pineiro, Vito Pansadoro, Paolo Emiliozzi, Anders Bjartell, Thomas Jiborn, Christopher Eden, Andrew J. Richards, Roland Van Velthoven, Jens-Uwe Stolzenburg, Robert Rabenalt, Li-Ming Su, Christian P. Pavlovich, Adam W. Levinson, Karim A. Touijer, Andrew Vickers et Bertrand Guillonneau : The learning curve for laparoscopic radical prostatectomy : An international multicenter study. The Journal of Urology, 184(6):2291–2296, 2010. 29

Laurents P.S. Stassen, Willem A. Bemelman et Jeroen Meijerink : Risks of minimally invasive surgery underestimated : a report of the dutch health care inspectorate. Surgical Endoscopy, 24(3):495–498, 2010. 29

K. Subramonian et G. Muir : The 'learning curve' in surgery : what is it, how do we measure it and can we influence it ? BJU International, 93(9):1173–1174, 2004. 26

Shabnam Undre : Observational assessment of surgical teamwork : a feasibility study. World journal of surgery, 30(10):1774–1783, 2006. 24

Guy Vallancien : L'évaluation de la sécurité, de la qualité et de la continuité des soins chirurgicaux dans les petits hôpitaux publics en france. Rapport technique, Ministère de la Santé et des Solidarités, 2006. 23, 27

Andrew J. Vickers, Fernando J. Bianco, Mithat Gonen, Angel M. Cronin, James A. Eastham, Deborah Schrag, Eric A. Klein, Alwyn M. Reuther, Michael W. Kattan, J. Edson Pontes et Peter T. Scardino : Effects of pathologic stage on the learning curve for radical prostatectomy : Evidence that recurrence in Organ-Confined cancer is largely related to inadequate surgical technique. European Urology, 53(5):960–966, 2008. 29

Andrew J Vickers, Caroline J Savage, Marcel Hruza, Ingolf Tuerk, Philippe Koenig, Luis Martínez-Piñeiro, Gunther Janetschek et Bertrand Guillonneau : The surgical learning curve for laparoscopic radical prostatectomy : a retrospective cohort study. The Lancet Oncology, 10(5):475–480, 2009. 29

L. Villegas, B. E. Schneider, M. P. Callery et D. B. Jones : Laparoscopic skills training. Surgical Endoscopy And Other Interventional Techniques, 17(12):1879–1888, décembre 2003. 29

Charles Vincent, Krishna Moorthy, Sudip K. Sarker, Avril Chang et Ara W. Darzi : Systems approaches to surgical quality and safety : From concept to measurement. Annals of Surgery, 239(4):475, avril 2004. 24

S. Voros, A. Moreau-Gaudry, B. Tamadazte, G. Custillon, R. Heus, M.-P. Montmasson, F. Giroud, O. Gaiffe, C. Pieralli, G. Fiard, J.-A. Long, J.-L. Descotes, C. Vidal, A. Nguyen-Dinh et P. Cinquin : Devices and systems targeted towards augmented robotic radical prostatectomy. IRBM, 2013. 28

CHAPITRE 2

CONTEXTE SCIENTIFIQUE

Sommaire

2.1 Méthodes de description d'une procédure chirurgicale . . . 34
2.1.1 Notion de granularité et modélisation de procédures chirurgicales . . . 34
2.1.2 Reconnaissance de tâches en fonction du contexte . . . 36
2.2 Méthodes de quantification existantes . . . 38
2.2.1 Quelles sont les propriétés d'une méthode d'évaluation idéale ? 39
2.2.2 Systèmes de notation structurée . . . 40
2.2.3 Évaluation à l'aide de statistiques descriptives . . . 47
2.2.4 Modélisations statistiques . . . 56
2.3 Conclusion et problématique de la thèse . . . 58

L'objectif de ce chapitre est de présenter les outils existants dans le domaine de l'évaluation des gestes chirurgicaux, par rapport auxquels je me situe et sur lesquels je m'appuie. Le développement d'une méthode d'évaluation automatique de la qualité des gestes, telle que proposée dans cette thèse, nécessite plusieurs étapes :

1. identification des informations a priori concernant le geste étudié (reconnaissance du temps opératoire à un instant donné, des structures anatomiques d'intérêt) ;
2. détermination de la trajectoire temporelle des instruments chirurgicaux ;
3. évaluation automatique de la qualité du geste par un score basé sur l'analyse de la trajectoire, par comparaison avec des scores déjà validés.

Je décris spécifiquement ici les méthodes d'identification des connaissances a priori nécessaires à mon approche, puis les différentes méthodes existantes pour quantifier la qualité d'un geste, qu'elles soient supervisées ou automatiques.

2.1 Méthodes de description d'une procédure chirurgicale

L'évaluation de la qualité du geste et la détection d'erreurs au cours de la réalisation d'une procédure chirurgicale peuvent s'effectuer à différents niveaux d'analyse, et nécessitent d'être capable d'identifier la tâche réalisée à un instant donné par le chirurgien parmi le flux de tâches réalisées au cours de l'intervention.

2.1.1 Notion de granularité et modélisation de procédures chirurgicales

L'analyse scientifique du flux des tâches effectuées par le chirurgien, ou *surgical workflow* dans la littérature [Jannin et Morandi, 2007], a reçu depuis une dizaine d'années une attention accrue de la part de la communauté scientifique, avec deux objectifs principaux :

- constituer une base de données permettant d'évaluer et de développer des dispositifs d'aide à la chirurgie,
- optimiser le système d'informations disponibles pour le personnel du bloc opératoire afin de créer la salle d'opération du futur [Cleary et al., 2005].

i) Objectifs

L'objectif théorique de la formalisation du processus chirurgical à travers une modélisation est d'établir une connaissance du flux des actions, afin de réaliser plusieurs objectifs concrets : améliorer la connaissance de l'intervention, évaluer l'apport de nouveaux outils pour la visualisation et la manipulation, évaluer les compétences des chirurgiens dans les différentes étapes d'une procédure, développer des nouveaux dispositifs d'entraînement basés sur la décomposition de la procédure chirurgicale, améliorer la planification de l'intervention et apporter les informations pertinentes au bon moment au chirurgien pour l'aider à réaliser sa tâche.

ii) Formalisme et définitions

Cette analyse du processus chirurgical se décompose en deux dimensions : une décomposition hiérarchique verticale de l'intervention en une succession de "temps", dont la précision définit le niveau de granularité de l'observation, et une classification horizontale entre les différentes composantes de chaque "temps".

La granularité est définie par Trelhu [2011] comme la taille du plus petit élément d'un système, et définit la précision de la description du système. Plus la granularité est élevée, plus l'information obtenue sera précise, incluant par exemple le geste réalisé, l'outil utilisé, l'organe cible, la main effectrice... Le niveau de granularité souhaité est fonction des applications visées, et conditionne la méthodologie d'enregistrement du *workflow*.

Les classifications horizontales sont décrites par Neumuth et al. [2006] dans cinq domaines complémentaires : organisation (qui effectue la tâche ?), fonction (l'unité

constituée de l'activité et de la structure sur laquelle elle s'applique), information (données requises ou produites), opération (outils et instruments chirurgicaux utilisée pour la tâche) et comportement (temporel : enchaînement des tâches, logique : tâches synchrones et boucles, et causal : rapport de causalité entre tâches).

Cette méthodologie rigoureuse permet de modéliser la procédure chirurgicale dans un référentiel commun à plusieurs équipes, indépendamment du lieu ou du type d'intervention envisagé, grâce à un formalisme partagé par tous. Ceci afin de créer des modèles de procédures chirurgicales génériques (gSPM pour *generic surgical process model*) pouvant servir de bases de travail à l'optimisation de la réalisation du geste chirurgical, pouvant être enrichis au cours du temps avec de nouvelles connaissances sous forme d'ontologies, et tolérant une variabilité dans la séquence des étapes chirurgicales en fonction des différentes techniques chirurgicales.

Les premières modélisations de procédures chirurgicales consistant en plusieurs niveaux de granularité ont été proposées par McKenzie [Cao et al., 1996, MacKenzie et al., 2001], définissant des étapes, sous-étapes, tâches, sous-tâches et activités pour 3 différentes procédures chirurgicales. Cette approche ne présentait en revanche pas de classification horizontale autre que la succession temporelle des tâches, qui aurait permis de rendre la description plus riche et plus adaptable à d'autres chirurgies.

De nombreuses procédures chirurgicales ont depuis été décrites sous la forme de SPM, on se rapportera au travail de Lalys [2012] pour avoir une description exhaustive des travaux dans ce domaine.

Ce dernier identifie cinq caractères nécessaires à la définition d'un SPM :

- **le niveau de granularité de l'observation**, qui permet de structurer la procédure en une succession de mouvements (description d'une trajectoire, sans signification associée), qui définissent une activité (une action ayant un sens dans la progression de la chirurgie, définie par le triplet {*outil, cible, action*}), dont l'enchaînement définit à son tour une étape de la procédure ;
- **le formalisme utilisé par la modélisation**, allant du plus complexe (ontologie, c'est-à-dire une représentation explicite d'un concept permettant un traitement rationnel) au plus simple (liste d'actions non ordonnées) ;
- **la méthode d'acquisition des données**, qui dépend du niveau de granularité souhaité (décomposition de la procédure dans son ensemble, de certaines étapes, de certaines activités ou de mouvements) et de l'opérateur observé (chirurgien, anesthésiste, infirmière, ou les trois). L'acquisition peut être réalisée en temps réel ou en post-opératoire à partir de vidéos ou de données de suivi d'instruments, et selon deux méthodes principales : acquisition par un observateur ou acquisition par des capteurs (donc automatique) ;
- **les méthodes d'analyse des données collectées** : elles dépendent de la méthode choisie pour intégrer les données collectées à un modèle en complétant les informations manquantes entre les deux. Cette intégration peut se

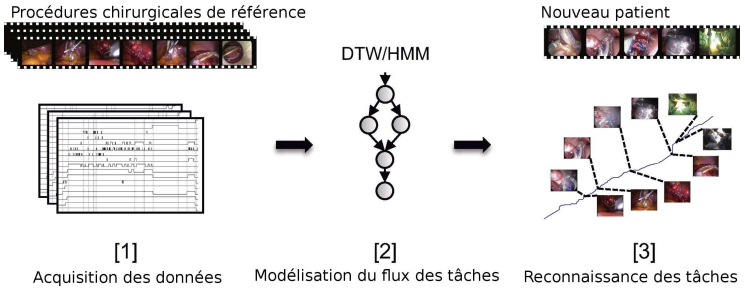


FIGURE 2.1. – Exemple de système de reconnaissance de tâches, basé sur les instruments utilisés à un instant donné, extrait de [Padoy et al., 2012]

faire selon une approche ascendante *bottom-up*), en partant de données de bas-niveau (position des instruments par exemple) pour aller vers des informations de haut-niveau telles que l'étape courante, ou selon une approche descendante *top-down*), en partant de données de haut-niveau (informations spécifiques au patient par exemple) pour aller vers des informations de bas-niveau, telles des détails concernant les gestes effectués. On peut alors aggréger plusieurs SPM pour générer des modélisations généraiques (gSPM), ou encore comparer plusieurs processus entre eux pour souligner leurs différences. Enfin on peut vouloir simplement représenter graphiquement le déroulé (temporel ou logique) d'une procédure afin d'en faciliter la compréhension ;

- **le domaine d'application du modèle** (quel type de chirurgie ?) et son objectif (voir 2.1.1i)).

2.1.2 Reconnaissance de tâches en fonction du contexte

Dans le contexte de cette thèse, je fais l'hypothèse qu'il est possible, à partir de travaux décrits dans la littérature et que je présente ici, de développer un système permettant d'identifier précisément l'activité effectuée à un instant donné par le chirurgien, et de la situer dans l'enchaînement d'activités dans la procédure, grâce aux SPM. La figure 2.1 illustre ce principe. Pour remplir cet objectif, le formalisme des SPM étudiés doit donc être au minimum une liste ordonnée d'activités, et l'acquisition de l'activité courante doit pouvoir être effectuée en temps réel de façon automatique. L'analyse des données acquises permet alors de se situer dans le déroulement d'un processus modèle, afin de permettre l'évaluation de la qualité du geste au sein d'un système contextualisé.

Dans ce cadre, le groupe CAMP de Munich propose plusieurs pistes, basées sur la connaissance des instruments utilisés à un instant donné de la procédure. Ahmadi et al. [2006] proposent de recaler temporellement les différentes tâches d'une procédure chirurgicale en appliquant un algorithme de *dynamic time warping* [Sakoe et Chiba, 1971] à des vecteurs d'état multidimensionnels correspondant à la présence

ou non d'un instrument dans la cavité abdominale, assortis d'informations visuelles (vue endoscopique, présence de clips chirurgicaux). Cette méthode est appliquée à une cholécystectomie et permet de recaler les étapes d'une procédure avec une précision proche de la seconde. Cette méthode est cependant limitée par les variations possibles d'une procédure à l'autre (répétition d'une étape imprévue par exemple). Padoy et al. [2007, 2008] et Blum et al. [2008], en utilisant le même type de données, construisent leur modèle à partir de modèles de Markov cachés (HMM) [Rabiner, 1989]. Padoy et al. [2012] combinent les deux approches de *dynamic time warping* et de modèles de Markov cachés. La limite de ces travaux reste l'annotation manuelle de la présence ou non des instruments dans la cavité abdominale.

Une piste pour résoudre cette difficulté est explorée par Neumuth et al. [2009], qui proposent une assistance pour la création de modèles de procédure chirurgicale (pour une chirurgie fonctionnelle endoscopique des sinus) en détectant automatiquement l'utilisation d'un instrument grâce à des capteurs électroniques, puis en proposant à l'observateur de construire son modèle en s'aidant des concepts liés à l'usage de ces instruments, puisés dans des ontologies existantes.

Ko et al. [2007] proposent une approche bottom-up pour asservir l'orientation d'une caméra à l'étape active d'une chirurgie, en analysant également les instruments utilisés à l'aide d'un diagramme de changements d'états. Ici aussi l'annotation des transitions d'instruments est manuelle.

Le groupe de Yang propose deux approches distinctes. James et al. [2007] utilisent des réseaux de neurones pour reconnaître différentes étapes d'un modèle de cholécystectomie par laparoscopie chez le porc, grâce à une approche bottom-up intégrant comme données des primitives visuelles du flux vidéo et la direction du regard du chirurgien, avec un taux de reconnaissance de 75%. Lo et al. [2003] sont capables de reconnaître 4 phases distinctes (attente, rétraction, cautérisation, suture) au cours d'une chirurgie, sans les ordonner, en combinant la position 2D des instruments dans l'image, des informations sur la déformation des tissus, sur les reflets spéculaires et sur la présence du fil de suture, dans une approche bayésienne.

Lin et al. [2006], en partant des données de cinématique du robot chirurgical DaVinci (Intuitive Surgical, Inc), parviennent à identifier plusieurs activités au cours d'une tâche de suture grâce à une approche bottom-up basée sur un classifieur bayésien, en réduisant la dimensionnalité de son jeu de données par une analyse discriminante linéaire.

Enfin le groupe de Jannin a proposé plusieurs approches pour identifier automatiquement les tâches au cours d'une procédure chirurgicale. On citera entre autres Lalys et al. [2011a] qui extraient des informations du flux vidéo d'une chirurgie hypophysaire à l'aide de séparateurs à vaste marge, puis entraînent un modèle de Markov caché discret pour identifier les différentes phases de la procédure afin de pouvoir générer automatiquement des comptes-rendus. Dans [Lalys et al., 2011b, Lalys, 2012, Lalys et al., 2013, Bouget et al., 2012], une méthode permettant d'identifier les tâches et les activités (donc intégrant deux niveaux de granularité distincts)

est présentée, basée sur l'analyse de vidéos de chirurgie de la cataracte, et combine un séparateur à vaste marge à un algorithme de *dynamic time warping*.

Ces travaux montrent que des méthodes permettant d'identifier la phase et l'activité d'une procédure chirurgicale à un instant donné existent, qu'elles autorisent une certaine variabilité dans le déroulement de la procédure, et qu'elles peuvent être améliorées en combinant différents types d'informations (information visuelle, cinématique des instruments, type d'instrument utilisé,...).

2.2 Méthodes de quantification existantes

Les méthodes de quantification existantes pour mesurer la qualité technique d'un geste chirurgical peuvent se répartir en trois grands domaines :

- les systèmes de notation structurée ("*structured grading*")
- les statistiques descriptives
- les modélisations statistiques

Je ne traite pas ici des méthodes d'évaluation de type *In-Training Evaluation Reports* et *log-books* car il a été démontré que leurs résultats n'étaient pas fiables [Turnbull et al., 1998] ou ne mesuraient pas effectivement la qualité de la chirurgie (en particulier les *log-books* à cause de l'absence de retour sur la qualité : *"Experience can mean doing the wrong thing over and over again", Charles Mayo*).

▷ **Plan d'analyse des différentes méthodes :** Pour chaque type d'évaluation étudié dans cette partie, je m'attacherai à préciser autant que possible ses caractéristiques parmi les domaines suivants :

- type de chirurgie : la méthode étudiée s'intéresse-t-elle à la chirurgie ouverte, à la chirurgie minimallement invasive du type laparoscopie ou encore à la chirurgie robotisée ?
- domaine de chirurgie : s'agit-il de chirurgie digestive, d'urologie,de gynécologie ou encore de neurochirurgie ou d'orthopédie ? Certaines méthodes étudiées sont génériques par rapport à ce domaine et peuvent s'appliquer indifféremment à plusieurs types de chirurgie, tandis que d'autres sont conçues spécifiquement pour une discipline donnée.
- contexte d'utilisation : la méthode peut-elle s'appliquer au bloc opératoire, est-elle conçue pour un usage sur banc d'entraînement ou pour un simulateur de réalité virtuelle ?
- niveau de détail exploré : à quel niveau de granularité (procédure, tâche, activité,...) l'observation de la qualité du geste est-elle effectuée ? La méthode

est-elle alors spécifique à certaines tâches ou peut-elle être généralisée à la réalisation d'autres tâches ?

- méthode d'acquisition des données : il s'agit d'un point crucial pour pouvoir espérer une implémentation large des méthodes d'évaluation au sein de la pratique chirurgicale : l'acquisition dépend-elle d'un observateur externe, ou bien utilise-t-elle des techniques de suivi de trajectoire par marqueurs (magnétiques, visuels, mécaniques) ou par analyse d'image, ou encore utilise-t-elle des données de cinématiques obtenus par un simulateur ou un robot ?
- paramètres étudiés

2.2.1 Quelles sont les propriétés d'une méthode d'évaluation idéale ?

Les propriétés recherchées des méthodes d'évaluation des compétences psychomotrices en chirurgie s'articulent autour de cinq axes principaux :

▷ **La fiabilité** La fiabilité d'une méthode de mesure est principalement dépendante de la reproductibilité de ses résultats, qui peut être mesurée sur une échelle de 0 à 1. Le niveau visé de reproductibilité est dépendant de l'application cible : on considère dans le domaine de l'évaluation des compétences chirurgicales qu'une reproductibilité ≥ 0.80 est nécessaire si l'objectif est de valider des compétences (on peut se contenter de valeurs plus faibles si l'objectif est d'accompagner une progression) [Wanzel et al., 2002]. On distingue la reproductibilité intra-observateur (les résultats sont-ils identiques si le test est répété plusieurs fois avec le même individu sans modification de ses compétences ?), qui peut être mesurée par le coefficient de corrélation interne α de Cronbach, et la reproductibilité inter-observateurs mesurée par un score de corrélation classique (Pearson ou Spearman) ou par un coefficient de corrélation intra-classe.

▷ **La validité** La validité teste si la méthode mesure effectivement ce que l'on souhaite qu'elle mesure. On peut décomposer la validité selon plusieurs dimensions :

validité concurrente : à quel degré les résultats de la méthode testée sont-ils corrélés à ceux d'une autre méthode considérée comme *gold-standard* pour le domaine exploré ?

validité de construction : la méthode mesure-t-elle la caractéristique (non observable) qu'elle est censée mesurer ? Par exemple dans le cas de l'évaluation de compétences techniques on cherchera à s'assurer que les meilleurs techniciens obtiennent de meilleurs résultats que les moins bons. La définition du bon technicien doit alors être obtenue par d'autres méthodes que le test étudié (expérience, notoriété,...) avec l'existence possible de biais.

validité de contenu : la méthode d'évaluation est-elle exacte quant au contenu qu'elle propose ? On vérifie si les questions reflètent le concept à mesurer

de manière représentative. La plupart du temps, cette forme de validité (subjective) est évaluée lors de l'élaboration de la méthode par un panel d'experts.

validité apparente : le contexte dans lequel est effectuée la mesure représente-t-il de manière réaliste la situation réelle que l'on souhaite modéliser ? Il s'agit pour les évaluations sur banc d'essai ou sur simulateur de réalité virtuelle de vérifier qu'ils miment effectivement les situations rencontrées au bloc opératoire.

▷ **La faisabilité** La méthode d'évaluation de la qualité du geste chirurgical idéale du point de vue de la faisabilité doit pouvoir être aisément portée vers différents contextes d'utilisation (du laboratoire vers le bloc opératoire et inversement), ainsi que vers d'autres lieux géographiques [Wanzel et al., 2002]. La problématique du coût est également importante dans le contexte de la santé à l'heure actuelle, et doit être prise en compte lors de l'élaboration des méthodes d'évaluation [Gallagher et al., 1977, Cusimano et al., 1994].

▷ **L'exhaustivité** Cette propriété est complexe à atteindre, et est généralement antinomique avec la propriété précédente. La méthode d'évaluation idéale vis-à-vis de cette propriété doit explorer toutes les facettes du domaine étudié avec un niveau de détail fin (granularité importante), et mesurer la performance du chirurgien dans chacune de ces facettes.

▷ **La flexibilité** La flexibilité se rapproche de la faisabilité : la méthode d'évaluation doit pouvoir s'adapter à différents types d'interventions, qu'elles soient simples ou complexes, et à différents domaines chirurgicaux.

Le tableau 2.4 résume pour chaque méthode décrite ci-dessous ces propriétés.

2.2.2 Systèmes de notation structurée

Ces dispositifs sont destinés à évaluer la performance technique de l'opérateur de la chirurgie en limitant les biais de subjectivité de l'observateur et en améliorant la reproductibilité inter- et intra-observateur. La notation se réalise à l'aide soit de *check-lists*, soit à l'aide de scores globaux évaluant différentes composantes de l'intervention, soit à l'aide des deux. Une revue de la littérature a été réalisée par Ahmed et al. [2011] en reprenant pour chaque méthode la validité, la faisabilité et la reproductibilité. Je décris ici plus précisément trois méthodes, qui sont soit les plus utilisées dans le cadre de simulateurs chirurgicaux ou au bloc opératoire, soit d'un intérêt particulier dans le cadre de cette thèse.

i) Objective Structured Assessment of Technical Skills (OSATS)

Ce score proposé par l'équipe de Reznick [Faulkner et al., 1996, Martin et al., 1997, Reznick et al., 1997] s'inspire des méthodes développées pour évaluer la pra-

tique clinique des médecins de façon plus générale (*Objective Strucutured Clinical Evaluations* [van der Vleuten et Swanson, 1990]). Il s'articule autour de *check-lists* spécifiques à chaque tâche effectuée au cours d'une procédure chirurgicale simulée sur banc d'essai ou sur modèle animal, ainsi que sur des scores globaux pour l'ensemble de la procédure. Chaque *check-list* permet de contrôler la réalisation de plusieurs points critiques au sein de la tâche évaluée (tableau 2.1a par exemple pour une plaie de veine cave inférieure). Le score global s'articule autour de la notation de 7 items sur une échelle de Likert (tableau 2.1b), avec des notes réparties de 1 à 5, avec des libellés descriptifs en 1, 3 et 5, en s'inspirant des travaux précurseurs de Kopta [1971] dont la validité n'avait jamais été démontrée. Un exemple de *check-list* et la grille de notation du score global sont donnés dans le tableau 2.1.

▷ **Type et domaine de chirurgie** Ce score a été initialement conçu pour la chirurgie ouverte, mais de nombreuses variantes des exercices proposés ont été développées pour permettre d'évaluer également la qualité des gestes réalisés en laparoscopie ou en chirurgie robotisée [Goff et al., 2002, Reznick et MacRae, 2006]. La structure du score permet de le porter aisément vers différentes spécialités chirurgicales : gynécologie [Goff et al., 2002, Swift et Carter, 2006], chirurgie vasculaire [Beard, 2005, Beard et al., 2007, Pandey et al., 2006] et cardiaque [Hance et al., 2005], urologie [Herati et al., 2012], orthopédie [Leong et al., 2008],... Ces échelles ont été initialement pensées pour l'évaluation des internes, mais leur validité a été secondairement démontrée pour des niveaux de pratique supérieurs [Aggarwal et al., 2007, Hance et al., 2005, Swift et Carter, 2006]. Munz et al. [2004] montrent cependant un effet de seuil concernant le niveau de pratique, le système de notation ne parvenant pas à discriminer les chirurgiens seniors entre eux.

▷ **Contexte d'utilisation** Le contexte d'utilisation initial de ce score était la simulation de procédures chirurgicales sur banc d'essai (inanimé ou modèle animal), mais plusieurs travaux ont montré qu'il pouvait être utilisé dans le contexte du bloc opératoire [Hiemstra et al., 2011, Aggarwal et al., 2008].

▷ **Niveau de détail exploré** L'analyse de la qualité du geste à l'aide d'une grille de notation type OSATS s'effectue à deux niveaux de granularité distincts : au niveau de la procédure dans son ensemble à l'aide de l'échelle de Likert globale, et au niveau des différentes tâches à réaliser pour compléter la procédure à l'aide des *check-lists* spécifiques à chacune de ces tâche. Les deux échelles sont équivalentes dans la mesure où Martin et al. [1997] montrent qu'elles classent les internes en fonction de leur niveau de pratique de la même façon (validité de construction) sur modèle animal. Leur interprétation peut donc identifier des lacunes concernant spécifiquement une tâche (prise en charge non optimale d'une plaie de veine cave inférieure par exemple, en oubliant de contrôler que l'ensemble des instruments nécessaires à la réalisation du geste sont disponibles) mais peut également identifier des lacunes plus générales

ITEM	Not done/ Done incorrectly	Done correctly
CONTROL OF HEMORRHAGE		
1. Applies pressure to stop bleeding first	0	1
2. Asks assistant for suction field	0	1
3. Inspects injury by carefully releasing the IVC	0	1
4. Ensures all equipment needed for repair is at hand before starting	0	1
5. Control of bleeding point (use deBakey forceps/Satinsky clamp or prox/distal pressure	0	1
REPAIR		
6. Select appropriate suture (4.0/5.0/6.0 polyproylene)	0	1
7. Select appropriate needle driver (vascular)	0	1
8. Select appropriate forceps (deBakey)	0	1
9. Needle loaded 1/3-2/3 from tip 90% of time	0	1

(a) Extrait de la liste de vérification pour l'hémostase d'une plaie de veine cave inférieure (IVC) dans l'OSATS. Chaque item listé reçoit une note (1 ou 0) en fonction de sa réalisation correcte ou non.

	1	2	3	4	5
Respect for tissue	Frequently used unnecessary force on tissue or caused damage by inappropriate use of instruments		Careful handling of tissue but occasionally caused inadvertent damage		Consistently handled tissues appropriately with minimal damage
Time and motion	Many unnecessary moves		Efficient time/motion but some unnecessary moves		Economy of movement and maximum efficiency
Instrument handling	Repeatedly makes tentative or awkward moves with instruments		Competent use of instruments although occasionnally appeared stiff or awkward		Fluid moves with instruments and no awkwardness
Knowledge of instruments	Frequently asked for the wrong instrument or used an inapprpriate instrument		Knew the names of most instruments and used appropriate instrument for the task		Obviously familiar with the instruments required and their names
Use of assistants	Consistently placed assistants poorly or failed to use assistants		Good use of assistants most of the time		Strategically used assistant to the best advantage at all times
Flow of operation and forward planning	Frequently stopped operation or needed to discuss next move		Demonstrated ability for forward planning with steady progreesion of operative procedure		Obviously planned course of operation with effortless flow from one move to the next
Knowledge of specific procedure	Deficient knowledge. Needed specific instruction at most operative steps		Knew all important aspects of the operation		Demonstrated familiarity with all aspects of the operation

(b) Échelle d'évaluation globale de l'OSATS. Il s'agit d'une échelle de Likert notée de 1 à 5, les ancres en 1, 3 et 5 permettent à l'évaluateur d'avoir des decriptions de la qualité attendue pour ces notes.

Tab. 2.1. – Exemples d'échelles d'évaluation de l'OSATS (extrait de Martin et al. [1997]).

(mauvaise communication avec le reste de l'équipe chirurgicale). L'exploration des deux niveaux de granularité rend donc ces deux composantes du système de notation complémentaires.

▷ **Méthode d'acquisition des données** Cette méthode d'évaluation nécessite la présence d'un observateur externe pour analyser la réalisation de la procédure de la façon la plus objective possible. En effet, l'auto-évaluation de la performance n'a pas montré de résultats satisfaisants [Pandey et al., 2008, Sidhu et al., 2006], sauf pour les populations de chirurgiens les plus expérimentés dans une étude [Moorthy et al., 2006]. En revanche, l'évaluation ne doit pas nécessairement être effectuée en temps réel, plusieurs travaux ayant montré que l'analyse d'enregistrements vidéo de la procédure respectaient les critères de validité de construction [Herati et al., 2012, Beard et al., 2005, Dath et al., 2004, Aggarwal et al., 2008, Driscoll et al., 2008]. Scott et al. [2000] montrent cependant que l'édition des enregistrements vidéo afin de ne conserver que les phases considérées comme critiques pour la procédure altéraient les scores donnés par les examinateurs et ne permettaient pas de démontrer une validité de construction. Des versions modifiées de cette échelle permettent également d'évaluer la qualité de la chirurgie à partir de ses résultats "physiques", c'est à dire l'analyse des sutures, anastomoses ou résections réalisées sur les bancs d'essais [Szalay et al., 2000], permettant de différer le moment de l'évaluation.

▷ **Paramètres étudiés** Les paramètres étudiés sont très variables en ce qui concerne les *check-lists*, celles-ci étant adaptées à chaque procédure analysée. En revanche, le score de notation global n'est pas modifié d'une procédure à l'autre, ni même d'un domaine chirurgical à l'autre. Les différentes composantes du geste chirurgical analysé sont décrites dans le tableau 2.1b : respect des tissus, vitesse et fluidité d'exécution, manipulation des instruments, connaissance des instruments, interaction avec le reste de l'équipe chirurgicale, planification de la suite de l'intervention, connaissance de la procédure. Ces paramètres sont volontairement génériques, et ont participé au déploiement large de cette méthode d'évaluation dans les programmes d'apprentissage des internes dans les pays anglo-saxons.

ii) Global Operative Assessment of Laparoscopic Skills (GOALS)

Cette méthode d'évaluation a été proposée par Vassiliou et al. [2005], dans le but de concevoir une échelle d'évaluation spécifique aux interventions chirurgicales par laparoscopie, et pensée directement pour une utilisation au bloc opératoire. Cette méthode utilise également un système de notation évaluant 5 items sur une échelle de Likert allant de 1 à 5, avec des ancres en 1, 3 et 5 (donc assez proche de l'échelle de notation globale de la méthode précédente). Les 5 composantes du geste analysées sont la dextérité bimanuelle, la perception de la profondeur, l'efficience, la préhension des tissus et l'autonomie. Cette échelle est présentée dans le tableau 2.2.

	1	2	3	4	5
Depth perception	Consistently overshoots target, wide swings, slow to correct		Some overshooting or missing of target, but quick to correct		Accurately directs instruments in the correct plane to target
Bimanual dexterity	Uses only one hand, ignores nondominant hand, poor coordination between hands		Uses both hands, but does not optimize interaction between hands		Expertly uses both hands in a complementary manner to provide optimal exposure
Efficiency	Uncertain, inefficient efforts ; many tentative movements ; constantly changing focus or persisting without progress		Slow, but planned movements are reasonably organized		Confident, efficient and safe conduct, maintains focus on task until it is better performed by way of an alternative approach
Tissue handling	Rough movements, tear tissue, injuries adjacent structures, poor grasper control, grasper frequently slips		Handles tissues reasonably well, minor trauma to adjacent tissue (ie. occasionnal unnecessary bleeding or slipping of the grasper		Handles tissues well, applies appropriate traction, negligeable injury to adjacent structures
Autonomy	Unable to complete entire task, even with verbal guidance		Able to complete task safely with moderate guidance		Able to complete task independently without prompting

Tab. 2.2. – Échelle d'évaluation GOALS (extrait de Vassiliou et al. [2005]).

▷ **Type et domaine de chirurgie** Cette méthode d'évaluation a été développée spécifiquement pour la chirurgie minimalement invasive, en particulier pour la laparoscopie. Elle a été évaluée pour différentes procédures de chirurgie digestive (cholécystectomie [Vassiliou et al., 2005], appendicectomie [Gumbs et al., 2007], colectomie [Sidhu et al., 2006], réparation de hernie inguinale [Kurashima et al., 2011]) avec de meilleures performances en terme de validité de construction que les *check-lists* auxquelles elle était comparée.

▷ **Contexte d'utilisation** Ce score a été conçu pour être utilisé au bloc opératoire, mais sa proximité avec le score global de la méthode OSATS amène à penser qu'il pourrait être utilisé lors d'entraînements sur bancs d'essai.

▷ **Niveau de détail exploré** Ce score, par son aspect générique, ne permet pas d'explorer en détail chaque tâche effectuée au cours d'une procédure, mais évalue la procédure dans son ensemble. Il reste cependant valide pour analyser des phases dans la procédure (Vassiliou et al. [2005] analysent uniquement la phase de dissection du plan postérieur de la vésicule biliaire) pour atteindre un niveau de granularité plus fin, mais nécessite alors de répéter la mesure pour chaque phase si on veut avoir une évaluation de l'ensemble de la procédure, ce qui peut être plus contraignant pour l'examinateur.

▷ **Méthode d'acquisition des données** Ici encore, l'acquisition des données est effectuée par un observateur externe ; Vassiliou et al. [2005] montrent que la reproductibilité inter-observateur est bonne si les observateurs sont expérimentés en

chirurgie (*observateur-expert*). Afin de rendre l'analyse des gestes moins contraignante, Chang et al. [2007] ont montré pour cette méthode que l'enregistrement vidéo de la procédure puis la lecture différée cet enregistrement n'entravait pas la validité de construction du score. Leur étude souffre cependant d'un important défaut de puissance, seules deux procédures ayant été comparées (une réalisée par expert, l'autre par un novice). Choy et al. [2013] montrent que l'analyse *in-situ* et l'analyse différée par enregistrement vidéo sont équivalentes, et que le taux de corrélation entre les deux augmente avec la bande-passante de la vidéo, soulignant le point critique qu'est la qualité des données. Cette faisabilité de différer l'analyse des vidéos présente un avantage supplémentaire : la possibilité d'anonymiser les vidéos afin de limiter le biais de subjectivité lors de leur analyse.

▷ **Paramètres étudiés** Les paramètres étudiés ont été décrits plus haut, ils sont génériques et peuvent s'appliquer pour tous les types de chirurgie laparoscopique, et permettent d'analyser la qualité du geste de façon qualitative en soulignant les domaines psychomoteurs dans lesquels le chirurgien n'est pas optimal.

L'intérêt de ce score pour ma thèse est son aspect générique et sa validité pour évaluer les compétences techniques du chirurgien au bloc opératoire, ce qui en fait un outil utile pour évaluer la validité des scores que j'envisage de construire.

iii) Observational Clinical Human Reliability Assessment (OCHRA)

Cette approche proposée par l'équipe de Cuschieri [Joice et al., 1998, Tang et al., 2006] est intéressante dans le cadre de cette thèse, car elle combine une décomposition (manuelle) de la procédure chirurgicale à une analyse des erreurs susceptibles de se produire au cours de chaque tâche, selon une méthodologie inspirée de l'aéronautique. Les étapes successives de la construction de l'échelle sont décrites en détail dans la revue de littérature de [Cuschieri et Tang, 2010] :

- décomposition temporelle ordonnée de la procédure en tâches, sous-tâches et étapes par un panel de chirurgiens experts, avec définition d'un enchaînement considéré comme idéal pour une réalisation optimale de la procédure
- identification par le panel d'experts des tâches critiques qui déterminent l'issue de la procédure
- description des erreurs, ainsi que de leur cause, leur mode (erreur dans la procédure/dans l'exécution), leur portée (avec conséquence/sans conséquence).

▷ **Type et domaine de chirurgie** Cette méthode a été initialement conçue pour analyser le déroulement d'une cholécystectomie par voie laparoscopique, mais a été adaptée pour également analyser le déroulement d'une pyloromyotomie [Tang et al., 2004] ou d'une chirurgie de la cataracte [Gauba et al., 2008]. L'adaptation n'est pas aussi rapide que pour le score GOALS par exemple (en raison de la nécessaire dé-

composition spécifique de la procédure), ce qui limite la disponibilité de la méthode pour un usage courant.

▷ **Contexte d'utilisation** Toutes les études citées précédemment ont évalué la méthode au bloc opératoire.

▷ **Niveau de détail exploré** Le niveau de granularité atteint par cette méthode est celui de l'étape chirurgicale [Joice et al., 1998], mais l'analyse de l'enchaînement des étapes, puis des sous-tâches permet d'évaluer également la connaissance de la procédure dans son ensemble.

▷ **Méthode d'acquisition des données** Les procédures sont enregistrées au bloc opératoire, puis analysées secondairement en aveugle par un observateur externe dans toutes les études citées ci-dessus. Le niveau de granularité exploré et la complexité de l'échelle (analyse de différents types d'erreur simultanément) rend l'analyse longue et exigeante, mais reproductible et objective (observateur aveugle, critères très détaillés).

▷ **Paramètres étudiés** Cette méthode analyse à la fois des compétences techniques (force appliquée aux tissus, déplacement/orientation des instruments,...) et des compétences cognitives (connaissance du déroulé de la procédure, positionnement des instruments,...).

L'intérêt principal de cette méthodologie dans le cadre de cette thèse est qu'elle pose les bases théoriques sur lesquelles je souhaite construire ma méthode d'évaluation de la qualité du geste en chirurgie. Ces bases sont d'une part l'identification des erreurs portant à conséquence se produisant dans le déroulé de la procédure, qu'elles soient techniques ou procédurales, d'autre part l'analyse de conduites non-optimales par rapport à un déroulé "idéal", qui même si elles ne sont pas préjudiciables pour le résultat final de la procédure rendent celle-ci sub-optimale. Cependant, cette méthode ne peut pas être utilisée en pratique courante en raison des coûts importants lors de sa construction (décomposition manuelle de la procédure) et de sa mise en œuvre (ressources humaines non disponibles pour évaluer les enregistrements).

De plus, cette méthode présente l'inconvénient de proposer une analyse différée, ce qui ne permet pas au chirurgien de rectifier son action afin de s'approcher de la procédure optimale. Enfin, l'analyse par un observateur humain limite la construction de modèles de procédures complexes, dans lesquels plusieurs scenarii peuvent être équivalents.

Les SPMs présentés dans la section 2.1 apparaissent alors comme un support indispensable à l'utilisation de ce type de score, aussi bien à la phase de construction du modèle (pour décomposer la procédure en tâches, sous-tâches et étapes s'articulant selon un ou plusieurs scenarii optimaux) qu'au moment de l'analyse (pour

extraire le déroulé de la procédure étudiée et identifier les erreurs de procédure).

L'ensemble des méthodes d'évaluation proposées ci-dessus présente la même limite, rendant leur implémentation en pratique courante inenvisageable, à savoir le coût associé à une évaluation par un observateur externe expert. De nombreuses équipes se sont donc orientées vers la recherche de grandeurs qui seraient corrélées à l'évaluation par un observateur expert, mais dont l'acquisition pourrait être automatisée. Ces grandeurs pourraient ensuite servir de base à des classifieurs basés sur l'analyse statistique de leur distribution parmi les populations de différents niveaux de chirurgie.

2.2.3 Évaluation à l'aide de statistiques descriptives

L'analyse de la qualité du geste basée sur des statistiques descriptives permet, à partir de variables quantitatives, de classer le geste ou le chirurgien l'effectuant par rapport à des standards issus de répétitions antérieures du même geste. Les variables quantitatives analysées peuvent être multiples, et peuvent être acquises par un observateur (non-nécessairement expert car il s'agit de grandeurs mesurables objectives) ou par des systèmes automatisés.

On peut distinguer deux grandes catégories dans ces méthodes d'évaluation : l'analyse du résultat et l'analyse du mouvement.

i) Analyse du résultat

La principale méthode d'analyse de la qualité du geste à partir de l'étude d'un résultat est le test *Fundamentals of Laparoscopic Skills* (FLS) qui est basé sur le simulateur *McGill Inanimate System for Training and Evaluation of Laparoscopic Skills* (MISTELS) mis au point par le groupe de Fried [Derossis et al., 1998]. MISTELS est un banc d'entraînement permettant d'analyser la qualité de réalisation de 5 tâches basiques de laparoscopie (transfert d'objets, découpe d'un cercle, placement d'un endoloop, suture extra-corporelle, suture intra-corporelle).

Le score est basé sur une combinaison du temps de réalisation et sur une mesure de la précision (pénalités associées à la réalisation d'erreurs).

Cette méthode, peu chère, a été adoptée par la *Society of American Gastrointestinal and Endoscopic Surgeons* (SAGES) et l'*American College of Surgeons* (ACS) pour évaluer les qualités techniques des internes en chirurgie digestive. Elle a donc été largement étudiée, avec des arguments forts pour la validité de construction : Derossis et al. [1998, 1999] démontrent l'existence d'une corrélation entre le score obtenu et l'année de formation des internes évalués à l'aide d'une régression linéaire, Fraser et al. [2003] parviennent à séparer des chirurgiens experts et des internes en définissant un score seuil, Fried et al. [2004] et Swanstrom et al. [2006] montrent que les résultats obtenus diffèrent de façon statistiquement significative entre experts et internes (ANOVA), et que les scores des internes progressaient avec l'entraînement

(test de Wilcoxon). Enfin McCluney et al. [2007] démontrent que le score FLS est corrélé au niveau de pratique (ANOVA) et dans le même temps comparent pour chaque participant le score obtenu sur le banc d'essai avec le score obtenu au bloc opératoire avec la méthode d'évaluation GOALS. Ils démontrent une validité concurrente entre les scores FLS sur banc d'essai et GOALS au bloc opératoire, dans la mesure où ces scores sont significativement corrélés avec une analyse par régression linéaire.

Feldman et al. [2004] et Fried et al. [1999, 2004] ont également démontré une validité concurrente en comparant le score FLS respectivement au score ITERs (*In-Training Evaluation Reports*) et à des scores de performance au bloc opératoire (score FLS modifié pour s'adapter à un modèle animal et GOALS).

Enfin Vassiliou et al. [2006] ont étudié la fiabilité des mesures et ont mis en évidence une reproductibilité inter-observateur excellente.

Ce score est associé à des valeurs seuil permettant une évaluation certificative (validation d'acquis), qui ont été déterminés par McCluney et al. [2007] et Fraser et al. [2003]. Malheureusement, les scores publiés ne sont pas ceux utilisés par la SAGES dans son module de formation à la chirurgie laparoscopique, ce qui limite la disponibilité des plate-formes d'évaluation à celles agréées par la société.

Un autre dispositif similaire au score FLS est le *Advanced Dundee Endoscopic Psychomotor Tester* (ADEPT) proposé par [Hanna et al., 1998]. Ce banc d'entraînement ne reproduit pas des tâches basiques de chirurgie laparoscopique, mais demande au volontaire d'effectuer des gestes simples avec une main en libérant l'espace pour accéder à la zone de travail avec l'autre main (de la même façon qu'un chirurgien doit parfois écarter une structure anatomique pour exposer sa cible, le volontaire doit translater un plan). Le score combine un temps nécessaire à la réalisation du geste à des pénalités liées à des contacts non prévus entre un instrument et le plan ou le fond de la zone de travail. Macmillan et Cuschieri [1999] et Francis et Hanna [2002] ont démontré respectivement la validité concurrente et la validité de construction de ce score.

Ce dispositif présente cependant l'inconvénient de proposer des exercices qui sont très éloignés de ce que peut rencontrer un chirurgien au cours d'une intervention, avec une validité apparente qui est donc médiocre. Il est également techniquement plus complexe à installer que le banc d'entraînement MISTELS.

Une méthode intéressante d'analyse de la qualité du geste, valable uniquement dans le cadre du laboratoire, est proposée par Frischknecht et al. [2013]. Ils parviennent à classer des volontaires en fonction de leur niveau en utilisant un algorithme de traitement d'image pour extraire les caractéristiques d'une suture (nombre de points, longueur des point, orientation des points, ratio longueur plaie/ longueur suture, symétrie des points,...). Cette méthodologie pourrait permettre de ne pas utiliser d'observateur externe pour mesurer les erreurs commises lors du test

FLS par exemple.

Enfin, des simulateurs de réalité virtuelle peuvent avoir une détection automatique de la précision des gestes effectués avec un décompte automatique des erreurs (cible manquée, lâcher d'objet par exemple pour le daVinci Skills Simulator® [Lyons et al., 2013], le MIST®-VR [Gallagher et Satava, 2002] ou le LapSim [Eriksen et Grantcharov, 2005]) mais aucun score n'a encore été validé pour exploiter ces informations.

ii) Analyse du mouvement

L'analyse du mouvement lors de la réalisation de gestes opératoires est une des méthodes les plus prometteuses pour évaluer ces derniers, à condition d'arriver d'une part à acquérir les paramètres du mouvement sans perturber le déroulement du geste et d'autre part à les corréler à une mesure fiable de la qualité [Mason et al., 2012].

Les mouvements considérés peuvent être des déplacements des effecteurs (des instruments) ou des déplacements des commandes (les mains). Cette différence est importante, en particulier si on considère des gestes effectués en laparoscopie, dans la mesure où des petits mouvements des mains peuvent se traduire par de grands déplacements de la partie effectrice de l'instrument (pince, ciseaux,...) ou inversement (en cas de rotation autour de l'axe de l'instrument par exemple). On peut également étudier l'effet des mouvements, c'est à dire les forces exercées sur les tissus ou les structures de support (trocarts par exemple).

De nombreuses études ont porté sur ce sujet, identifiant plusieurs paramètres pouvant servir de support pour l'analyse de la qualité du geste. Dans la suite de ce manuscrit, je nomme ces paramètres "métriques", dans le sens défini par Cotin et al. [2002]. Je présente leurs principales caractéristiques ci-après.

▷ **Métriques étudiées** Le tableau 2.3 résume les différentes métriques associées aux méthodes d'acquisition de la trajectoire.

le temps nécessaire à la réalisation d'un geste est le paramètre qui paraît le plus accessible et logique pour analyser la qualité de ce dernier, et est effectivement presque systématiquement identifié comme paramètre pertinent pour quantifier la qualité du geste dans la littérature ([Tokunaga et al., 2012, Moorthy et al., 2004, Van Sickle et al., 2005, Aggarwal et al., 2007, Tanoue et al., 2010] entres autres). De façon intéressante, Cotin et al. [2002] montrent que, pour la réalisation de trois tâches spécifiques, le temps ne permet pas isolément de différencier les participants en fonction de leur niveau de pratique. Des variantes de la mesure simple du temps existent pour certaines tâches spécifiques, tels que le temps d'approche [Egi et al., 2010, 2012, Tokunaga et al., 2012], le temps passé avec les instruments hors du champ de vision [Lyons et al., 2013]

ou le temps passé dans la zone de travail et le temps d'immobilité [Oropesa et al., 2013].

la longueur de chemin parcouru, aussi bien par l'extrémité de l'instrument que par la main, est également un paramètre fréquemment analysé. Elle est une première approche pour mesurer l'économie de mouvement, une distribution "compacte" étant par exemple caractéristique du mouvement d'un expert selon Cotin et al. [2002]. Ce paramètre est identifié dans plusieurs contextes d'évaluation, par Smith et al. [2002] pour la cholecystectomie dans un modèle animal, par Moorthy et al. [2004] pour des sutures dans un banc d'entraînement, par Xeroulis et al. [2009] pour les tâches du banc d'entraînement MISTELS, par Van Sickle et al. [2005], Pellen et al. [2009b,a] et Oostema et al. [2008] dans le cadre d'un simulateur de réalité virtuelle, ou encore Judkins et al. [2009] pour des tâches d'exercice avec le robot DaVinci. Aggarwal et al. [2007] identifient également ce facteur comme élément prédictif du niveau de pratique du chirurgien lors de cholecystectomies sur patient au bloc opératoire. De même que pour le temps, Cotin et al. [2002] trouvent que la longueur de chemin parcouru n'est pas un facteur suffisamment discriminant pour prédire la qualité de la chirurgie. Taffinder et al. [1998] étudient le rapport entre la longueur de chemin parcouru et la longueur du chemin idéal pour des sous-tâches simples permettant la définition d'un chemin optimal, tandis que Egi et al. [2012] étudient la déviation par rapport à un chemin de ce type.

le nombre de mouvements effectués est une autre façon d'aborder l'économie de mouvement. Ce paramètre est isolé par Smith et al. [2002], Taffinder et al. [1998], Gallagher et Satava [2002], Xeroulis et al. [2009], Broe et al. [2006], Lyons et al. [2013] et Aggarwal et al. [2007]. La définition d'un mouvement peut changer d'une publication à l'autre en fonction du dispositif d'acquisition du mouvement.

la vitesse moyenne des mouvements est identifiée comme facteur prédictif par Smith et al. [2002] et Judkins et al. [2009] ainsi que par Oropesa et al. [2013].

la fluidité du geste mesurée par le *jerk*, c'est à dire la dérivée troisième de la trajectoire, dans les travaux de Cotin et al. [2002] et Chmarra et al. [2010], ou plus grossièrement par le nombre de changements de la vélocité dans les travaux de Van Sickle et al. [2005], Oostema et al. [2008] et Pellen et al. [2009b,a], a également été identifiée dans ces publications comme variable prédictive de la qualité du geste. Oropesa et al. [2013] ne trouvent en revanche pas de corrélation entre cette mesure et le niveau des participants à l'étude.

le déplacement selon l'axe de l'instrument est une façon d'étudier la perception de la profondeur du chirurgien. Ce paramètre est prédictif de la qualité du geste dans les travaux de Cotin et al. [2002], Chmarra et al. [2010] et de Oropesa et al. [2013].

la rotation autour de l'axe de l'instrument mesure la capacité à orienter de façon

optimale l'instrument pour se servir de la pince ou des ciseaux situés à son extrémité. Ce paramètre est isolé par Cotin et al. [2002].

l'accélération moyenne des instruments, au même titre que la vitesse moyenne, est identifié comme paramètre prédictif par Oropesa et al. [2013].

l'économie de volume est calculée comme le rapport entre le volume exploré par la pointe de l'instrument et la longueur du chemin parcouru par cette pointe. On peut supposer qu'un chirurgien expert aura davantage l'habitude de travailler dans de petits volumes et s'y astreindra, afin de ne pas risquer de léser des structures adjacentes. Ce paramètre est identifié par Oropesa et al. [2013] comme prédictif de la qualité du geste. Le volume exploré et l'aire angulaire explorée, sans normalisation par la longueur de chemin parcouru, sont également retenus par Chmarra et al. [2010].

le rayon de courbure de la trajectoire a également été retenu par Judkins et al. [2009] pour quantifier la qualité d'un geste chirurgical.

les vitesses angulaires des instruments ou des membres du chirurgien, ainsi que des mesures dérivées de ces vitesses angulaires (fréquence des pics de vitesse, écart-type) ont été étudiées par Lin et al. [2012] qui montrent une corrélation entre ces paramètres et le niveau de pratique du chirurgien. Ces paramètres sont également retrouvés par Eriksen et Grantcharov [2005].

les forces appliquées aux instruments ou aux tissus apparaissent comme un paramètre candidat logique pour prédire la qualité du geste, les chirurgiens experts manipulant les tissus avec davantage de soin qu'un chirugien novice. Judkins et al. [2009] ne parviennent cependant pas à montrer cet effet sur un robot chirurgical (da Vinci Surgical System®), et mesurent même l'effet inverse. Une explication potentielle à ce comportement inattendu est que le robot ne propose pas de retour haptique, c'est à dire que le chirurgien ne parvient pas à sentir quelle force il applique sur les tissus. a contrario Lyons et al. [2013] parviennent sur un simulateur de ce robot, mais avec des exercices différents, à corréler ce facteur à la qualité du geste, ce qui souligne l'importance du type d'exercice proposé pour identifier des paramètres prédictifs de la qualité du geste (la validité apparente des exercices doit être importante afin de pouvoir généraliser ces mesures dans d'autres contextes, en particulier au bloc opératoire). De même, Yamauchi et al. [2002] montrent que l'intégrale des forces appliquées sur un tissu par un endoscope lors d'une simulation de chirurgie endoscopique sinusienne sur un banc d'entraînement (donc avec un retour haptique) est inversement corrélée au niveau de pratique du chirurgien.

Référence	Métriques explorées	Acquisition des données	Niveau de granularité	Contexte d'utilisation	Validité
Cotin et al. [2002]	Temps, longueur de chemin parcouru, fluidité du mouvement, perception de la profondeur, contrôle de l'orientation	*Virtual Laparoscopic Interface* modifié	Tâches	Laboratoire (banc d'entraînement)	Construction
Van Sickle et al. [2005], Oostema et al. [2008], Pellen et al. [2009b,a]	Temps, longueur de chemin parcouru, fluidité du mouvement	Simulateur chirurgical ProMIS	Sous-tâches	Laboratoire (simulateur de réalité virtuelle)	Construction et concurrente (contre notation structurée [Pellen et al., 2009a])
Broe et al. [2006]	Temps, nombre de mouvements	Simulateur chirurgical ProMIS	Tâches et sous-tâches	Laboratoire (simulateur de réalité virtuelle)	Construction
Smith et al. [2002]	Temps, longueur de chemin parcouru, vitesse des mouvements, nombre de mouvements	ICSAD (marqueurs magnétiques)	Tâches au cours d'une procédure unique (cholecystectomie)	Laboratoire (modèle animal)	Construction et apparente
Moorthy et al. [2004]	Temps, longueur de chemin parcouru	ICSAD	Tâches	Laboratoire (banc d'entraînement)	Construction et concurrente (contre *check-list*)
Xeroulis et al. [2009]	Temps, longueur de chemin parcouru, nombre de mouvements	ICSAD	Tâches	Laboratoire (banc d'entraînement MISTELS)	Construction et concurrente (contre score FLS)
Aggarwal et al. [2007]	Temps, longueur de chemin parcouru, nombre de mouvements	ICSAD	Procédure	Bloc opératoire	Construction et concurrente (contre OSATS)
Tanoue et al. [2010]	Temps	Simulateur chirurgical LapSim	Sous-tâches	Laboratoire (simulateur de réalité virtuelle)	Construction
Eriksen et Grantcharov [2005]	Temps, longueur de chemin parcouru, distance angulaire parcourue	Simulateur chirurgical LapSim	Tâches et sous-tâches	Laboratoire (simulateur de réalité virtuelle)	Construction
Taffinder et al. [1998], Gallagher et al. [2001], Gallagher et Satava [2002]	Temps, nombre de mouvements	Simulateur chirurgical MIST®-VR	Tâches et sous-tâches	Laboratoire (simulateur de réalité virtuelle)	Construction
Sarker et al. [2012]	Temps, longueur de chemin parcouru, nombre de mouvements	Simulateur chirurgical SimBionix	Tâches	Laboratoire (simulateur de réalité virtuelle	Construction
	Angles des articulations du chirurgien, vitesses angulaires, distribution fréquentielle des vitesses angulaires, fonction de distribution cumulative des vitesses angulaires, densité spectrale de puissance des vitesses angulaires	Dispositif de mesure inertielle (accéléromètres et gyromètres, WB-3)	Tâches	Laboratoire (banc d'entraînement)	Construction
Egi et al. [2010, 2012]	Temps, déviation du chemin parcouru par rapport à un chemin optimal	HUESAD (suivi mécanique)	Sous-tâches	Laboratoire (banc d'entraînement)	Construction

Tokunaga et al. [2012]	Temps, temps d'approche	HUESAD	Sous-tâches	Laboratoire (banc d'entraînement)	Construction
Chmarra et al. [2010]	Temps, longueur de chemin parcouru, variations de profondeur de l'instrument, fluidité du mouvement, aire angulaire explorée, volume exploré	TrEndo (suivi mécanique)	Tâches	Laboratoire (banc d'entraînement)	Construction
Oropesa et al. [2013]	Temps, longueur de chemin parcouru, variations de profondeur, vitesse moyenne, accélération moyenne, économie d'aire, économie de volume	*Endoscopic Video Analysis* (suivi par analyse d'images)	Tâches	Laboratoire (banc d'entraînement)	Construction et concurrente (contre TrEndo)
Judkins et al. [2009]	Temps ; longueur du chemin parcouru, vitesse moyenne, rayon de courbure, phase relative des déplacements des instruments	Données de cinématique du robot daVinci	Tâches	Laboratoire (banc d'entrainement)	Construction
Lyons et al. [2013]	Temps, longueur de chemin parcouru, volume exploré par les mains du chirurgien, collisions entre instruments, force appliquée, erreurs (lâchers, cibles manquées,...)	Simulateur de robot chirurgical daVinci	Tâches	Laboratoire (simulateur chirurgical)	Construction, contenu, apparente (évaluées par des scores globaux par des experts)
Yamauchi et al. [2002]	Pressions appliquées aux tissus	Capteurs de pressions sur simulateur d'endoscopie sinusienne	Tâches	Laboratoire (banc d'entraînement)	Construction

Tab. 2.3. – Métriques précédemment explorées dans la littérature pour l'analyse du mouvement, avec leurs principales caractéristiques d'utilisation
ICSAD : Imperial College Surgical Assessment Device ; HUESAD : Hiroshima University Endoscopic Surgical Assessment Device

▷ **Niveau de détail exploré** La plupart des études ayant été conduites au laboratoire, l'analyse de la qualité du geste chirurgical a porté sur des tâches, voire des activités, plutôt que sur l'ensemble de la procédure. Seule l'étude de Aggarwal et al. [2007] conduite au bloc opératoire a analysé la procédure chirurgicale comme un seul bloc. Or il apparaît que les paramètres prédictifs de la qualité du geste chirurgical ne sont pas les mêmes en fonction de la tâche étudiée, comme le démontrent Lyons et al. [2013] dans leur étude.

Le niveau de détail et la tâche pour lesquels des paramètres sont isolés sont donc des variables d'importance lors de l'implémentation d'une méthode d'évaluation de la qualité du geste ; la possibilité d'identifier la tâche active au cours d'une procédure grâce aux SPMs afin de sélectionner les paramètres pertinents pour analyser sa réalisation est un critère de succès majeur de la méthodologie.

▷ **Méthode d'acquisition des données** On distingue deux grandes familles dans les études présentées ci-dessus, celles utilisant des marqueurs pour suivre la trajectoire d'un instrument, et celles utilisant des données issues de simulateurs de réalité virtuelle. Je reviens plus loin dans ce manuscrit (section 3.1) sur les avantages et inconvénients des différentes façons envisageables pour acquérir le mouvement des objets suivis (mains ou instruments), insistant ici sur la diversité des méthodes ayant permis d'extraire les métriques décrites au point précédent.

Parmi les méthodes utilisant des marqueurs, Smith et al. [2002], Moorthy et al. [2004], Aggarwal et al. [2007], Xeroulis et al. [2009] utilisent des réflecteurs électromagnétiques (*Imperial College Surgical Assessment Device*, ICSAD) pour acquérir en temps réel la position des instruments ou de la main du chirurgien, Lin et al. [2012] utilisent des accéléromètres associés à des gyromètres (*WB-3 Inertial Measurement Unit*) pour suivre le mouvement des membres de l'opérateur, tandis que Cotin et al. [2002],Chmarra et al. [2010] et Egi et al. [2008, 2010, 2012], Tokunaga et al. [2012] ont développé indépendamment des systèmes de suivi mécanique (respectivement *Virtual Laparoscopic Interface* (modifié), *TrEndo Tracking System* et *Hiroshima University Endoscopic Surgical Assessment Device*). Judkins et al. [2009] récupèrent les données de cinématique du robot chirurgical daVinci grâce à un accès à l'API. Enfin Oropesa et al. [2013] proposent une méthode d'extraction de la trajectoire à partir des images endoscopiques.

À la frontière entre méthodes utilisant des marqueurs et simulateurs de réalité virtuelle, on peut citer le dispositif ProMIS™ qui extrait la trajectoire à partir de multiples caméras, avant de superposer à l'image une simulation de cavité abdominale. Ce dispositif est utilisé par Van Sickle et al. [2005], Oostema et al. [2008] et Pellen et al. [2009b,a].

L'intérêt des simulateurs de réalité virtuelle est la facilité de récupération des données de cinématique des instruments, le système étant construit autour d'un dispositif spécialement dédié à cette tâche (les mouvements du chirurgien sont les variables d'entrée de la simulation).

Parmi les résultats présentés ci-dessus, Eriksen et Grantcharov [2005] et Tanoue et al. [2010] utilisent le simulateur LapSim qui modélise des tâches de différentes difficultés (navigation de la caméra, des instruments, transfert d'objets, ligature par clip et section de vaisseaux, suture), tandis que Gallagher et al. [2001], Gallagher et Satava [2002] et Taffinder et al. [1998] utilisent le dispositif *Minimally Invasive Surgical Trainer-Virtual Reality* (MIST®-VR) et que Sarker et al. [2012] utilisent le simulateur SimBionix. Enfin Lyons et al. [2013] utilisent un simulateur chirurgical du robot daVinci dont la manipulation est très proche de celle du robot original (bonne validité apparente).

▷ **Type et domaine de chirurgie** La plupart des méthodes présentées ici ont été développées pour l'analyse du mouvement en chirurgie laparoscopique, y compris robotisée [Lyons et al., 2013], à l'exception de Yamauchi et al. [2002] qui s'intéressent à la chirurgie ORL endoscopique. Le dispositif ICSAD a par ailleurs été évalué pour des gestes de chirurgie ouverte, et les paramètres précédemment décrits étaient corrélés au niveau de pratique du chirurgien [Datta et al., 2001].

▷ **Contexte d'utilisation** Toutes les méthodes présentées ci-dessus ont été évaluées dans le contexte du laboratoire, à l'exception de l'étude de Aggarwal et al. [2007], qui a validé les métriques isolées précédemment à l'aide du dispositif ICSAD dans le contexte du bloc opératoire.

L'analyse du mouvement apparaît donc comme une alternative intéressante aux méthodes de notation structurée, dans la mesure où l'acquisition des données est réalisée en général par un système automatisé, ce qui élimine des biais de subjectivité potentiels et qui est plus économe en ressources. Cette automatisation de l'acquisition des données autorise également davantage de flexibilité, le chirurgien pouvant choisir plus librement le moment auquel il souhaite s'évaluer.

Les bancs d'entraînement de type FLS nécessitent un observateur non nécessairement expert pour évaluer la qualité du geste, les rendant moins coûteux et davantage disponibles, mais ne pourront jamais permettre d'évaluer directement la qualité d'un geste dans le contexte du bloc opératoire.

Dans l'état actuel du domaine, l'analyse du mouvement est encore fortement dépendante des dispositifs d'acquisition des trajectoires, dont le coût reste trop élevé pour permettre leur diffusion dans de nombreuses structures de chirurgie (environ 5000$ pour le dispositif ICSAD, plus de 100000$ pour des simulateurs de réalité virtuelle [Wanzel et al., 2002]), dont l'usage au bloc opératoire n'est pas forcément possible ou dont l'acceptabilité au sein des blocs opératoires pour enregistrer une procédure est médiocre (contraintes fortes sur l'environnement chirurgical, voir section 3.1).

2.2.4 Modélisations statistiques

Une limitation à l'utilisation des statistiques descriptives à l'aide des métriques décrites ci-dessus est la variabilité possible dans le déroulement d'une procédure chirurgicale : des différences dans le temps de réalisation d'une tâche ou dans la longueur de chemin parcouru peuvent parfois refléter davantage des variations dans la technique chirurgicale adoptée que des variations du niveau de qualité du geste. La modélisation de la procédure permet de contourner cette difficulté, soit en adaptant les métriques aux différentes tâches reconnues (voir section 2.1.2), soit en comparant directement les représentations du geste à l'aide de modèles.

Comme le notent Rosen et al. [2006], il existe des similarités entre l'analyse du déroulement d'un geste chirurgical et l'analyse de la parole. Dans ce dernier domaine, les modèles de Markov sont une méthode de reconnaissance de la structure du langage largement étudiée et validée. Il apparaît donc intéressant d'étudier l'application de ce type de modélisation statistique à l'analyse de la qualité du geste chirurgical.

Une autre approche, proposée par Forestier et al. [2011, 2012], utilise une distance pour comparer deux processus chirurgicaux à l'aide d'un algorithme de *dynamic time warping*.

i) Modèles de Markov et modèles de Markov cachés

▷ **Paramètres étudiés** Les premiers travaux publiés par l'équipe de Hannaford [Rosen et al., 1999, 2001a,b, 2002] analysent comment les forces et couples appliqués sur les instruments pouvaient servir de signatures pour différentes actions des instruments (immobilité, préhension, poussée, écartement,...). Ces paramètres servent d'observations pour la création de modélisations de la procédure grâce à des modèles de Markov dont les états sont les actions des instruments. L'analyse de la qualité du geste s'appuie d'une part sur l'étude du temps passé dans chaque état, sur les transitions utilisées pour passer d'un état à un autre, sur les états visités par le modèle et sur l'amplitude des forces et couples appliqués, d'autre part sur une mesure de la similarité entre les modèles créés pour chaque volontaire (mesure de la probabilité d'une série d'observations d'un chirurgien avec le modèle créé pour un autre chirurgien). Dans des travaux postérieurs [Rosen et al., 2006], la même équipe introduit comme observation supplémentaire la position de la pointe de l'instrument. Murphy et al. [2003] se basent également sur les données de cinématique et de force exercées par les instruments pour construire un modèle de Markov permettant d'identifier les mouvements du chirurgien et de classer celui-ci en fonction de leur nombre et de leur durée.

Leong et al. [2006] choisissent également une approche basée sur la position de la pointe de l'instrument comme observable (après transformation en *centroid distance function*), et utilisent des modèles de Markov cachés pour modéliser la trajectoire

(en ne définissant pas a priori les états du modèle). Ils définissent une mesure de similarité entre le modèle moyen issu des observations d'un groupe de chirurgiens et la trajectoire d'un chirurgien particulier. Cette mesure permet de classer les chirurgiens en fonction de la distance de leur trajectoire au modèle, classement qui est corrélé dans cet article à la qualité du geste mesurée avec le score OSATS. Megali et al. [2006] ont une approche similaire (HMM) en choisissant comme observation l'accélération de la pointe de l'instrument (après transformation par *short-time Fourier transform*). Ils calculent ensuite une mesure de similarité entre le modèle issu des observations de deux experts et les observations d'un chirurgien à évaluer.

Jayender et al. [2012] analysent des paramètres dérivés de la trajectoire de marqueurs fixés sur un endoscope souple de coloscopie (position absolue, position relative, orientation relative,...) pour entraîner des modèles de Markov dont les états sont les différents gestes réalisables par le manipulateur (insertion, rétraction, rotation,...). L'analyse de la qualité du geste est réalisée par le calcul de la probabilité qu'un modèle entraîné avec des données de chirurgien expert explique les observations d'un chirurgien à évaluer.

▷ **Niveau de détail exploré** L'utilisation d'une modélisation statistique du geste du chirurgien permet d'analyser sa qualité à différentes échelles. Rosen et al. [2001b,a, 2002] enregistrent une procédure dans son ensemble et utilisent les temps de réalisation des étapes chirurgicales comme mesure de la qualité de la procédure, sans analyser la qualité de réalisation des étapes prises séparément. Murphy et al. [2003] ont la même approche au niveau de granularité supérieur : ils reconnaissent au cours d'une tâche différentes activités et classent les participants en fonction du temps passé dans chaque activité.

Leong et al. [2006], Megali et al. [2006] et Rosen et al. [2006] comparent des modèles décrivant le geste respectivement au niveau du mouvement, de l'activité et de la tâche.

▷ **Acquisition des données** Dans les travaux de l'équipe de Hannaford, les forces et couples appliqués aux instruments sont acquis grâce à des capteurs installés sur ces derniers, tandis que les actions des instruments sont définies, pour l'entraînement du modèle, par l'annotation de la vidéo des procédures, image par image. Cette approche est coûteuse en temps, ce qui ne permet pas sa diffusion dans d'autres contextes que la recherche. Les derniers travaux de l'équipe [Rosen et al., 2006] intègrent également un dispositif de suivi mécanique de la position de la pointe de l'instrument (Blue DRAGON). L'annotation manuelle de vidéos pour retrouver les états à décrire dans le modèle de Markov est également utilisée dans les travaux de Murphy et al. [2003], qui acquièrent les données de cinématique et de force grâce à un dispositif haptique (3GM). Leong et al. [2006] utilisent un dispositif de marqueurs optiques (Polaris) pour suivre la trajectoire des instruments au cours du geste étudié. La vidéo de la procédure n'a pas besoin d'être annotée, les états du

modèle étant déterminés de façon statistique à partir de l'ensemble des trajectoires acquises pour l'entraîner. De même Megali et al. [2006] n'ont besoin d'acquérir que les données de trajectoire des instruments, ce qu'ils réalisent à l'aide du dispositif *Virtual Laparoscopic Interface* couplé au simulateur de réalité virtuelle LapSim.

▷ **Type et domaine de chirurgie** Toutes les études citées précédemment s'intéressent à la chirurgie laparoscopique, à l'exception des travaux de [Jayender et al., 2012] qui s'appliquent à l'endoscopie digestive basse. Les gestes de laparoscopie sont abordés de façon générique pour les études sur simulateurs (tâches peu spécifiques, de type pointage d'objets, ou plus spécifiques, de type sutures), en s'intéressant particulièrement à deux types d'intervention (fundoplicature de Nissen et cholecystectomie) pour les travaux de Rosen et al. [1999, 2001b,a, 2002] sur modèles animaux.

▷ **Contexte d'utilisation** Les travaux portant sur la laporoscopie ont tous été menés dans le contexte du laboratoire, sur des modèles animaux [Rosen et al., 1999, 2001b,a, 2002, 2006], sur des bancs d'entraînement [Leong et al., 2006] ou sur des simulateurs de réalité virtuelle [Megali et al., 2006, Murphy et al., 2003]. L'étude de Jayender et al. [2012] portant sur la coloscopie a été réalisée au bloc opératoire.

ii) Dynamic Time Warping

Forestier et al. [2011] proposent une analyse de la qualité du geste chirurgical basée sur l'algorithme de recalage temporel *dynamic time warping*. Les paramètres d'entrée de la méthode étudiée ici sont les activités définies par les triplets <*action, structure, instrument*>. Ces activités sont enregistrées par un observateur externe au cours d'une procédure de cure de hernie lombaire au bloc opératoire (une seule étape est analysée, la fermeture). La comparaison de la succession des activités après recalage temporel, par groupement à liaison moyenne, permet d'identifier deux groupes de procédures qui sont fortement corrélés au niveau d'expérience des chirurgiens conduisant ces procédures.

Cette approche permet de comparer efficacement la façon dont se déroulent des procédures en s'affranchissant du facteur temps grâce au recalage temporel.

2.3 Conclusion et problématique de la thèse

Il est important de noter, avant de conclure ce chapitre, que la majorité des méthodes existantes décrites ici, et en particulier celles utilisant des statistiques descriptives et des modélisations statistiques, sont encore dans le domaine de la recherche. Les performances atteintes en terme de reconnaissance ou de pouvoir prédictif sont encore trop faibles pour permettre une utilisation en routine, et les

Méthode	Fiabilité	Validité	Faisabilité	Exhausitivité	Flexibilité
Échelles de notation graduée					
OSATS	Reproductibilité inter- (0.64-0.72) et intra-observateur (0.61-0.85)	Construction, contenu, apparente	Moyenne : observateur externe, coût	Bonne : description fine de la procédure grâce aux *check-lists*	Moyenne : *check-lists* spécifiques à une procédure
GOALS	Reproductibilité inter- et intra-observateur	Construction, contenu, apparente	Moyenne : observateur externe, coût	Faible : description générique de la procédure	Bonne : utilisable sans modification pour plusieurs types de chirurgie
OCHRA	Reproductibilité inter- et intra-observateur	Construction, contenu, apparente	Faible : observateur externe, enregistrement vidéo, coût	Forte : explore différents niveaux de granularité, plusieurs types d'erreurs	Faible : nécessite une décomposition de la procédure (très spécifique)
Statistiques descriptives					
FLS	Reproductibilité intra- (0,66-,096) et inter-observateur (0,98-1,00) très bonnes	Construction, concurrente (contre GOALS, ITER), contenu, apparente	Bonne : observateur non expert, coût modique	Faible : n'explore que 5 tâches spécifiques, pas d'évaluation de la procédure	Faible : score spécifique pour 5 tâches prédéfinies
ADEPT	Excellente (acquisition automatique des erreurs)	Construction, concurrente	Faible : dispositif spécifique d'acquisition des données	Faible : n'explore que quelques composantes du geste (exposition et gestes simples)	Faible : dispositif conçu autour de ces gestes
Méthodes d'analyse du mouvement	Bonne en général (acquisition automatisée des données)	se reporter au tableau 2.3	Limitée au laboratoire en général sauf Aggarwal et al. [2007], perturbations fréquentes au bloc opératoire	Bonne en combinant les mesures	Faible (dispositifs d'acquisition des données spécifiques)
Modélisations statistiques	Bonne (acquisition automatisée des données)	Construction	Limitée (dispositifs d'acquisition des données spécifiques, non utilisables au bloc opératoire)	Bonne (plusieurs phases de la procédure explorées)	Faible (dispositifs d'acquisition des données spécifiques)

Tab. 2.4. – Comparaison des caractéristiques des différentes méthodes d'évaluation de la qualité du geste décrites dans ce chapitre. La grande variabilité des méthodes d'analyse du mouvement ne permet pas de donner des caractéristiques communes à l'ensemble de ces méthodes.
OSATS : Objective Structured Assessment of Technical Skills ; GOALS : Global Operative Assessment of Laparoscopic Skills ; OCHRA : Objective Clinical Human Reliability Assessment ; FLS : Fundamentals of Laparoscopic Skills ; ADEPT : Advanced Dundee Endoscopic Psychomotor Tester

évaluations ont été généralement réalisées dans des conditions de simulations.

Cet état de l'art permet d'isoler plusieurs problématiques relatives à la quantification de la qualité d'un geste en chirurgie. La première est le niveau de granularité auquel l'observation va se faire, ainsi que la façon de situer une observation effectuée à un niveau de granularité fin dans le déroulé de la procédure chirurgicale. La seconde est la méthode d'acquisition des données : afin de permettre une implémentation de dispositifs de mesure de la qualité du geste dans les centres d'enseignement et de pratique de la chirurgie, ces dispositifs doivent avoir un coût financier mais également humain limité (ce qui exclut les systèmes de notation graduée), être portables aisément d'un centre à un autre (pas de matériel spécifique) et ne pas interférer avec le déroulement de la procédure chirurgicale. Idéalement ces dispositifs devraient pouvoir être utilisés dans le cadre du laboratoire pour la phase d'entraînement initial (début de la courbe d'apprentissage) mais également au bloc opératoire pour la pratique courante (optimisation du geste). La dernière problématique est celle des paramètres analysés : nous avons vu que de nombreuses mesures pouvaient être effectuées autour du geste chirurgical, allant de paramètres physiques (temps et trajectoires) à des mesures de similarité entre modélisations statistiques du geste. La mesure idéale, c'est à dire valide, fiable et faisable, est probablement une combinaison de plusieurs des mesures précédemment citées.

Cette thèse se focalise sur les deux dernières problématiques. Je fais le postulat que le développement de modèles de la procédure chirurgicale permettront, en couplant plusieurs dispositifs complémentaires (reconnaissance automatique du type d'instrument employé, du type de geste, des organes visualisés, des modalités d'imagerie per-opératoire utilisées à un instant donné,...), d'identifier à un instant donné l'étape et la tâche courantes, et donc d'étudier spécifiquement la qualité de réalisation de ces dernières. Je n'aborderai donc pas dans ce manuscrit la comparaison de modèles de procédures chirurgicales entre eux, même si cette mesure est probablement pertinente pour évaluer la qualité globale du geste et identifier des erreurs de procédure.

Je me concentre sur la qualité d'exécution du geste au niveau de granularité de la tâche grâce à une méthode d'analyse non supervisée basée sur des connaissances à priori concernant la distribution statistique de métriques mesurables en fonction du niveau du chirurgien. Je présente dans un premier temps un dispositif d'acquisition de la position des instruments utilisés en chirurgie laparoscopique basé sur l'analyse des images endoscopiques, puis détermine dans un second temps quelles métriques peuvent être extraites des trajectoires enregistrées pour quantifier la qualité de réalisation de tâches. Je conclurai en présentant des pistes permettant d'améliorer cette analyse, en intégrant ce dispositif dans un environnement plus riche incluant des modalités d'imagerie per-opératoire afin d'étudier la trajectoire des instruments dans un contexte anatomique.

Références

R. Aggarwal, T. Grantcharov, K. Moorthy, T. Milland et A. Darzi : Toward feasible, valid, and reliable video-based assessments of technical surgical skills in the operating room. Annals of surgery, 247(2):372, 2008. 41, 43

R. Aggarwal, T. Grantcharov, K. Moorthy, T. Milland, P. Papasavas, A. Dosis, F. Bello et A. Darzi : An evaluation of the feasibility, validity, and reliability of laparoscopic skills assessment in the operating room. Annals of surgery, 245(6):992, 2007. 41, 49, 50, 52, 54, 55, 59

Seyed-Ahmad Ahmadi, Tobias Sielhorst, Ralf Stauder, Martin Horn, Hubertus Feussner et Nassir Navab : Recovery of surgical workflow without explicit models. In Rasmus Larsen, Mads Nielsen et Jon Sporring, éditeurs : Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006, Lecture Notes in Computer Science, pages 420–428. Springer Berlin Heidelberg, janvier 2006. 36

Kamran Ahmed, Danilo Miskovic, Ara Darzi, Thanos Athanasiou et George B. Hanna : Observational tools for assessment of procedural skills : a systematic review. The American Journal of Surgery, 202 (4):469–480.e6, octobre 2011. 40

J. D. Beard, S. Choksy et S. Khan : Assessment of operative competence during carotid endarterectomy. British Journal of Surgery, 94(6):726–730, 2007. 41

J.D. Beard : Setting standards for the assessment of operative competence. European Journal of Vascular and Endovascular Surgery, 30(2):215–218, 2005. 41

JD Beard, BC Jolly, DI Newble, WEG Thomas, J. Donnelly et LJ Southgate : Assessing the technical skills of surgical trainees. British Journal of Surgery, 92(6):778–782, 2005. 43

Tobias Blum, Nicolas Padoy, Hubertus Feussner et Nassir Navab : Workflow mining for visualization and analysis of surgeries. International Journal of Computer Assisted Radiology and Surgery, 3(5):379–386, novembre 2008. 37

David Bouget, Florent Lalys et Pierre Jannin : Surgical tools recognition and pupil segmentation for cataract surgical process modeling. In Studies in Health Technology and Informatics, volume 173, pages 78–84, février 2012. 37

D. Broe, P. F. Ridgway, S. Johnson, S. Tierney et K. C. Conlon : Construct validation of a novel hybrid surgical simulator. Surgical Endoscopy And Other Interventional Techniques, 20(6):900–904, juin 2006. 50, 52

C. Cao, C. MacKenzie et S. Payandeh : Task and motion analyses in endoscopic surgery. In Proceedings ASME Dynamic Systems and Control Division, page 583–590, 1996. 35

L. Chang, N.J. Hogle, B.B. Moore, M.J. Graham, M.N. Sinanan, R. Bailey et D.L. Fowler : Reliable assessment of laparoscopic performance in the operating room using videotape analysis. Surgical innovation, 14(2):122, 2007. 45

M.K. Chmarra, C.A. Grimbergen, F.W. Jansen et J. Dankelman : How to objectively classify residents based on their psychomotor laparoscopic skills ? Minimally Invasive Therapy & Allied Technologies, 19 (1):2–11, 2010. 50, 51, 53, 54

Ian Choy, Andras Fecso, Josephine Kwong, Tim Jackson et Allan Okrainec : Remote evaluation of laparoscopic performance using the global operative assessment of laparoscopic skills. Surgical Endoscopy, 27(2):378–383, février 2013. 45

Kevin Cleary, Audrey Kinsella et Seong K. Mun : OR 2020 workshop report : Operating room of the future. International Congress Series, 1281:832–838, 2005. 34

Stephane Cotin, Nicholas Stylopoulos, Mark Ottensmeyer, Paul Neumann, David Rattner, Steven Dawson, Takeyoshi Dohi et Ron Kikinis : Metrics for laparoscopic skills trainers : The weakest link ! Medical Image Computing and Computer-Assisted Intervention — MICCAI 2002, pages 35–43, janvier 2002. 49, 50, 51, 52, 54

A. Cuschieri et B. Tang : Human reliability analysis (HRA) techniques and observational clinical HRA. Minimally Invasive Therapy & Allied Technologies, 19(1):12–17, 2010. 45

M D Cusimano, R Cohen, W Tucker, J Murnaghan, R Kodama et R Reznick : A comparative analysis of the costs of administration of an OSCE (objective structured clinical examination). Academic Medicine, 69(7):571–6, juillet 1994. 40

D. Dath, G. Regehr, D. Birch, C. Schlachta, E. Poulin, J. Mamazza, R. Reznick et HM MacRae : Toward reliable operative assessment : the reliability and feasibility of videotaped assessment of laparoscopic technical skills. Surgical endoscopy, 18(12):1800–1804, 2004. 43

Vivek Datta, Sean Mackay, Mirren Mandalia et Ara Darzi : The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model. Journal of the American College of Surgeons, 193(5):479–485, novembre 2001. 55

A M Derossis, M Antoniuk et G M Fried : Evaluation of laparoscopic skills : a 2-year follow-up during residency training. Canadian journal of surgery. Journal canadien de chirurgie, 42(4):293–296, août 1999. PMID : 10459330. 47

Anna M. Derossis, Gerald M. Fried, Michal Abrahamowicz, Harvey H. Sigman, Jeffrey S. Barkun et Jonathan L. Meakins : Development of a model for training and evaluation of laparoscopic skills. The American Journal of Surgery, 175(6):482–487, 1998. 47

Peter J. Driscoll, Anna M. Paisley et Simon Paterson-Brown : Video assessment of basic surgical trainees' operative skills. The American Journal of Surgery, 196(2):265–272, 2008. 43

H. Egi, M. Okajima, T. Kawahara, M. Yoshimitsu, D. Sumitani, M. Tokunaga, H. Takeda, T. Itamoto et H. Ohdan : Scientific assessment of endoscopic surgical skills. Minimally Invasive Therapy & Allied Technologies, 19(1):30–34, 2010. 49, 52, 54

Hiroyuki Egi, Masazumi Okajima, Masanori Yoshimitsu, Satoshi Ikeda, Yoshihiro Miyata, Hirokazu Masugami, Tomohiro Kawahara, Yuichi Kurita, Makoto Kaneko et Toshimasa Asahara : Objective assessment of endoscopic surgical skills by analyzing direction-dependent dexterity using the hiroshima university endoscopic surgical assessment device (HUESAD). Surgery Today, 38(8):705–710, août 2008. 54

Hiroyuki Egi, Masakazu Tokunaga, Minoru Hattori et Hideki Ohdan : Evaluating the correlation between the HUESAD and OSATS scores : Concurrent validity study. Minimally invasive therapy & allied technologies : MITAT : official journal of the Society for Minimally Invasive Therapy, novembre 2012. PMID : 23134442. 49, 50, 52, 54

J. R. Eriksen et T. Grantcharov : Objective assessment of laparoscopic skills using a virtual reality stimulator. Surgical Endoscopy And Other Interventional Techniques, 19(9):1216–1219, septembre 2005. 49, 51, 52, 55

H Faulkner, G Regehr, J Martin et R Reznick : Validation of an objective structured assessment of technical skill for surgical residents. Academic Medicine, 71(12):1363–5, décembre 1996. 40

L.S. Feldman, V. Sherman et G.M. Fried : Using simulators to assess laparoscopic competence : ready for widespread use ? Surgery, 135(1):28, 2004. 48

Germain Forestier, Florent Lalys, Laurent Riffaud, Brivael Trelhu et Pierre Jannin : Assessment of surgical skills using surgical processes and dynamic time warping. septembre 2011. 56, 58

Germain Forestier, Florent Lalys, Laurent Riffaud, Brivael Trelhu et Pierre Jannin : Classification of surgical processes using dynamic time warping. Journal of Biomedical Informatics, 45(2):255–264, 2012. 56

NK Francis et GB Hanna : The performance of master surgeons on the advanced dundee endoscopic psychomotor tester : Contrast validity study. Archives of Surgery, 137(7):841–844, juillet 2002. 48

S.A. Fraser, D.R. Klassen, L.S. Feldman, G.A. Ghitulescu, D. Stanbridge et G.M. Fried : Evaluating laparoscopic skills. Surgical Endoscopy, 17(6):964–967, 2003. 10.1007/s00464-002-8828-4. 47, 48

G. M. Fried, A. M. Derossis, J. Bothwell et H. H. Sigman : Comparison of laparoscopic performance in vivo with performance measured in a laparoscopic simulator. Surgical Endoscopy, 13(11):1077–1081, novembre 1999. 48

Gerald M. Fried, Liane S. Feldman, Melina C. Vassiliou, Shannon A. Fraser, Donna Stanbridge, Gabriela Ghitulescu et Christopher G. Andrew : Proving the value of simulation in laparoscopic surgery. Annals of Surgery, 240(3):518–528, septembre 2004. PMID : 15319723 PMCID : PMC1356442. 47, 48

Adam C Frischknecht, Steven J Kasten, Stanley J Hamstra, Noel C Perkins, R Brent Gillespie, Thomas J Armstrong et Rebecca M Minter : The objective assessment of experts' and novices' suturing skills using an image analysis program. Academic medicine : journal of the Association of American Medical Colleges, 88(2):260–264, février 2013. PMID : 23269303. 48

A G Gallagher, K Richie, N McClure et J McGuigan : Objective psychomotor skills assessment of experienced, junior, and novice laparoscopists with virtual reality. World journal of surgery, 25(11):1478–1483, novembre 2001. PMID : 11760752. 52, 55

A. G. Gallagher et R. M. Satava : Virtual reality as a metric for the assessment of laparoscopic psychomotor skills. Surgical Endoscopy And Other Interventional Techniques, 16(12):1746–1752, décembre

2002. 49, 50, 52, 55

R Gallagher, M Donnelly, P M Scalzi et M Deighton : Toward a comprehensive methodology for resident evaluation. Annual Conference on Research in Medical Education. Conference on Research in Medical Education, 16:12–16, 1977. PMID : 606064. 40

Vinod Gauba, Peter Tsangaris, Charalambos Tossounis, Arijit Mitra, Chris McLean et George M Saleh : Human reliability analysis of cataract surgery. Archives of ophthalmology, 126(2):173–177, février 2008. PMID : 18268206. 45

Barbara A. Goff, Peter E. Nielsen, Gretchen M. Lentz, Greg E. Chow, Robert W. Chalmers, Dee Fenner et Lynn S. Mandel : Surgical skills assessment : A blinded examination of obstetrics and gynecology residents. American Journal of Obstetrics and Gynecology, 186(4):613–617, 2002. 41

Andrew A. Gumbs, Nancy J. Hogle et Dennis L. Fowler : Evaluation of resident laparoscopic performance using global operative assessment of laparoscopic skills. Journal of the American College of Surgeons, 204(2):308–313, février 2007. 44

Julian Hance, Rajesh Aggarwal, Rex Stanbridge, Christopher Blauth, Yaron Munz, Ara Darzi et John Pepper : Objective assessment of technical skills in cardiac surgery. European Journal of Cardio-Thoracic Surgery, 28(1):157–162, juillet 2005. 41

G. B. Hanna, T. Drew, P. Clinch, B. Hunter et A. Cuschieri : Computer-controlled endoscopic performance assessment system. Surgical Endoscopy, 12(7):997–1000, juillet 1998. 48

Amin S Herati, Sylvia Montag, Sero Andonian, Edan Y Shapiro, Meredith Akerman, Louis R Kavoussi et Lee Richstone : Audio and visual analysis of urologic laparoscopic and robotic skills : objective criteria for surgical skill evaluation. Urology, 80(6):1277–1282, décembre 2012. PMID : 23206774. 41, 43

Ellen Hiemstra, Wendela Kolkman, Ron Wolterbeek, Baptist Trimbos et Frank Willem Jansen : Value of an objective assessment tool in the operating room. Can J Surg, 54(2):116–122, avril 2011. 41

A. James, D. Vieira, B. Lo, A. Darzi et G.-Z. Yang : Eye-Gaze driven surgical workflow segmentation. In Nicholas Ayache, Sébastien Ourselin et Anthony Maeder, éditeurs : Medical Image Computing and Computer-Assisted Intervention – MICCAI 2007, numéro 4792 in Lecture Notes in Computer Science, pages 110–117. Springer Berlin Heidelberg, janvier 2007. 37

P Jannin et X Morandi : Surgical models for computer-assisted neurosurgery. NeuroImage, 37(3):783–791, septembre 2007. PMID : 17613249. 34

Jagadeesan Jayender, Inbar Spofford, Balazs I. Lengyel, Christopher C. Thompson et Kirby G. Vosburgh : Markov modeling of colonoscopy gestures to develop skill trainers. In Cristian A. Linte, John T. Moore, Elvis C. S. Chen et David R. Holmes III, éditeurs : Augmented Environments for Computer-Assisted Interventions, volume 7264 de Lecture Notes in Computer Science, pages 65–73. Springer Berlin Heidelberg, janvier 2012. 57, 58

P Joice, G.B Hanna et A Cuschieri : Errors enacted during endoscopic surgery—a human reliability analysis. Applied Ergonomics, 29(6):409–414, décembre 1998. 45, 46

Timothy N. Judkins, Dmitry Oleynikov et Nick Stergiou : Objective evaluation of expert and novice performance during robotic surgical training tasks. Surgical Endoscopy, 23(3):590–597, mars 2009. 50, 51, 53, 54

Seong-Young Ko, Jonathan Kim, Woo-Jung Lee et Dong-Soo Kwon : Surgery task model for intelligent interaction between surgeon and laparoscopic assistant robot. International Journal of Assitive Robotics and Mechatronics, 8(1):38–46, 2007. 37

J A Kopta : An approach to the evaluation of operative skills. Surgery, 70(2):297–303, août 1971. PMID : 5560189. 41

Yo Kurashima, Liane S. Feldman, Salman Al-Sabah, Pepa A. Kaneva, Gerald M. Fried et Melina C. Vassiliou : A tool for training and evaluation of laparoscopic inguinal hernia repair : the global operative assessment of laparoscopic Skills-Groin hernia (GOALS-GH). The American Journal of Surgery, 201 (1):54–61, janvier 2011. 44

Florent Lalys : Automatic recognition of low-level and high-level surgical tasks in the Operating Room from video images. Thèse de doctorat, Université Rennes 1, mai 2012. URL http://tel.archives-ouvertes.fr/tel-00695648. 35, 37

Florent Lalys, David Bouget, Laurent Riffaud et Pierre Jannin : Automatic knowledge-based recognition of low-level tasks in ophthalmological procedures. International Journal of Computer Assisted Radiology and Surgery, 8(1):39–49, janvier 2013. 37

Florent Lalys, Laurent Riffaud, David Bouget et Pierre Jannin : An Application-Dependent framework

for the recognition of High-Level surgical tasks in the OR. In Gabor Fichtinger, Anne Martel et Terry Peters, éditeurs : Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011, Lecture Notes in Computer Science, pages 331–338. Springer Berlin Heidelberg, janvier 2011a. 37

Florent Lalys, Laurent Riffaud, Xavier Morandi et Pierre Jannin : Surgical phases detection from microscope videos by combining SVM and HMM. In Bjoern Menze, Georg Langs, Zhuowen Tu et Antonio Criminisi, éditeurs : Medical Computer Vision. Recognition Techniques and Applications in Medical Imaging, Lecture Notes in Computer Science, pages 54–62. Springer Berlin Heidelberg, janvier 2011b. 37

J. J. H. Leong, D. R. Leff, A. Das, R. Aggarwal, P. Reilly, H. D. E. Atkinson, R. J. Emery et A. W. Darzi : Validation of orthopaedic bench models for trauma surgery. Journal of Bone & Joint Surgery, British Volume, 90-B(7):958–965, juillet 2008. 41

Julian J. H. Leong, Marios Nicolaou, Louis Atallah, George P. Mylonas, Ara W. Darzi et Guang-Zhong Yang : HMM assessment of quality of movement trajectory in laparoscopic surgery. In Rasmus Larsen, Mads Nielsen et Jon Sporring, éditeurs : Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006, Lecture Notes in Computer Science, pages 752–759. Springer Berlin Heidelberg, janvier 2006. 56, 57, 58

Henry C Lin, Izhak Shafran, David Yuh et Gregory D Hager : Towards automatic skill evaluation : detection and segmentation of robot-assisted surgical motions. Computer aided surgery : official journal of the International Society for Computer Aided Surgery, 11(5):220–230, septembre 2006. PMID : 17127647. 37

Z Lin, M Uemura, M Zecca, S Sessa, H Ishii, M Tomikawa, M Hashizume et A Takanishi : Objective skill evaluation for laparoscopic training based on motion analysis. IEEE transactions on bio-medical engineering, novembre 2012. PMID : 23204271. 51, 54

Benny P. L. Lo, Ara Darzi et Guang-Zhong Yang : Episode classification for the analysis of Tissue/Instrument interaction with multiple visual cues. In Randy E. Ellis et Terry M. Peters, éditeurs : Medical Image Computing and Computer-Assisted Intervention - MICCAI 2003, numéro 2878 in Lecture Notes in Computer Science, pages 230–237. Springer Berlin Heidelberg, janvier 2003. 37

Calvin Lyons, David Goldfarb, Stephen L Jones, Niraj Badhiwala, Brian Miles, Richard Link et Brian J Dunkin : Which skills really matter ? proving face, content, and construct validity for a commercial robotic simulator. Surgical endoscopy, février 2013. PMID : 23389060. 49, 50, 51, 53, 54, 55

C. L. MacKenzie, J. A. Ibbotson, C. G. L. Cao et A. J. Lomax : Hierarchical decomposition of laparoscopic surgery : a human factors approach to investigating the operating room environment. Minimally Invasive Therapy and Allied Technologies, 10(3):121–127, 2001. 35

Alasdair I.M Macmillan et Alfred Cuschieri : Assessment of innate ability and skills for endoscopic manipulations by the advanced dundee endoscopic psychomotor tester : predictive and concurrent validity. The American Journal of Surgery, 177(3):274–277, mars 1999. 48

JA Martin, G. Regehr, R. Reznick, H. MacRae, J. Murnaghan, C. Hutchison et M. Brown : Objective structured assessment of technical skill (OSATS) for surgical residents. British Journal of Surgery, 84 (2):273–278, 1997. 40, 41, 42

John D Mason, James Ansell, Neil Warren et Jared Torkington : Is motion analysis a valid tool for assessing laparoscopic skill ? Surgical endoscopy, décembre 2012. PMID : 23233011. 49

AL McCluney, MC Vassiliou, PA Kaneva, J. Cao, DD Stanbridge, LS Feldman et GM Fried : FLS simulator performance predicts intraoperative laparoscopic skill. Surgical endoscopy, 21(11):1991–1995, 2007. 48

G. Megali, S. Sinigaglia, O. Tonet et P. Dario : Modelling and evaluation of surgical performance using hidden markov models. Biomedical Engineering, IEEE Transactions on, 53(10):1911–1919, octobre 2006. 57, 58

K. Moorthy, Y. Munz, A. Dosis, F. Bello, A. Chang et A. Darzi : Bimodal assessment of laparoscopic suturing skills. Surgical endoscopy, 18(11):1608–1612, 2004. 49, 50, 52, 54

Krishna Moorthy, Yaron Munz, Sally Adams, Vikas Pandey et Ara Darzi : Self-assessment of performance among surgical trainees during simulated procedures in a simulated operating theater. The American Journal of Surgery, 192(1):114–118, 2006. 43

Yaron Munz, Krishna Moorthy, Simon Bann, Jyoti Shah, Sneizana Ivanova et Sir Ara Darzi : Ceiling effect in technical skills of surgical residents. The American Journal of Surgery, 188(3):294–300, septembre 2004. 41

Todd E. Murphy, Chet M. Vignes, David D. Yuh et Allison M. Okamura : Automatic motion recognition and skill evaluation for dynamic tasks. Proceedings Eurohaptics, pages 363–373, 2003. 56, 57, 58

T. Neumuth, M. Czygan, D. Goldstein, G. Strauss, J. Meixensberger et O. Burgert : Computer assisted acquisition of surgical process models with a sensors-driven ontology. In M2CAI workshop, MICCAI, London, volume 44, 2009. 37

Thomas Neumuth, Gero Strauss, Jürgen Meixensberger, Heinz U. Lemke et Oliver Burgert : Acquisition of process descriptions from surgical interventions. In Stéphane Bressan, Josef Küng et Roland Wagner, éditeurs : Database and Expert Systems Applications, Lecture Notes in Computer Science, pages 602–611. Springer Berlin Heidelberg, janvier 2006. 34

J. Adam Oostema, Matthew P. Abdel et Jon C. Gould : Time-efficient laparoscopic skills assessment using an augmented-reality simulator. Surgical Endoscopy, 22(12):2621–2624, décembre 2008. 50, 52, 54

Ignacio Oropesa, Patricia Sánchez-González, Magdalena K Chmarra, Pablo Lamata, Alvaro Fernández, Juan A Sánchez-Margallo, Frank Willem Jansen, Jenny Dankelman, Francisco M Sánchez-Margallo et Enrique J Gómez : EVA : laparoscopic instrument tracking based on endoscopic video analysis for psychomotor skills assessment. Surgical endoscopy, 27(3):1029–1039, mars 2013. PMID : 23052495. 50, 51, 53, 54

N. Padoy, T. Blum, I. Essa, Hubertus Feussner, M.-O. Berger et Nassir Navab : A boosted segmentation method for surgical workflow analysis. In Nicholas Ayache, Sébastien Ourselin et Anthony Maeder, éditeurs : Medical Image Computing and Computer-Assisted Intervention – MICCAI 2007, numéro 4791 in Lecture Notes in Computer Science, pages 102–109. Springer Berlin Heidelberg, janvier 2007. 37

Nicolas Padoy, Tobias Blum, Seyed-Ahmad Ahmadi, Hubertus Feussner, Marie-Odile Berger et Nassir Navab : Statistical modeling and recognition of surgical workflow. Medical Image Analysis, 16(3):632–641, 2012. 11, 36, 37

Nicolas Padoy, Blum Tobias, Hubertus Feussner, Marie-Odile Berger et Nassir Navab : On-line recognition of surgical activity for monitoring in the operating room. pages 1718–1724, 2008. 37

V. A. Pandey, J. H. N. Wolfe, C. D. Liapis et D. Bergqvist : The examination assessment of technical competence in vascular surgery. British Journal of Surgery, 93(9):1132–1138, 2006. 41

VA Pandey, JHN Wolfe, SA Black, M Cairols, CD Liapis et D Bergqvist : Self-Assessment of technical skill in surgery : The need for expert feedback. Annals of The Royal College of Surgeons of England, 90 (4):286–290, mai 2008. PMID : 18492390 PMCID : PMC2647188. 43

M. Pellen, L. Horgan, J. Roger Barton et S. Attwood : Laparoscopic surgical skills assessment : can simulators replace experts ? World journal of surgery, 33(3):440–447, 2009a. 50, 52, 54

Michael G. C. Pellen, Liam F. Horgan, J. Roger Barton et Stephen E. Attwood : Construct validity of the ProMIS laparoscopic simulator. Surgical Endoscopy, 23(1):130–139, janvier 2009b. 50, 52, 54

L.R. Rabiner : A tutorial on hidden markov models and selected applications in speech recognition. Proceedings of the IEEE, 77(2):257 –286, 1989. 37

Richard Reznick, Glenn Regehr, Helen MacRae, Jenepher Martin et Wendy McCulloch : Testing technical skill via an innovative "bench station" examination. The American Journal of Surgery, 173 (3):226–230, mars 1997. 40

Richard K. Reznick et Helen MacRae : Teaching surgical skills – changes in the wind. The New England Journal of Medicine, 355(25):2664, décembre 2006. 41

J. Rosen, B. Hannaford, C.G. Richards et M.N. Sinanan : Markov modeling of minimally invasive surgery based on tool/tissue interaction and force/torque signatures for evaluating surgical skills. IEEE Transactions on Biomedical Engineering, 48(5):579 –591, mai 2001a. 56, 57, 58

J Rosen, M MacFarlane, C Richards, B Hannaford et M Sinanan : Surgeon-tool force/torque signatures–evaluation of surgical skills in minimally invasive surgery. Studies in health technology and informatics, 62:290–296, 1999. PMID : 10538374. 56, 58

J Rosen, M Solazzo, B Hannaford et M Sinanan : Objective laparoscopic skills assessments of surgical residents using hidden markov models based on haptic information and tool/tissue interactions. Studies in health technology and informatics, 81:417–423, 2001b. PMID : 11317782. 56, 57, 58

Jacob Rosen, Jeffrey D Brown, Lily Chang, Mika N Sinanan et Blake Hannaford : Generalized approach for modeling minimally invasive surgery as a stochastic process using a discrete markov model. IEEE Trans Biomed Eng, 53(3):399–413, mars 2006. 56, 57, 58

Jacob Rosen, Massimiliano Solazzo, Blake Hannaford et Mika Sinanan : Task decomposition of laparoscopic surgery for objective evaluation of surgical residents' learning curve using hidden markov model.

Computer Aided Surgery, 7(1):49–61, 2002. 56, 57, 58

Hiroaki Sakoe et Seibi Chiba : A dynamic programming approach to continuous speech recognition. In Proceedings of the Seventh International Congress on Acoustics, volume 3, page 65–69, 1971. 36

Shah-Jalal Sarker, Muwaffaq Mezeil Telfah, Louisa Onuba et Bijendra P Patel : Objective assessment of skills acquisition during laparoscopic surgery courses. Surgical innovation, décembre 2012. PMID : 23242519. 52, 55

Daniel J. Scott, Robert V. Rege, Patricia C. Bergen, Weidun A. Guo, Royce Laycock, Seifu T. Tesfay, R. James Valentine et Daniel B. Jones : Measuring operative performance after laparoscopic skills training : Edited videotape versus direct observation. Journal of Laparoendoscopic & Advanced Surgical Techniques, 10(4):183–190, août 2000. 43

Ravi S. Sidhu, Elena Vikis, Rona Cheifetz et Terry Phang : Self-assessment during a 2-day laparoscopic colectomy course : can surgeons judge how well they are learning new skills ? The American Journal of Surgery, 191(5):677–681, mai 2006. 43, 44

S. G. T. Smith, J. Torkington, T. J. Brown, N. J. Taffinder et A. Darzi : Motion analysis : a tool for assessing laparoscopic dexterity in the performance of a laboratory-based laparsocopic cholecystectomy. Surgical Endoscopy And Other Interventional Techniques, 16(4):640–645, avril 2002. 50, 52, 54

Lee L. Swanstrom, Gerald M. Fried, Kaaren I. Hoffman et Nathaniel J. Soper : Beta test results of a new system assessing competence in laparoscopic surgery. Journal of the American College of Surgeons, 202(1):62–69, janvier 2006. 47

Steven E. Swift et James F. Carter : Institution and validation of an observed structured assessment of technical skills (OSATS) for obstetrics and gynecology residents and faculty. American Journal of Obstetrics and Gynecology, 195(2):617–621, 2006. 41

David Szalay, Helen MacRae, Glenn Regehr et Richard Reznick : Using operative outcome to assess technical skill. The American Journal of Surgery, 180(3):234–237, septembre 2000. 43

N Taffinder, C Sutton, R J Fishwick, I C McManus et A Darzi : Validation of virtual reality to teach and assess psychomotor skills in laparoscopic surgery : results from randomised controlled studies using the MIST VR laparoscopic simulator. Studies in health technology and informatics, 50:124–130, 1998. PMID : 10180527. 50, 52, 55

B. Tang, G. B. Hanna, N. M. A. Bax et A. Cuschieri : Analysis of technical surgical errors during initial experience of laparoscopic pyloromyotomy by a group of dutch pediatric surgeons. Surgical Endoscopy And Other Interventional Techniques, 18(12):1716–1720, décembre 2004. 45

B. Tang, G. B. Hanna, F. Carter, G. D. Adamson, J. P. Martindale et A. Cuschieri : Competence assessment of laparoscopic operative and cognitive skills : Objective structured clinical examination (OSCE) or observational clinical human reliability assessment (OCHRA). World Journal of Surgery, 30 (4):527–534, mars 2006. 45

K. Tanoue, M. Uemura, H. Kenmotsu, S. Ieiri, K. Konishi, K. Ohuchida, M. Onimaru, Y. Nagao, R. Kumashiro, M. Tomikawa et al. : Skills assessment using a virtual reality simulator, LapSim™, after training to develop fundamental skills for endoscopic surgery. Minimally Invasive Therapy & Allied Technologies, 19(1):24–29, 2010. 49, 52, 55

Masakazu Tokunaga, Hiroyuki Egi, Minoru Hattori, Masanori Yoshimitsu, Daisuke Sumitani, Tomohiro Kawahara, Masazumi Okajima et Hideki Ohdan : Approaching time is important for assessment of endoscopic surgical skills. Minimally Invasive Therapy & Allied Technologies, 21(3):142–149, mai 2012. 49, 53, 54

Brivaël Trelhu : Modélisation, analyse et comparaison de procédures chirrugicales, application à la neurochirurgie. PhD thesis, Université de Rennes 1, 2011. 34

Jeffrey Turnbull, Jean Gray et John MacFadyen : Improving In-Training evaluation programs. Journal of General Internal Medicine, 13(5):317–323, mai 1998. PMID : 9613887 PMCID : PMC1496958. 38

C. P. M. van der Vleuten et David B. Swanson : Assessment of clinical skills with standardized patients : State of the art. Teaching and Learning in Medicine, 2(2):58–76, 1990. 41

K. R. Van Sickle, D. A. McClusky Iii, A. G. Gallagher et C. D. Smith : Construct validation of the ProMIS simulator using a novel laparoscopic suturing task. Surgical Endoscopy And Other Interventional Techniques, 19(9):1227–1231, septembre 2005. 49, 50, 52, 54

M. Vassiliou, G. Ghitulescu, L. Feldman, D. Stanbridge, K. Leffondré, H. Sigman et G. Fried : The MISTELS program to measure technical skill in laparoscopic surgery. Surgical Endoscopy, 20(5):744–747, 2006. 10.1007/s00464-005-3008-y. 48

M.C. Vassiliou, L.S. Feldman, C.G. Andrew, S. Bergman, K. Leffondré, D. Stanbridge et G.M. Fried : A global assessment tool for evaluation of intraoperative laparoscopic skills. The American Journal of Surgery, 190(1):107–113, 2005. 43, 44

K.R. Wanzel, M. Ward et R.K. Reznick : Teaching the surgical craft : from selection to certification. Current problems in surgery, 39(6):573–660, 2002. 39, 40, 55

George Xeroulis, Adam Dubrowski et Ken Leslie : Simulation in laparoscopic surgery : a concurrent validity study for FLS. Surgical Endoscopy, 23(1):161–165, janvier 2009. 50, 52, 54

Yasushi Yamauchi, Juli Yamashita, Osamu Morikawa, Ryoichi Hashimoto, Masaaki Mochimaru, Yukio Fukui, Hiroshi Uno et Kazunori Yokoyama : Surgical skill evaluation by force data for endoscopic sinus surgery training system. In Takeyoshi Dohi et Ron Kikinis, éditeurs : Medical Image Computing and Computer-Assisted Intervention — MICCAI 2002, numéro 2488 in Lecture Notes in Computer Science, pages 44–51. Springer Berlin Heidelberg, janvier 2002. 51, 53, 55

Deuxième partie

Développement d'une méthode de suivi des instruments chirurgicaux à partir d'images endoscopiques

CHAPITRE 3

DÉTECTION DES INSTRUMENTS CHIRURGICAUX À PARTIR DE MODÈLES STATISTIQUES ET GÉOMÉTRIQUES DANS DES IMAGES ENDOSCOPIQUES

Sommaire

3.1 Suivi des instruments chirurgicaux : méthodes existantes . 72

3.1.1 Quelles sont les propriétés d'une méthode de mesure idéale ? . 72

3.1.2 Méthodes basées sur des informations non-visuelles 72

3.1.3 Méthodes basées sur le traitement d'images endoscopiques . . 78

3.2 Approche développée 81

3.2.1 Cadre de travail général 81

3.2.2 Calibrage de la caméra 83

3.2.3 Représentation géométrique de l'instrument 85

3.2.4 Traitement de l'image 88

3.2.5 Détermination de la position du point d'insertion dans le repère lié à la caméra 105

3.2.6 Représentation spatiale de l'orientation de l'instrument 109

3.2.7 Algorithme de CONDENSATION 114

3.2.8 Détermination profondeur d'insertion 119

3.2.9 Améliorations apportées à la méthode 124

3.3 Résultats bruts 124

3.1 Suivi des instruments chirurgicaux : un bref aperçu des méthodes existantes

Je m'intéresse dans cette partie de ma thèse à la détermination de la trajectoire des instruments de chirurgie laparoscopique au cours d'une intervention chirurgicale. Cette trajectoire peut-être décrite soit par la position de la pointe de l'instrument, soit par le couple {position de la pointe, direction de l'axe} de l'instrument, en deux dimensions (dans l'image endoscopique) ou en trois dimensions.

De nombreux travaux se sont intéressés à cette problématique, on peut en trouver une description détaillée mais incomplète dans la revue de la littérature proposée par Chmarra et al. [2007]. Je présente brièvement dans ce chapitre les méthodes disponibles et expose leurs limites pour l'usage recherché dans le cadre de cette thèse avant de présenter la méthode développée pour répondre à mes besoins. Les trajectoires ainsi obtenues doivent être filtrées pour permettre l'évaluation de la précision de la méthode développée ; ce sera l'objet du chapitre suivant.

3.1.1 Quelles sont les propriétés d'une méthode de mesure idéale ?

De la même manière que j'ai défini dans la section 2.2.1 les propriétés idéales des méthodes d'évaluation, je définis ici les propriétés que devrait satisfaire une méthode d'acquisition des données idéale :

▷ **La fiabilité** La mesure des données doit être reproductible, et être robuste aux variations de l'environnement dans lequel son utilisation est envisagée.

▷ **La précision** La méthode de mesure des données doit être capable de séparer deux valeurs voisines dont la signification diffère.

▷ **L'acceptabilité et la portabilité** Le dispositif de mesure ne doit pas interférer avec le déroulé habituel de la procédure chirurgicale et ne doit pas modifier la façon dont le chirurgien réalise son geste. Il doit être utilisable dans différents contextes et sites, et satisfaire aux contraintes d'espace et d'hygiène du bloc opératoire. Son coût financier et humain doit être limité.

▷ **L'acquisition en temps réel** La mesure doit pouvoir être effectuée en temps réel afin de retourner les informations pertinentes au chirurgien lors de son geste pour lui permettre de le modifier si nécessaire.

3.1.2 Méthodes basées sur des informations non-visuelles

Ces méthodes sont basées sur des dispositifs permettant d'extraire la position de la pointe de l'instrument, ou des mains du chirurgien, indépendamment de l'image

endoscopique. Certaines ont un contexte d'utilisation limité (robots, simulateurs), d'autres peuvent s'adapter à différentes situations (localisateurs).

i) Données de cinématique des robots chirurgicaux

Lorsque l'accès à l'interface de programmation de l'application (API) contrôlant le robot chirurgical daVinci est possible (accord de recherche entre le fabricant et le laboratoire), les données de cinématique des commandes et des effecteurs peuvent être récupérées par l'utilisateur pour procéder à une analyse du mouvement. Cependant le calcul de la position de la pointe à partir des commandes passées par l'utilisateur est approximatif en raison d'incertitudes au niveau des articulations du robot (chaîne comprenant 18 articulations pour plus de deux mètres de longueur effective). Reiter et al. [2012] mesurent une erreur pouvant dépasser 2cm entre la position retournée par l'API et la position réelle de la pointe. Cette erreur est tolérable pour l'usage habituel du robot car le chirurgien intervient dans la boucle de rétro-contrôle, adaptant la commande qu'il passe à la position actuelle de l'effecteur, mais limite l'utilisation de ces données pour l'analyse de la qualité du geste (en particulier si on cherche à mesurer des distances entre l'instrument et des structures anatomiques).

Ce dispositif, bien que pratique (accès direct à de nombreuses informations), présente cependant plusieurs limites : sa disponibilité est limitée aux centres disposant d'un robot chirurgical dédié aux applications de recherche bénéficiant d'un accord d'utilisation avec Intuitive Surgical (je n'ai pas trouvé de littérature rapportant l'utilisation de l'API dans le contexte du bloc opératoire), et de nombreuses interventions de laparoscopie se déroulent sans robot, car le temps de mise en place de ce dernier est trop important par rapport à la durée totale de l'intervention. Enfin les courbes d'apprentissage de la chirurgie robotisée et de la laparoscopie classique ne sont pas forcément comparables car les compétences requises ne sont pas identiques. De plus le caractère récent de la chirurgie robotisée ne permet pas de définir pour le moment une courbe d'apprentissage consensuelle, les critères de définition de cette dernière variant fortement d'une publication à l'autre (la durée de la phase d'apprentissage varie ainsi de 8 interventions [Ahlering et al., 2003] à 250 interventions [Herrell et Smith, 2005] selon les critères retenus). Ces différences ne permettent pas de pronostiquer la qualité d'un geste en laparoscopie classique à partir de métriques définies en laparoscopie robotisée.

ii) Données des simulateurs de réalité virtuelle

Les simulateurs de réalité virtuelle permettent également d'accéder facilement aux données de position de la pointe de l'instrument ainsi qu'à son axe puisque ces données sont nécessaires à la reconstruction de l'image de l'instrument dans l'environnement virtuel. La position des instruments est généralement acquise par un système mécanique de type pivot avant d'être envoyée vers l'application de si-

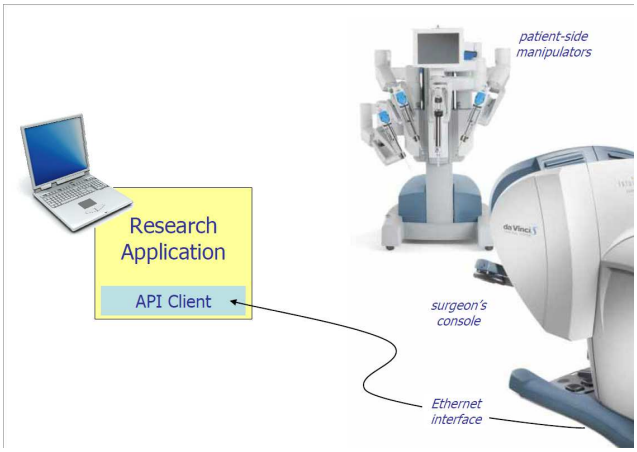


FIGURE 3.1. – Robot chirurgical daVinci avec API, extrait de DiMaio et Hasser [2008]

mulation. On peut citer parmi ces interfaces celles proposées par Immersion Inc. (*Laparoscopic Surgical Workstation*, *Virtual Laparoscopic Interface* (VLI), *Laparoscopic Impulse Engine* (LIE), figure 3.2).

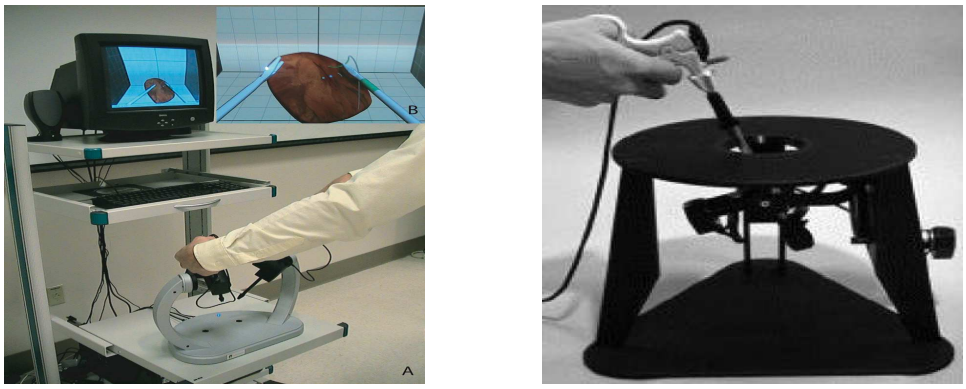
L'étude du geste chirurgical sur simulateur de réalité virtuelle a cependant plusieurs limitations :

- l'environnement modélisé par la simulation est moins réaliste que l'environnement que le chirurgien rencontrera au bloc opératoire ;
- le retour haptique est soit inexistant (VLI), soit beaucoup moins riche que dans la réalité (pas de modélisation biomécanique fine des organes pour les simulations en temps réel, par exemple pour le LIE ou le LapSim [Acosta et Temkin, 2005]) ;
- la portabilité du système d'acquisition des données vers le bloc opératoire est nulle ;
- le nombre de scenarii modélisés est faible, avec une validité apparente souvent limitée (modélisation de tâches au mieux, de sous-tâches peu réalistes le plus souvent)

iii) Localisateurs optiques, magnétiques, acoustiques ou mécaniques

L'utilisation d'instruments laparoscopiques réels permet d'analyser le geste dans des conditions plus proches de la réalité, en particulier grâce au retour haptique qui permet au chirurgien d'adapter son geste à son environnement. Plusieurs dispotifs peuvent être ajoutés aux instruments du chirurgien afin de suivre leur position dans le temps, dans différents contextes (bloc opératoire, banc d'entraînement).

Marqueurs optiques : il s'agit de faire communiquer entre eux une ou plusieurs caméras et un ensemble de marqueurs, actifs (diodes électroluminescentes) ou



(a) Virtual Laparoscopic Interface (b) Laparoscopic Impulse Engine

FIGURE 3.2. – Interfaces d'acquisition de la position des instruments pour les simulateurs de réalité virtuelle (Immersion Inc.)

passifs (réflecteurs de lumière infra-rouge émise par la caméra), de géométrie connue, pour connaître la position de l'un dans le référentiel de l'autre. La position des marqueurs dans l'image de la caméra permet de calculer leur position en trois dimensions. En général les caméras (rigidement liées entre elles) servent de "repère monde" tandis que les marqueurs sont fixées rigidement aux instruments, une phase de calibrage (méthode du pivot par exemple [Lavallée et al., 1997]) permettant de connaître la position de la pointe connaissant la position du repère formé par les marqueurs. Cette méthode de suivi est précise [Chassat et Ph.D, 1998] si le calibrage est de bonne qualité (difficile en pratique), mais sensible aux occultations : la ligne de vue entre les caméras et les marqueurs ne doit pas être interrompue, par un membre de l'équipe chirurgicale ou par une orientation spécifique de l'instrument, sous peine de ne plus pouvoir acquérir la position de celui-ci. Cette condition est difficilement réalisable au bloc opératoire sans modifier l'organisation habituelle de l'équipe, dont l'ergonomie est garante de conditions optimales pour le déroulement de l'intervention chirurgicale. Cette méthode reste cependant largement utilisée au laboratoire, comme étalon pour évaluer d'autres méthodes, grâce à sa disponibilité, sa précision et sa facilité d'utilisation, ainsi qu'au bloc opératoire pour certaines applications spécifiques, en orthopédie et neuro-chirurgie (figure 3.3a).

Marqueurs acoustiques : cette méthode est basée sur le repérage d'un émetteur d'ultrasons haute-fréquence dans un repère formé par des récepteurs de géométrie connue (principe similaire aux marqueurs optiques). La position est calculée par temps de vol ou par décalage de phase. Cette méthode est dépendante de la vitesse du son dans l'air, elle-même dépendante de la température et des mouvements de masses d'air dans la salle (engendrés par des déplacements de personnel). Ces dépendances peuvent conduire à des erreurs de me-

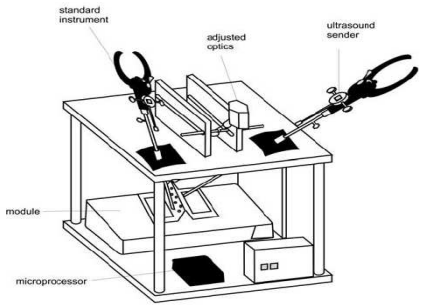
sure [Wehn et Belanger, 1997] si les récepteurs sont placés loin des émetteurs. Elle est également sensible aux réflexions parasites et au phénomène d'occultation décrit pour les marqueurs optiques. Ces contraintes ne permettent pas actuellement de se servir de ce type de dispositif au bloc opératoire, mais plusieurs systèmes ont été développés pour un usage de recherche [Tatar et al., 2002, Sokollik et al., 2004] (figure 3.3b).

Marqueurs électromagnétiques : on détecte ici le mouvement d'un récepteur constitué de trois bobines formant un repère orthogonal dans un champ électromagnétique basses-fréquences généré par un émetteur lui-même constitué de 3 bobines. Le dispositif ICSAD utilise cette technologie (IsoTrak II, Polhemus) pour suivre les mains du chirurgien ; Cristancho et al. [2009] ont adapté des marqueurs du même type (Fastrak, Polhemus) pour les fixer sur des instruments de laparoscopie et extraire la position de la pointe grâce à une phase de calibrage (figure 3.3c). Ce type de marqueurs n'est pas sensible aux occultations mais la qualité des mesures peut être altérée par la présence de matériel ferromagnétique dans le voisinage de l'émetteur ou des récepteurs (Polhemus recommande de ne pas avoir de matériel ferromagnétique à une distance inférieure à trois fois la distance émetteur/récepteur), ce qui peut être problématique dans le contexte du bloc opératoire. De plus, ce type de dispositif n'est pas certifié pour une utilisation médicale.

Dispositifs mécaniques : je regroupe sous cette dénomination l'ensemble des dispositifs utilisant la propriété de pivot du point d'insertion pour déterminer l'orientation de l'outil en trois dimensions. En général, une structure comprenant plusieurs bagues articulées entre elles mime le point d'insertion et sert de référence dans l'espace, différents types de capteurs permettant ensuite de calculer l'orientation de ces bagues les unes par rapport aux autres, la translation de l'instrument le long de l'axe et parfois la rotation de l'outil autour de son propre axe. De nombreux dispositifs de ce type ont été développés pour analyser la qualité des gestes en chirurgie : CELTS (*Computer Enhanced Laparoscopic Training System*, version modifiée du VLI décrit plus haut) [Cotin et al., 2002], ADEPT [Macmillan et Cuschieri, 1999], TrEndo [Chmarra et al., 2006] entre autres, le dernier cité ayant l'intérêt d'être compact et utilisable sur n'importe quel type de banc d'entraînement contrairement aux deux premiers. Le dispositif BlueDRAGON proposé par Rosen et al. [2006] est également utilisable sur tout type de banc d'entraînement, mais est volumineux donc peu transportable (figure 3.3d). D'une façon générale, ces dispositifs ont un encombrement trop important pour pouvoir être utilisés au bloc opératoire.

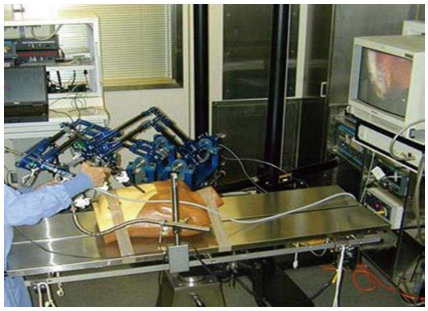
Toutes ces méthodes nécessitent des ajouts par rapport à la configuration habituelle du bloc opératoire : elles ne sont pas aisément portables et ont une acceptabilité modérée voire faible de la part des intervenants de la procédure chirurgicale (contraintes d'encombrement dans un environnement où l'ergonomie est capitale, ajouts de fils pour permettre la communication entre capteurs et ordinateurs, pro-

(a) Station de suivi par marqueurs optiques Polaris



(b) Détection de la position d'un instrument par ultrasons [Sokollik et al., 2004]

(c) Marqueur électromagnétique pour le suivi d'instruments laparoscopiques [Cristancho et al., 2009]



(d) Dispositif mécanique BlueDRAGON [Rosen et al., 2006]

FIGURE 3.3. – Dispositifs de suivi d'instruments laparoscopiques permettant une utilisation sur banc d'entraînement ou au bloc opératoire

blèmes de stérilisation...). Or il a été démontré que la simplicité de l'organisation d'un bloc opératoire permettait d'améliorer l'efficacité de l'équipe chirurgicale et la manipulation des outils [Reiley et al., 2011]. De plus certains dispositifs peuvent entraîner des modifications dans la façon même de réaliser le geste (problème d'occultation de la ligne de vue avec les marqueurs optiques ou acoustiques), perturbant le déroulement du processus chirurgical habituel que l'on souhaite analyser.

Je recherche une méthode nécessitant le minimum d'adaptation par rapport au matériel existant et n'interférant pas avec le déroulement habituel de la procédure chirurgicale que je cherche à analyser (hormis une phase de calibrage préalable), ni avec l'activité et l'ergonomie du bloc opératoire.

La détermination de la trajectoire par l'analyse d'images endoscopiques répond à cette contrainte, le flux vidéo étant aisément disponible et sa récupération ne nécessitant pas de matériel spécifique.

3.1.3 Méthodes basées sur le traitement d'images endoscopiques

Le domaine de la vision par ordinateur pour identifier des structures dans des images est trop large pour que l'ensemble de ses aspects soient abordés dans ce travail. D'une façon générale, les travaux de ce domaine se basent sur des indices de couleur [Gong et al., 1998], de texture (avec des descripteurs d'Haralick [Haralick et al., 1973] par exemple) ou de forme (approche par moments [Hu, 1962] par exemple) pour identifier des structures d'intérêt dans des images. Pour davantage de détails, on pourra se rapporter au livre de Cocquerez et Philipp-Foliguet [1995], qui présentent les méthodes les plus courantes.

Une première étape vers la détection de la position d'instruments à partir d'images endoscopiques consiste à modifier ces instruments pour y ajouter des indices (bague de couleur au niveau de la pointe par exemple) pour faciliter leur segmentation dans une image vidéo, puis à observer la scène sous plusieurs angles grâce à différentes caméras dont la position relative est connue pour trianguler la position de ces indices. Cette approche est utilisée par le simulateur chirurgical ProMIS par exemple [Van Sickle et al., 2005]. Même si l'encombrement est moindre que pour la plupart des méthodes décrites ci-dessus, l'ajout d'indices ne permet pas d'utiliser du matériel standard et stérilisable. L'ajout de caméras supplémentaires dans la cavité abdominale n'est pas impossible, et pourrait même apporter des informations pertinentes pour certains types de chirurgie [Boschet, 2010], mais leur repérage dans un référentiel commun n'a pas de solution pratique pour le moment.

Les méthodes basées sur l'image endoscopique sans modification du matériel chirurgical procèdent d'une façon générale en deux phases successives :

- localisation de la caméra en 3D dans le contexte du bloc opératoire (c'est-à-dire définir les paramètres de calibrage extrinsèque de la caméra), la majorité des études publiées utilisant une caméra fixe ;

– segmentation des instruments dans l'image endoscopique et reconstruction de leur position en trois dimensions dans le repère de la caméra à partir des paramètres de calibrage intrinsèque de cette dernière. Les organes peuvent également être segmentés et localisés au cours de cette phase.

i) Segmentation de l'instrument

Le processus de segmentation de l'instrument dans l'image endoscopique est une étape délicate, en raison de conditions d'illumination variables, de la possibilité d'avoir de la fumée dans la cavité abdominale (utilisation de pinces bipolaires pour cautériser) et des phénomènes d'occultation et de superposition des instruments.

La segmentation peut s'effectuer en identifiant les pixels formant le manche de l'instrument (de couleur sombre par rapport au reste de la scène en règle générale). Doignon et al. [2005] utilisent un espace colorimétrique combinant la saturation et la luminance pour retrouver les régions pouvant correspondre au manche de l'instrument (croissance de région à partir des bords de l'image, seuillage par l'algorithme de Otsu et classement des régions candidates par des descripteurs). Cette méthode permet de détecter les instruments de façon robuste, avec une vitesse égale à la moitié du débit vidéo. Cette méthode est secondairement utilisée dans [Doignon et al., 2007] pour retrouver la position de l'axe de l'instrument en trois dimensions et déterminer la position du point d'insertion. Allan et al. [2012] classent les pixels de l'image en s'appuyant sur des arbres de décision prenant comme variables un grand nombre de paramètres colorimétriques et de forme de l'image. Un modèle géométrique de l'instrument est utilisé pour identifier ce dernier dans l'image segmentée, fondé sur les dimensions de ses petit et grand axes. Cette approche nécessite un entraînement préalable de l'arbre de décision dans des conditions similaires à celles de la chirurgie. Speidel et al. [2006, 2008] adoptent une approche similaire en séparant les instruments et le fond grâce à un classifieur bayésien préalablement entraîné sur des images similaires à celles rencontrées au cours de la chirurgie, puis identifient l'axe de l'instrument par filtrage, croissance de régions, squelettisation et transformée de Hough [Duda et Hart, 1972]. La reconstruction de la position de l'instrument en trois dimensions est obtenue grâce à l'utilisation d'un endoscope permettant la stéréo-vision.

Plusieurs autres travaux recherchent l'instrument dans l'image en se basant sur les propriétés de ses bords. Climent et Marés [2004] extraient les orientations du gradient pour chaque pixel d'une image endoscopique avant d'appliquer une transformation de Hough pour identifier les bords de l'instrument. Cette méthode est très sensible aux fausses détections (lignes parasites dans l'image), les auteurs appliquent donc un filtre heuristique aux résultats, ce qui limite la portabilité de la méthode (conditions du filtre très spécifiques au dispositif). De plus cette méthode n'est pas robuste si plusieurs instruments sont présents dans l'image.

Voros et al. [2007] proposent de réduire l'espace d'exploration pour réaliser la transformée de Hough en considérant que les droites définissant les bords de l'ins-

trument passent nécessairement à proximité de la projection du point d'insertion dans le plan image. Les bords de l'instrument sont recherché dans une image de gradient binarisée en fonction de la direction de celui-ci. La position de l'instrument n'est déterminée qu'en deux dimensions (l'objectif était l'asservissement d'un robot porte-endoscope). Cano et al. [2006, 2008] et Oropesa et al. [2012] proposent une approche basée sur la détection des bords de l'instrument (par augmentation du contraste, segmentation en fonction de la couleur, filtrage de Sobel et transformée de Hough) et la connaissance de la géométrie de ce dernier pour estimer sa position en trois dimensions. Ils parviennent à suivre deux instruments simultanément en réduisant les angles accessibles pour la transformée de Hough, dans une approche similaire à celle présentée dans Voros et al. [2007] et Allen et al. [2011]. Leur méthode est cependant très sensible aux paramètres choisis pour l'étape de traitement d'image, ceux-ci devant être déterminés par l'utilisateur de façon empirique pour chaque situation d'utilisation (testé sur banc d'entraînement uniquement, avec un fond simple), et les auteurs ne donnent pas d'informations sur le temps de calcul. Sánchez-González et al. [2011] s'appuient sur des paramètres de couleur et de forme de l'instrument pour le segmenter dans des images chirurgicales avant d'extraire ses bords par filtrage laplacien et transformée de Hough, puis déterminent sa position en trois dimensions avec le même modèle que les citations précédentes. Les auteurs précisent cependant que la performance du suivi est très dépendante de la qualité de l'image des instruments, la rendant peu robuste. Loukas et al. [2013] combinent deux approches pour détecter la position de l'instrument dans le repère lié à la caméra : un suivi de marqueur ajouté à sa pointe basé sur ses caractéristiques colorimétriques dans l'espace HSV, et une recherche des droites formant les bords de l'instrument dans un sous-espace de l'image mis à jour grâce à un filtre de Kalman. Cette approche fonctionne en temps réel, est robuste mais n'a été évaluée que sur un fond de banc d'entraînement (uni).

La plupart de ces méthodes sont coûteuses en temps de calcul car l'ensemble de l'image doit être exploré pour extraire les bords ou les régions d'intérêt. Il est donc fondamental de réduire l'espace des observations possibles en prédisant la position de l'instrument d'une image à l'autre.

ii) Optimisation de la vitesse du suivi

Afin de réduire l'espace des observations, et dans une approche similaire à celle que j'ai présentée dans Wolf et al. [2011], Climent et Hexsel [2012] utilisent un algorithme de CONDENSATION [1] [Isard et Blake, 1998a] pour représenter les couples (ρ, θ) à explorer avec la transformée de Hough dans l'image (obtenue de la même façon que [Climent et Marés, 2004]). Speidel et al. [2006] utilisent le même algorithme pour suivre le centre de gravité de régions d'intérêt représentant l'instrument

1. *CONditional DENsity propagaTION*, décrit plus loin dans ce manuscrit, section 3.2.7

en deux dimensions, sans utiliser de modèle géométrique pour retrouver la position de l'instrument en trois dimensions. Doignon et al. [2006] réduisent le domaine de recherche des germes pour la croissance de région et excluent les erreurs de mesure en déterminant la projection du point d'insertion dans le plan image, ce qui limite les points d'entrée possibles de l'instrument dans celle-ci. Oropesa et al. [2012] définissent une fenêtre de suivi dont la taille est basée sur la position et le déplacement de l'instrument à l'étape précédente, et réduisent le traitement d'image à cette fenêtre. Si aucun instrument n'est détecté, l'algorithme est relancé pour la totalité de l'image. Enfin Loukas et al. [2013] réduisent l'espace de recherche en appliquant un filtre de Kalman sur les coordonnées des bords de l'instrument dans l'espace de Hough.

Toutes ces méthodes n'exploitent pas pleinement l'information contenue dans la relative fixité du point d'insertion de l'instrument dans la cavité abdominale pour réduire l'espace de recherche des bords potentiels de l'instrument. Je propose ici une méthode modélisant les positions successives possibles de l'instrument en m'appuyant sur cette propriété.

3.2 Approche développée

3.2.1 Cadre de travail général

La connaissance de la position du point d'insertion de l'instrument de laparoscopie dans la cavité abdominale permet de réduire le nombre de degrés de liberté des mouvements de ce dernier de 7 degrés (translation selon 3 directions, rotation selon 3 directions, rotation sur son propre axe) à 4 degrés (latitude, longitude, translation, rotation sur son propre axe, voir figure 3.4).

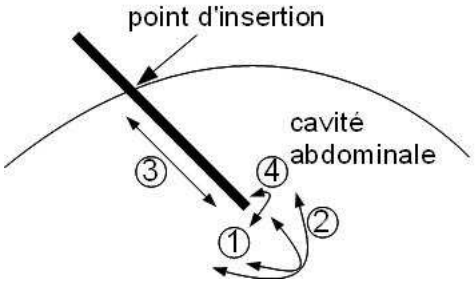


FIGURE 3.4. – Degrés de liberté d'un instrument de laparoscopie contraint par le passage par un point d'insertion dans la cavité abdominale : ① → longitude, ② → latitude, ③ → translation, ④ → rotation sur son axe.

Cette hypothèse de fixité du point d'insertion n'est pas parfaitement vérifiée dans les conditions du bloc opératoire : les déplacements de la paroi abdominale liés à

la respiration du patient, aux variations de la pression du pneumo-péritoine ou aux mouvements du chirurgien sont faibles, mais ne peuvent être entièrement négligés. J'explique plus loin comment compenser cette incertitude sur la position du point d'insertion.

De plus, la position du point d'insertion est fixe dans le repère lié au patient, mais n'est pas fixe dans le repère lié à la caméra endoscopique, qui est le repère que nous utiliserons dans la suite de ce travail. Afin de simplifier la compréhension de la méthode développée, nous nous plaçons ici dans le cas d'une caméra fixe. Toute l'approche est généralisable au bloc opératoire à condition de pouvoir calculer à chaque instant la position de la caméra (donc en connaissant sa matrice de calibrage extrinsèque). Cette condition peut être remplie en utilisant par exemple un robot porte-endoscope (nous utilisons le robot ViKY®, développé par notre équipe). La connaissance des déplacements effectués par le robot nous permet à chaque instant de calculer la nouvelle position du point d'insertion dans le repère lié à la caméra (calibrage main-œil [Horaud et Dornaika, 1995, Krupa et al., 2003]).

Afin de pouvoir passer dans les trocarts sans causer d'exsufflation du pneumo-péritoine, les instruments de laparoscopie doivent avoir une forme standardisée, cylindrique. Le diamètre le plus couramment utilisé est un diamètre de 5mm. On peut donc générer aisément un modèle géométrique cylindrique de l'instrument, invariant par translation et rotation sur son propre axe, afin de réduire encore davantage l'espace de recherche des droites formant ses bords dans l'image endosocopique. L'invariance par rotation d'un cylindre ne permet pas à ma méthode de retrouver le degré de rotation de l'instrument sur son propre axe, limite que je discuterai plus loin.

Mon approche est résumée figure 3.5.

Toute la méthode s'appuie sur la combinaison du modèle géométrique de l'instrument et du modèle projectif de la caméra (calibrage intrinsèque) pour faire correspondre à une orientation donnée de l'instrument en trois dimensions une paire de droites dans l'image endoscopique, approche initialement proposée par Duchateau et al. [2007].

Après une phase de calibrage permettant de déterminer la position du point d'insertion dans le repère lié à la caméra, je discrétise les orientations possibles que peut prendre l'instrument par l'utilisation d'une sphère géodésique, puis étudie la probabilité que l'instrument prenne une orientation donnée grâce à un algorithme de CONDENSATION. Je détermine enfin le dernier degré de liberté, c'est-à-dire la profondeur d'insertion.

Je détaille ma méthode étape par étape dans la suite de cette section.

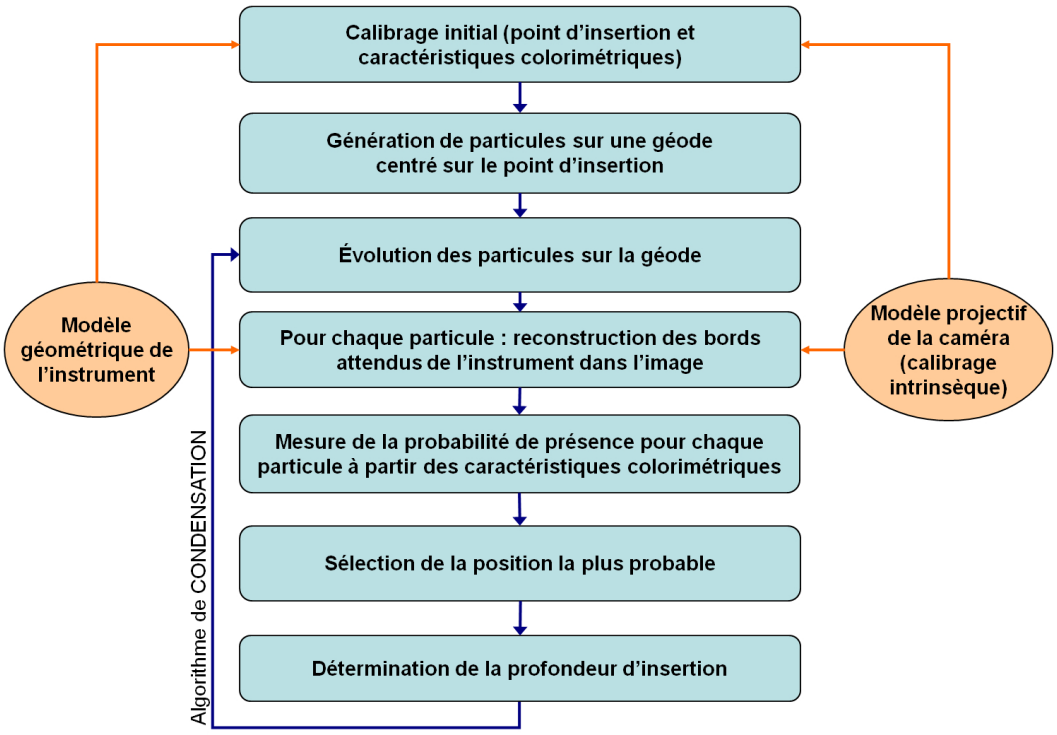


FIGURE 3.5. – Diagramme global de la méthode de suivi d'instrument.

3.2.2 Calibrage de la caméra

Nous adoptons pour représenter la relation faite par la caméra entre le monde en trois dimensions et l'image en deux dimensions le modèle du sténopé [2] (ou modèle à projection centrale), qui présente l'avantage d'être simple et linéaire [Hartley et Zisserman, 2003]. Ce modèle est valable si le système optique de la caméra respecte les conditions de Gauss, ce qui est le cas des systèmes utilisés en chirurgie (combinaison d'un capteur CCD et de lentilles dans l'endoscope).

Dans ce modèle, illustré figure 3.6, un point de coordonnées $\mathbf{X} = (X, Y, Z)^T$ sera projeté sur un point $\mathbf{x}$ du plan image Π correspondant à l'intersection avec ce plan de la droite reliant le centre optique C au point dans l'espace. On définit le repère caméra $\Re_c$ par :

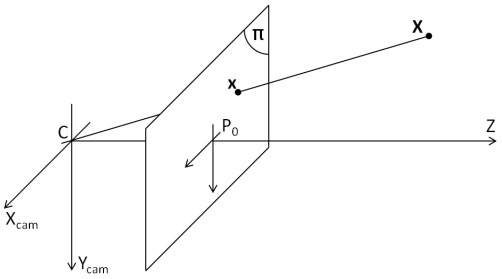
- son origine, le *centre optique* C
- la direction de l'axe $\mathbf{z}$ perpendiculaire au plan image, avec $z_{planimage} = f$, f étant la *distance focale* du système optique. La droite définie par $(C, \mathbf{z})$ est l'*axe optique* du système
- le point P_0 situé à l'intersection de l'axe optique et du plan image est le *point principal* du système

2. *pin-hole model* dans la littérature anglo-saxonne

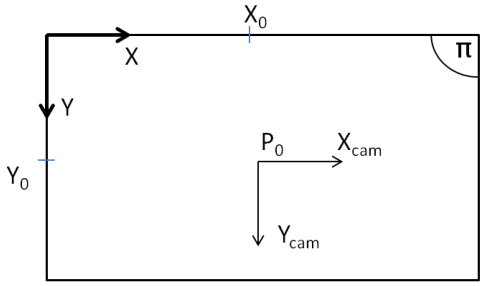
On note que $\mathbf{x} = (fX/Z, fY/Z, f)^T$ dans le repère lié à la caméra $\Re_c$ (figure 3.6a). En prenant en compte le décalage entre l'origine des coordonnées décrivant un point de l'espace euclidien $\mathbb{R}^2$ du plan image et le point principal P_0 de coordonnées $(p_x, p_y)^T$ dans ce repère (figure 3.6b), la transformation entre un point de l'espace $\mathbf{X}$ et sa projection $\mathbf{x}$ dans le repère image peut s'écrire :

$$\mathbf{x} = K[I|\mathbf{0}]\mathbf{X} \quad \text{avec} \quad K = \begin{bmatrix} f & & p_x \\ & f & p_y \\ & & 1 \end{bmatrix}, \mathbf{X} = \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}, \mathbf{x} = \begin{pmatrix} x \\ y \\ Z \end{pmatrix}$$

où $\mathbf{X}$ et $\mathbf{x}$ sont exprimés en coordonnées homogènes.



(a) Illustration du modèle du sténopé.



(b) Position du point principal dans le repère lié au plan image.

FIGURE 3.6. – Modèle du sténopé pour la caméra.

On introduit les coefficients m_x et m_y qui sont le nombre de pixels par unité de distance selon les axes du capteur (pas forcément égaux) pour permettre la transformation entre les coordonnées euclidiennes dans le plan image et la position d'un point sur le capteur CCD. On a alors :

$$\mathbf{x} = K[I|\mathbf{0}]\mathbf{X} \quad \text{avec} \quad K = \begin{bmatrix} \alpha_x & & x_0 \\ & \alpha_x & y_0 \\ & & 1 \end{bmatrix} \tag{3.1}$$

où $\alpha_x = fm_x$, $\alpha_y = fm_y$, $x_0 = m_x p_x$, $y_0 = m_x p_y$.

La matrice K est appelée matrice de calibrage *intrinsèque* de la caméra.

On peut grâce à ce modèle procéder à la reprojection de points de l'espace sur l'image, et faire une projection inverse, c'est à dire déterminer la droite de l'espace contenant les points se projetant sur un pixel de l'image.

▷ **Reprojection** Comme expliqué ci-dessus, un point $\mathbf{X}$ est projeté sur le point $\mathbf{x}$ (points exprimés en coordonnées homogènes) selon la transformation $\mathbf{x} = K[I|\mathbf{0}]\mathbf{X}$.

Cette propriété nous permet de projeter des droites sur l'image pour effectuer des mesures sur les pixels représentant ces droites (voir 3.2.7).

▷ **Projection inverse** Un point de coordonnées $(\tilde{x}, \tilde{y})$ dans l'image se situe sur une droite passant par le centre optique C de vecteur directeur

$$\mathbf{X} = \mathbf{n_u} \otimes \mathbf{n_v} \qquad \text{où } \otimes \text{ est le produit vectoriel} \qquad (3.2)$$

avec

$$\begin{cases} \mathbf{n_u} = K_1^T - \tilde{x} K_3^T \\ \mathbf{n_v} = K_2^T - \tilde{y} K_3^T \end{cases} \qquad \text{où } K_i \text{ est la } i^{\text{ème}} \text{ ligne de } K$$

Le modèle linéaire représentant la caméra n'est toutefois pas exact pour décrire entièrement la transformation entre le monde et l'image. La principale perturbation est généralement la distorsion radiale. Cette erreur peut être corrigée en appliquant à l'image dès son acquisition une transformation replaçant les pixels à leur position attendue dans le modèle linéaire (voir [Hartley et Zisserman, 2003], chap. 7, p.189 pour davantage de détails).

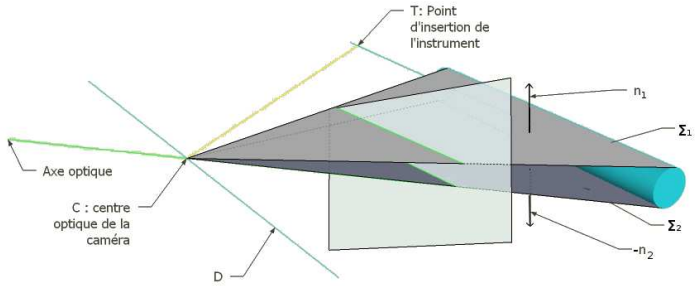
La matrice intrinsèque et les coefficients de distorsion sont déterminés dans le cadre de cette thèse par la méthode de Zhang [2000] à l'aide d'une mire 2D plane en utilisant une librairie OpenCV [3] implentée sous Visual Studio C++ [4]. Cette méthode est très utilisée par la communauté de la vision par ordinateur en raison de sa facilité de mise en œuvre, les paramètres pouvant être calculés à partir d'observations de la mire sous plusieurs vues sans avoir besoins de connaître leur position.

3.2.3 Représentation géométrique de l'instrument

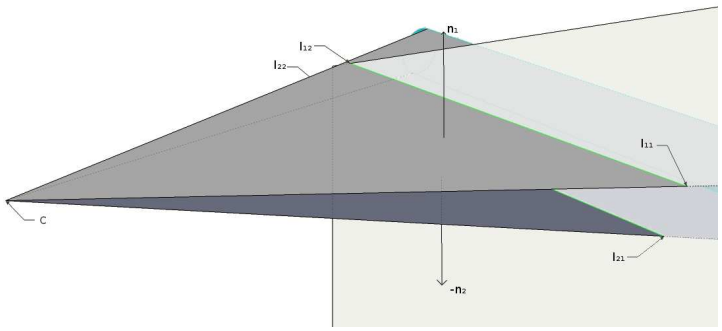
Comme expliqué ci-dessus, un instrument de laparoscopie peut être modélisé comme un cylindre de rayon connu. Ses bords dans l'image apparaissent alors comme deux segments de droites (indexées par i), qui définissent en trois dimensions deux plans tangents à l'instrument passant par le centre optique C de la caméra dans le modèle du sténopé (figure 3.7). Je décris ces plans Σ_i par les couples $(C, \mathbf{n_i})$, $\mathbf{n_i}$ étant la normale au plan i. Connaissant deux points dans l'image appartenant aux bords (I_{i1} et I_{i2}), il est aisé de construire $\mathbf{n_i} = \frac{\overrightarrow{CI_{i1}} \otimes \overrightarrow{CI_{i2}}}{||\overrightarrow{CI_{i1}} \otimes \overrightarrow{CI_{i2}}||}$ à partir de la matrice de calibrage intrinsèque de la caméra (éq. (3.2)).

Par construction des plans tangents à l'instrument, l'intersection de ces deux plans est une droite D parallèle à l'axe central de l'instrument, passant par le point

3. http ://opencv.willowgarage.com/wiki/
4. http ://msdn.microsoft.com/fr-fr/express



(a) Les plans $\Sigma_i : (C, \mathbf{n_i})$ sont les plans tangents à l'instrument passant par le centre optique, $D = \Sigma_1 \cap \Sigma_2$ est la droite parallèle à l'axe de l'instrument, $D : (C, \mathbf{e_1})$ avec $\mathbf{e_1} = \mathbf{n_1} \otimes \mathbf{n_2}$.



(b) Les normales $\mathbf{n_i}$aux plans Σ_i sont définies par $\mathbf{n_i} = \frac{\overrightarrow{CI_{i1}} \otimes \overrightarrow{CI_{i2}}}{||\overrightarrow{CI_{i1}} \otimes \overrightarrow{CI_{i2}}||}$.

FIGURE 3.7. – Représentation géométrique de l'instrument dans l'espace.

d'insertion, décrite par $(C, \mathbf{e_1})$, avec $\mathbf{e_1} = \frac{\mathbf{n_1} \otimes \mathbf{n_2}}{||\mathbf{n_1} \otimes \mathbf{n_2}||}$ son vecteur directeur normalisé.

Je détermine la position d'un point de passage de l'axe directeur du cylindre (appelé dorénavant point d'appui, noté P) représentant l'instrument grâce à la connaissance du rayon ρ de ce dernier. Dans le plan perpendiculaire à l'axe de l'instrument passant par le centre optique (Δ), et en posant $\mathbf{m_i} = \mathbf{e_1} \otimes \mathbf{n_i}$, le point d'appui doit satisfaire la condition :

$$\lambda \mathbf{m_1} - \rho \mathbf{n_1} = \lambda \mathbf{m_2} - \rho \mathbf{n_2}$$

$$\text{d'où} \qquad \lambda = \rho \frac{||\mathbf{n_1} + \mathbf{n_2}||^2}{(\mathbf{m_1} - \mathbf{m_2}) \cdot (\mathbf{n_1} + \mathbf{n_2})}$$

$$\text{et} \qquad \overrightarrow{CP} = \rho \frac{||\mathbf{n_1} + \mathbf{n_2}||^2}{(\mathbf{m_1} - \mathbf{m_2}) \cdot (\mathbf{n_1} + \mathbf{n_2})} \mathbf{m_1} - \rho \mathbf{n_1}$$

avec λ la distance du centre optique au point tangent (figure 3.8).

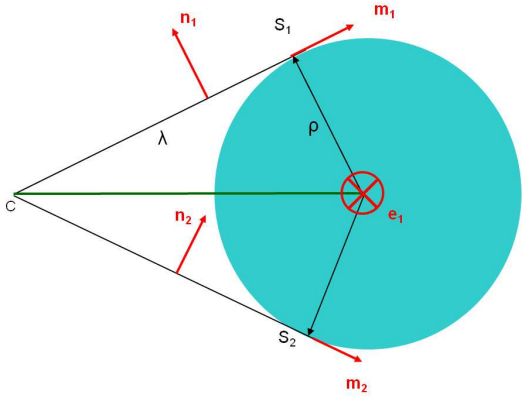


FIGURE 3.8. – Détermination de la position dans l'espace de l'instrument à partir de ses bords dans l'image.

La connaissance de quatre points définissant deux bords de l'instrument dans l'image suffit donc à déterminer la position de l'axe central de celui-ci dans l'espace.

Inversement, on peut reconstruire les deux bords de l'instrument dans l'image à partir du couple $(T, \mathbf{e_1})$, T étant un point appartenant à l'axe de l'instrument (le point d'insertion par exemple) et $\mathbf{e_1}$ son orientation dans l'espace (figure 3.9). Soit $d = \left|\left|\overrightarrow{CT} \otimes \mathbf{e_1}\right|\right|$, $\mathbf{e_2} = \frac{\overrightarrow{CT} \otimes \mathbf{e_1}}{||\overrightarrow{CT} \otimes \mathbf{e_1}||}$, $\mathbf{e_3} = \mathbf{e_1} \otimes \mathbf{e_2}$. En notant S_i les points appartenant à la surface de l'instrument et à l'intersection de Σ_i et Δ, on a $\overrightarrow{PS_i} = \pm\alpha \mathbf{e_2} + \beta \mathbf{e_3}$

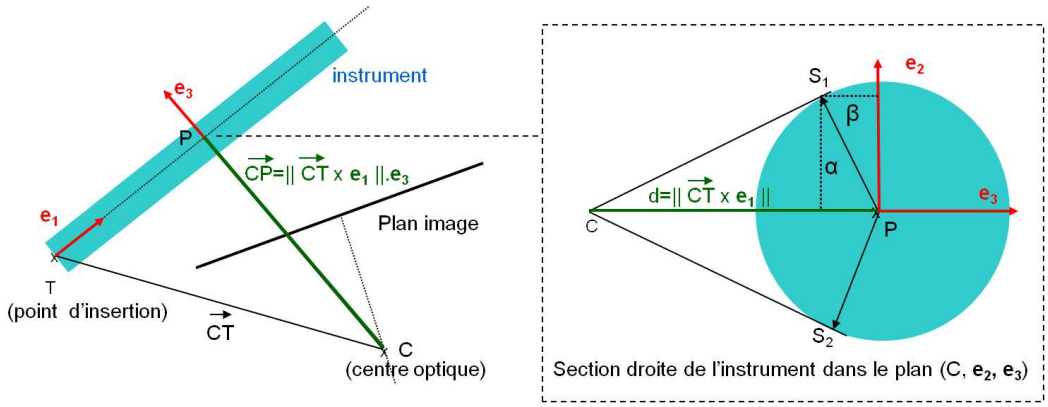


FIGURE 3.9. – Détermination de la position des bords à partir de l'axe et d'un point de passage de l'instrument.

avec :

$$\begin{cases} \alpha = \rho\sqrt{1 - \frac{\rho^2}{d^2}} \\ \beta = -\frac{\rho^2}{d} \end{cases}$$

Les deux droites tangentes à l'instrument et passant par le centre optique sont alors définies par :

$$\begin{cases} \Gamma_1 : M/\overrightarrow{CM} = \overrightarrow{CT} + \alpha\mathbf{e_2} + \beta\mathbf{e_3} + \gamma\mathbf{e_1}, \gamma \in \mathbb{R} \\ \Gamma_2 : M/\overrightarrow{CM} = \overrightarrow{CT} - \alpha\mathbf{e_2} + \beta\mathbf{e_3} + \gamma\mathbf{e_1}, \gamma \in \mathbb{R} \end{cases} \tag{3.3}$$

La matrice de calibrage intrinsèque de la caméra permet de projeter les points appartenant à ces droites sur le plan image (éq. (3.1)).

3.2.4 Traitement de l'image

Mon approche s'appuie sur la détection des bords de l'instrument dans une image endoscopique. Un traitement de l'image est donc nécessaire pour prédire la probabilité qu'un pixel donné appartienne ou non à un de ces bords. Je décris dans cette section les différentes étapes de ce traitement de l'image.

i) Choix de l'espace colorimétrique

L'image acquise par l'endoscope est codée dans l'espace colorimétrique RGB [Wyszecki et Stiles, 2000], qui est généralement utilisé pour le stockage des fichiers vidéos en informatique.

Doignon et al. [2005] présentent cependant de nombreux arguments incitant à ne pas se placer dans cet espace colorimétrique pour analyser les instruments de laparoscopie, et proposent d'utiliser l'espace TSL (pour Teinte, Saturation, Luminance) pour identifier les régions grises appartenant à l'instrument.

Cet espace est plus proche de la perception physiologique humaine, dans la mesure où il représente les trois paramètres théoriques de la vision. Le modèle TSL peut être représenté par deux cônes inversés placés l'un au dessus de l'autre (figure 3.10a). La teinte (*hue*) correspond à la forme pure d'une couleur, sans adjonction de blanc ou de noir. Elle est définie comme un angle sur le cercle chromatique de Newton, compris entre 0 et 2π.

$$\begin{cases} T = \theta & \text{si} \quad G \geq B \\ T = 2\pi - \theta & \text{si} \quad G \leq B \end{cases} \quad \text{avec} \quad \theta = \cos^{-1}\left[\frac{\frac{1}{2}\left[(R-G)+(R-G)\right]}{\sqrt{(R-G)^2+(R-B)(G-B)}}\right]$$

La saturation est une mesure du degré de pureté d'une couleur, ou encore de la quantité de gris dans un pixel. La Commission Internationale de l'Éclairage la définit comme "*the colourfulness of an area judged in proportion to its brightness*". Elle varie de 0 (gris) à 1 (couleur pure), et est représentée par le rayon d'une section circulaire du cône, c'est à dire une distance à l'axe achromatique. Elle est calculée dans l'espace TSL [Smith, 1978] par :

$$S = 1 - 3 \cdot \frac{\min[R,G,B]}{R+G+B} \in [0,1]$$

Smith [1978] définit également la luminance, qui représente l'intensité de la lumière reçue par le récepteur. Elle est définie par :

$$L = 0.299 \cdot R + 0.587 \cdot G + 0.114 \cdot B \in [0,255],$$

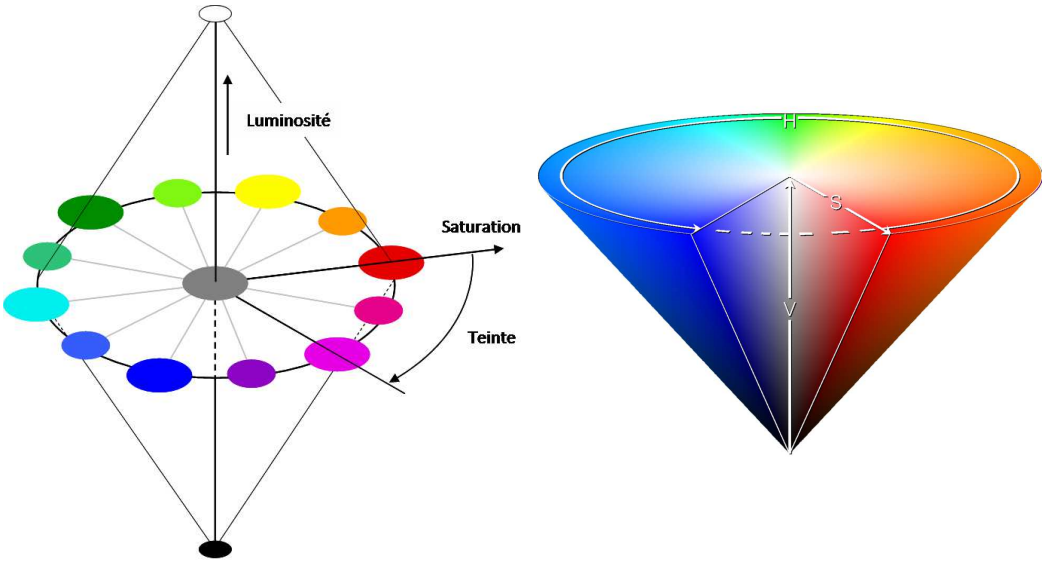
les coefficients traduisant les différentes sensibilités des récepteurs de la rétine aux différentes longueurs d'onde. L correspond aux niveaux de gris dans l'image.

Les composantes de teinte et de saturation définissent la chromacité qui, comme le soulignent Doignon et al. [2004], est décorrélée de la luminosité, celle-ci pouvant être changeante au cours d'une intervention chirurgicale. En revanche on ne s'attend pas à ce que la saturation d'un instrument chirurgical, qui est généralement noir, gris ou argenté, varie au cours de cette intervention. Il paraît donc judicieux de se baser sur cette composante pour identifier l'instrument dans les images endoscopiques.

En adoptant une approche inspirée des travaux cités ci-dessus, je me place dans l'espace TSV (Teinte, Saturation, Valeur), un modèle perceptuel proche du modèle TSL, pour détecter les instruments dans l'image. Cet espace n'utilise pas directement la luminosité, mais définit la variable *valeur* comme la "brillance" d'un pixel, $V = \max[R,G,B]$. Diminuer la valeur revient à augmenter la "noirceur" du pixel. Le calcul de la saturation et de la teinte sont également modifiés, avec en particulier pour la saturation :

$$S' = 1 - \frac{\min[R,G,B]}{\max[R,G,B]} \in [0,1]$$

La principale conséquence de cette modification, comme l'expliquent Doignon et al. [2005], est d'augmenter la saturation des pixels les plus colorés, sans modifier pro-



(a) extrait de Anselme et Gadal [2008] (b) H : Teinte, S : Saturation, V : Valeur

FIGURE 3.10. – Représentations graphiques des espaces colorimétriques {*Teinte, Saturation, Luminance*} (a) et {*Teinte, Saturation, Valeur*} (b).

fondément la saturation des pixels moins chromatiques. Un effet parallèle est d'augmenter la sensibilité de la saturation aux changements de valeur pour les pixels chromatiques, mais cet effet est marginal sur les pixels achromatiques qui nous intéressent ici. La différence entre les deux modes de calcul de la saturation est illustrée figure 3.11.

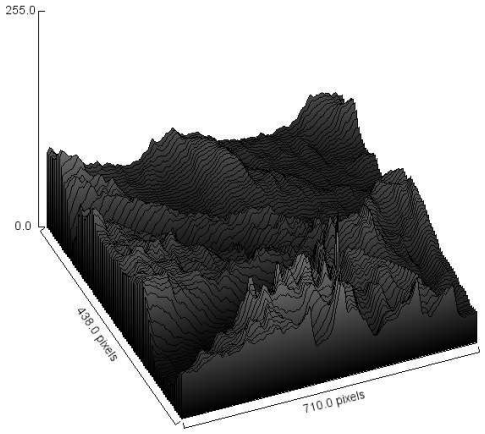
Toujours à partir de ces travaux, j'affecte aux pixels très lumineux ou très sombres ($L < 0.1 * 255$ ou $L > 0.9 * 255$ avec $L = 0.299 \cdot R + 0.587 \cdot G + 0.114 \cdot B$ la luminance) une saturation de 1, les classant dans le groupe des pixels n'appartenant probablement pas à l'instrument.

ii) Filtrage

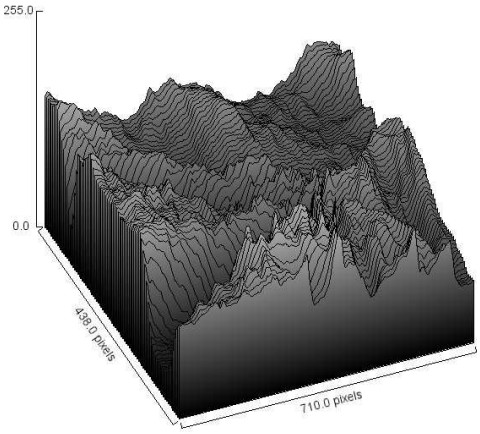
J'applique à l'image obtenue à l'étape précédente un filtre permettant de réduire les bruits de mesure, afin d'améliorer la détection des angles lors des prochaines étapes. Les filtres linéaires habituels de type gaussien font l'hypothèse que les propriétés d'une image (la saturation dans ce travail) varient relativement lentement dans l'espace, et que des pixels proches auront donc sensiblement les mêmes valeurs. Cette hypothèse permet alors de fixer la valeur d'un pixel à la moyenne de son voisinage, avec un poids décroissant en fonction de la distance. Cette hypothèse n'est pas valable lorsqu'on recherche des angles dans l'image, car on souhaite conserver l'information de variation spatiale rapide de la valeur entre deux pixels voisins, information diminuée en cas de filtrage gaussien. Cette problématique in-

(a) Image de saturation S dans l'espace TSL

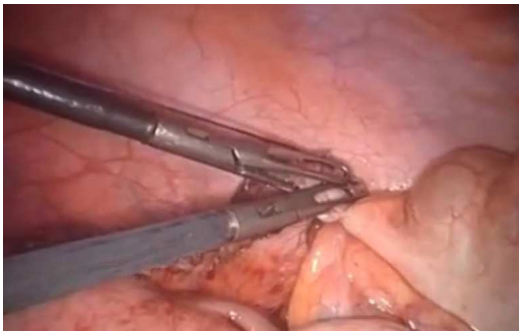
(b) Image de saturation S' dans l'espace TSV



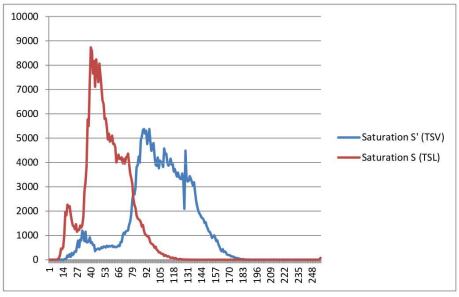
(c) Représentation 3D de la saturation S dans l'image



(d) Représentation 3D de la saturation S' dans l'image



(e) Image originale, extraite d'une vidéo d'appendicectomie.



(f) Comparaison des histogrammes des images (a) et (b)

FIGURE 3.11. – Comparaison entre les deux modes de calcul de la saturation. On note que la différence entre les "vallées" (correspondant aux instruments) et les "plateaux" (correspondant au fond de l'image) dans la représentation 3D des valeurs de saturation est plus marquée avec l'utilisation de la saturation S'.

cite donc à s'orienter vers un filtrage non-linéaire. Une première implémentation du filtre Sigma [Lee, 1983] avec l'optimisation proposée par Doignon et al. [2005] a montré des résultats satisfaisants mais le temps de calcul rendait son utilisation non compatible avec un suivi de l'instrument en temps réel.

Afin de rendre le traitement d'image compatible avec la contrainte du temps réel, je me suis donc orienté vers l'utilisation d'un filtre bilatéral [Tomasi et Manduchi, 1998]. Le principe de ce filtre est de combiner deux mesures de distance pour pondérer le poids d'un pixel dans le voisinage du pixel filtré : une distance dans l'espace au sens classique du terme (ou *proximité*) c, et une distance en terme de valeur affectée au pixel (ou *similarité*) s. Mathématiquement on écrit :

$$\mathbf{h}(\mathbf{x}) = k^{-1}(\mathbf{x}) \int_{-\frac{d}{2}}^{\frac{d}{2}} \int_{-\frac{d}{2}}^{\frac{d}{2}} \mathbf{f}(\xi) c(\xi, \mathbf{x}) s(\mathbf{f}(\xi), \mathbf{f}(x)) \mathrm{d}\xi \tag{3.4}$$

$$\text{avec} \quad k(\mathbf{x}) = \int_{-\frac{d}{2}}^{\frac{d}{2}} \int_{-\frac{d}{2}}^{\frac{d}{2}} c(\xi, \mathbf{x}) s(\mathbf{f}(\xi), \mathbf{f}(\mathbf{x})) \mathrm{d}\xi, \tag{3.5}$$

où $\mathbf{f}(x)$ est la valeur du pixel x, d est le diamètre du filtre

En choisissant comme mesures de proximité c et de similarité s des gaussiennes de la distance euclidienne entre leurs arguments, on obtient :

$$c(\xi, \mathbf{x}) = \mathrm{e}^{-\frac{1}{2}\left(\frac{d(\xi, \mathbf{x})}{\sigma_d}\right)} \qquad \text{avec} \quad d(\xi, \mathbf{x}) = ||\xi - \mathbf{x}||_{\mathbf{L_2}}$$

$$s(\mathbf{f}(\xi), \mathbf{f}(\mathbf{x})) = \mathrm{e}^{-\frac{1}{2}\left(\frac{\delta(\mathbf{f}(\xi), \mathbf{f}(\mathbf{x}))}{\sigma_r}\right)} \qquad \text{avec} \quad \delta(\mathbf{f}(\xi), \mathbf{f}(\mathbf{x})) = ||\mathbf{f}(\xi) - \mathbf{f}(\mathbf{x})||_{\mathbf{L_2}}$$

La figure 3.12 illustre le principe du filtre bilatéral au niveau d'une marche.

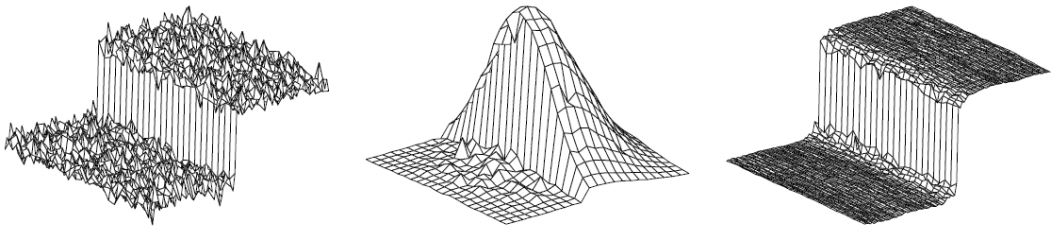


FIGURE 3.12. – Illustration du fonctionnement du filtre bilatéral, extraite de [Tomasi et Manduchi, 1998].
L'image originale (a) représente une marche avec un bruit gaussien, (b) correspond à la représentation du poids de chaque pixel dans le voisinage d'un pixel à droite de la marche pour le calcul de la valeur de ce dernier, (c) représente la marche après filtrage.

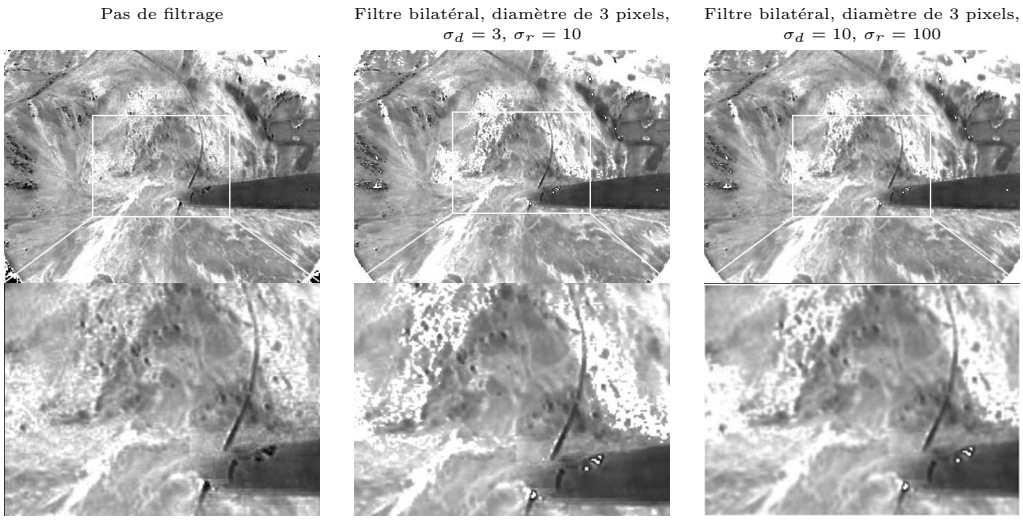


FIGURE 3.13. – Variations du filtrage bilatéral en fonction de σ_d et σ_r sur une image de saturation S'.

La taille du filtre, d, sera ajustée plus tard en fonction des résultats de la détection d'angle (si trop d'angles potentiels sont détectés, l'image est davantage filtrée). Une phase d'ajustement permet de fixer les coefficients $\sigma_d = 10$ et $\sigma_r = 100$, valeurs qui permettent d'avoir un filtrage visuellement de bonne qualité quel que soit le type d'image en entrée (figure 3.13). On note que l'augmentation du diamètre du filtre bilatéral, coûteux en temps de calcul (figure 3.14), aboutit pour un diamètre supérieur à 7 à un filtrage trop important pour la détection des bords de l'instrument (figures 3.15).

Comme on peut le remarquer sur la figure 3.14, une meilleure implémentation en termes d'allocations mémoire des filtres Sigma et Sigma optimisé me permet finalement d'avoir des vitesses de calcul similaires voire supérieures à celles obtenues par l'emploi du filtre bilatéral. Cette implémentation n'a pas été utilisée pour l'analyse des trajectoires de la partie III. Je comparerai cependant les performances de ces différents filtres pour la détection des bords de l'instrument plus loin.

iii) Détection des angles dans l'image par l'algorithme de Canny

La première méthode utilisée pour détecter les bords de l'instrument était une mesure du gradient en x et en y pour chaque pixel. Cette méthode s'est avérée très sensible au bruit, et a été abandonnée au profit d'une détection des angles des instruments basée sur un filtre de Canny [Canny, 1986].

Ce filtre développé pour la détection des contours est largement utilisé dans la

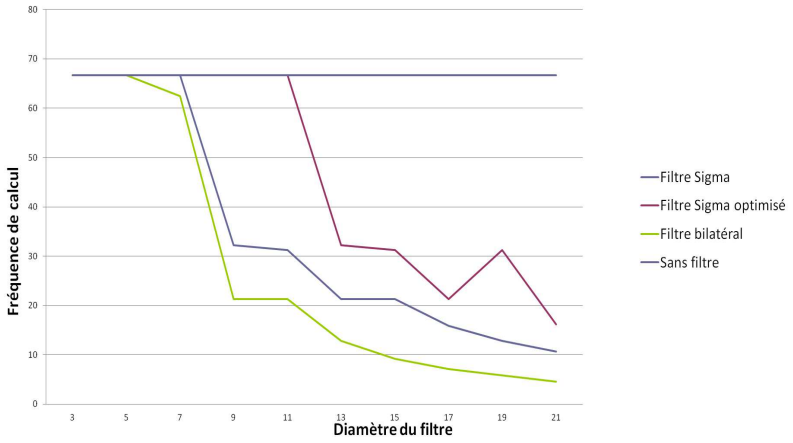


FIGURE 3.14. – Évolution du temps de calcul pour différents filtres de bruit en fonction de leur diamètre

communauté de la vision par ordinateur, car il permet d'être optimal pour les trois critères suivants :

bonne détection : faible taux d'erreur, c'est à dire peu de faux négatifs (angles non détectés) et peu de faux positifs (détection d'angles inexistants).

bonne localisation : la distance entre les points considérés comme des angles et la position réelle de l'angle dans l'image doit être minimale.

clarté de la réponse : absence de plusieurs réponses pour un angle unique

Dans son article, Canny explicite ces trois critères sous forme mathématique, et propose une succession d'étapes permettant l'optimisation de ces critères :

▷ **Filtrage du bruit :** Canny propose l'utilisation d'un filtre gaussien, filtre qui présente les limites décrites ci-dessus. Je lui préfère le filtrage non linéaire afin de conserver davantage d'information au niveau des angles, avec les filtres présentés plus haut.

▷ **Calcul des gradients d'intensité :** Cette étape est réalisée en appliquant à l'image des masques de convolution décrivant les dérivées de l'intensité en x et y. Je choisis ici d'utiliser des filtres de Sobel de dimension 3x1 et 1x3 ($S_x = [-1 \quad 0 \quad 1]$ et $S_y = [1 \quad 0 \quad -1]^T$ afin de ne pas filtrer à nouveau l'image (les filtres de Sobel 3x3 ont une composante de filtre gaussien). Pour chaque pixel, on calcule $G = \sqrt{S_x^2 + S_y^2}$ la valeur du gradient et $\theta = \tan^{-1}\left[\frac{S_y}{S_x}\right]$ sa direction.

▷ **Suppression des non-maxima locaux :** La valeur du gradient peut être élevée pour un pixel donné sans que celui-ci soit nécessairement situé sur un angle : il peut

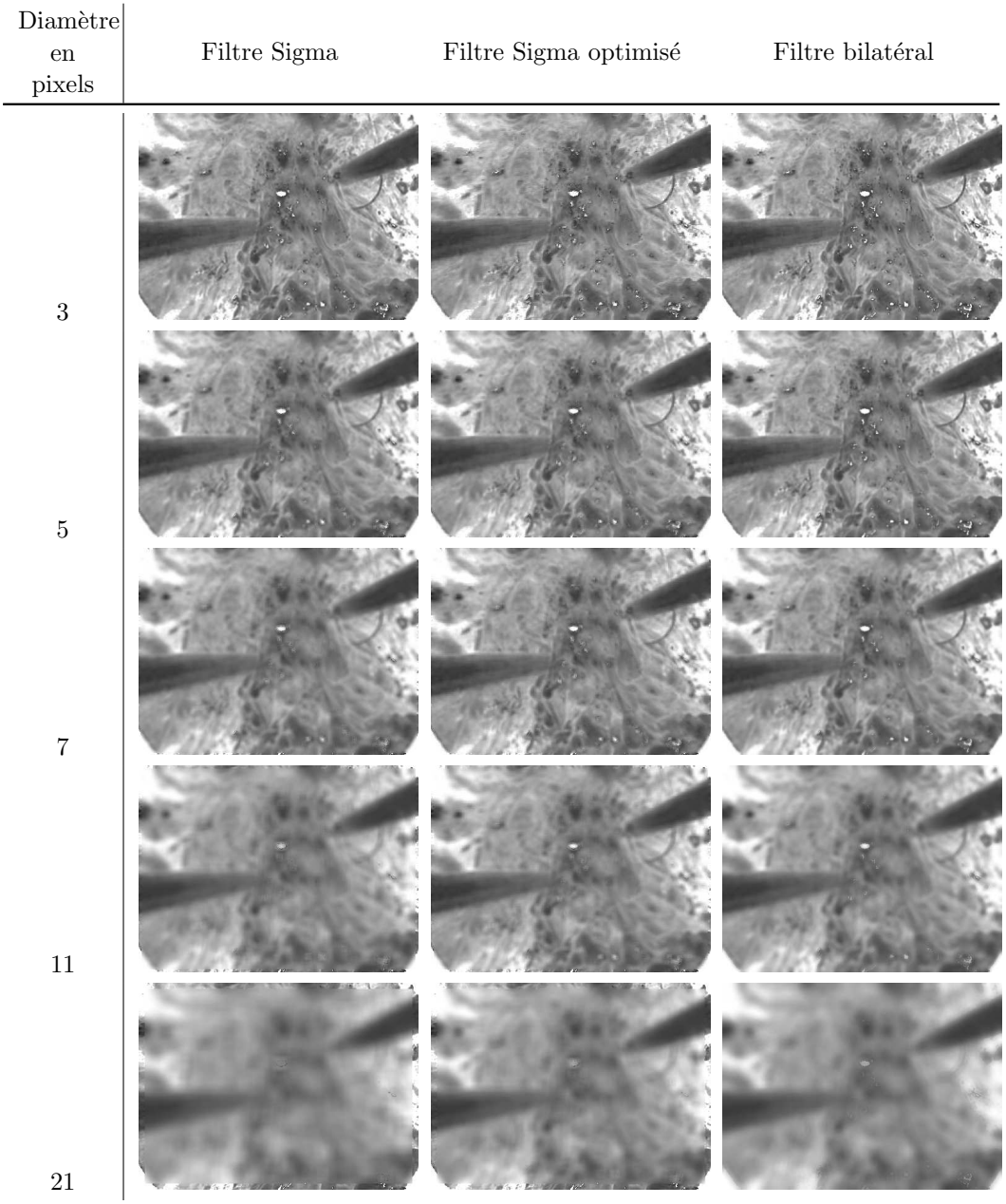
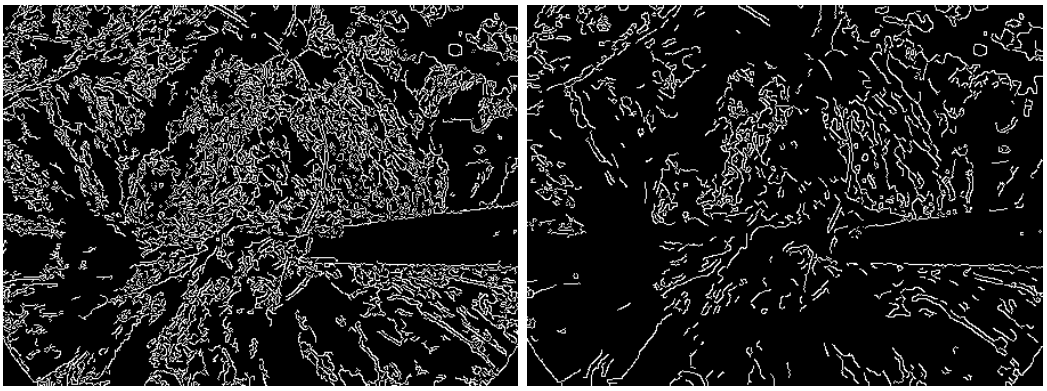


FIGURE 3.15. – Effet du diamètre du filtre bilatéral sur l'image de saturation. Pour les filtres Sigma et Sigma optimisé, la valeur du σ est déterminée de façon analogue à Doignon et al. [2005]

par exemple appartenir à une zone dont l'intensité décroît de façon homogène selon une direction donnée. On ne conserve donc pour la détection des angles que les maxima locaux du gradient, c'est à dire les points où sa dérivée spatiale s'annule.



(a) (b)

FIGURE 3.16. – Illustration de l'effet du filtrage bilatéral sur la détection des angles par l'algorithme de Canny. (a) : sans filtrage préalable ; (b) : avec filtre bilatéral. On constate que le bruit du fond de l'image est nettement diminué, mais que les bords de l'instrument restent détectés avec l'utilisation du filtre bilatéral. Les seuils de l'algorithme de Canny sont fixés à l'aide de la méthode présentée paragraphe iv).

▷ **Seuillage par hystérésis :** La définition d'un seuil de gradient permettant de séparer les pixels marqués comme maxima en deux classes, ceux appartenant aux angles et ceux n'appartenant pas aux angles, est une opération délicate : si le seuil est proche de la valeur moyenne du gradient au niveau des angles, alors une partie des pixels appartenant aux angles seront exclus, conduisant à un contour tronqué. À l'inverse si le seuil est trop bas, on observera de nombreux faux positifs. Canny contourne cette difficulté en proposant un seuillage par hystérésis, c'est à dire comportant deux seuils. Les pixels dont la valeur de gradient est située sous le seuil bas sont exclus. Si des pixels sont au dessus du seuil haut, ils sont immédiatement marqués comme appartenant à un contour, de même que tous les pixels reliés par un segment de contour à ce point et dont la valeur du gradient est supérieure au seuil bas. Canny propose un ratio de 2 à 3 sur 1 pour le rapport entre le seuil haut et le seuil bas. On obtient donc une image binarisée. Je stocke par ailleurs lors de cette étape les orientations de gradient pour chaque pixel.

La figure 3.16 illustre l'intérêt de procéder à un filtrage de l'image avant de rechercher les angles dans celle-ci, les angles de l'instrument restant détectés tandis que les angles du fond de l'image sont filtrés.

iv) Définition des seuils du filtre de Canny

La détection des angles avec la méthode présentée par Canny [1986] reste fortement dépendante du choix des paramètres pour le filtrage (diamètre d du filtre) et

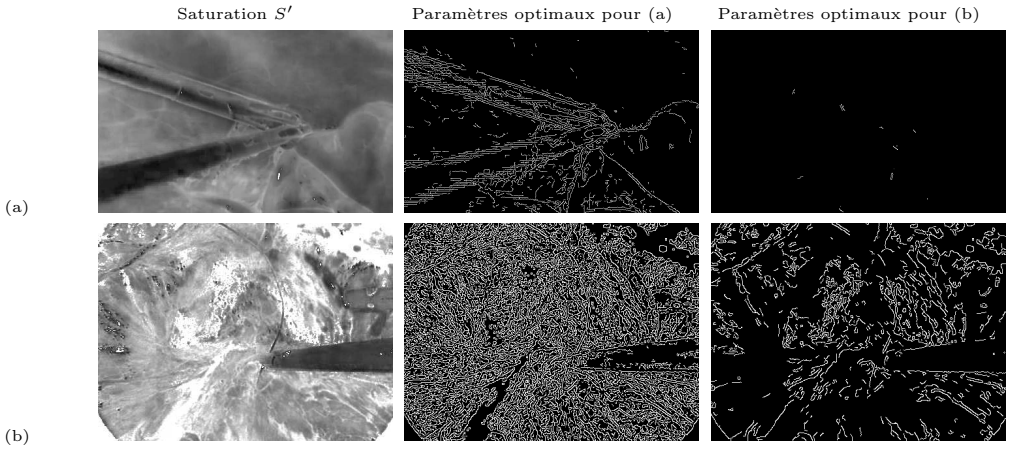


FIGURE 3.17. – Variations de la détection des angles avec la méthode Canny en fonction des seuils T_h et T_b et du diamètre d du filtre de bilatéral. Ces paramètres optimaux pour chaque image sont déterminés avec la méthode proposée par Yitzhaky et Peli [2003].
Paramètres optimaux pour (a) : $\{d = 3, T_h = 9, T_b = 6\}$; paramètres optimaux pour (b) : $\{d = 5, T_h = 40, T_b = 26\}$

pour le seuillage par hystérésis (seuils haut T_h et bas T_b).

Un des objectifs de ma méthode de suivi d'instrument est d'être utilisable dans différents contextes (banc d'entraînement, bloc opératoire) sans nécessiter d'ajustements particuliers dans ses paramètres. De plus, les conditions d'illumination et les saturations et teintes globales d'une scène chirurgicale peuvent varier selon la procédure envisagée (voir par exemple la figure 3.17). Des paramètres mal ajustés conduiront soit à des faux négatifs (image (a) de la figure 3.17 traitée avec les paramètres optimaux de l'image (b)), soit à des faux positifs (image (b) traitée avec les paramètres optimaux de l'image (a)). Je dois donc développer une stratégie permettant de définir ces paramètres de façon non supervisée au lancement de la procédure.

Je me base pour définir ces paramètres sur la méthode proposée par Yitzhaky et Peli [2003]. Ils réalisent une analyse statistique non supervisée de la pertinence des détections d'angle à partir de différents paramètres, pour évaluer de façon objective la performance d'un jeu de paramètres et sélectionner les paramètres optimaux. L'analyse statistique revient à estimer un "étalon-or" comportant les pixels appartenant de façon la plus probable aux angles, puis à identifier le jeu de paramètres dont l'image de détection est la plus proche de cet "étalon-or". Je résume ici leur approche en précisant les adaptations effectuées pour s'adapter au contexte de ce travail.

▷ **Sélection d'une image "typique" de la procédure :** Le chirurgien sélectionne au début de sa procédure, c'est à dire après création du pneumo-péritoine et insertion

des trocarts, une image dite "typique". Cette image est constituée d'un arrière-plan constituant la scène chirurgicale, et d'un avant-plan constitué par les instruments qu'il utilisera.

▷ **Définition des jeux de paramètres :** On construit un jeu de N combinaisons des paramètres *diamètre* d du filtre bilatéral, *seuil bas* T_b et *seuil haut* T_h du seuillage par hystérésis que l'on note $PS^i = \{d^i, T_b^i, T_h^i\}_{i=1..N}$ (*parameter set*).

Comme vu dans la section ii), les diamètres à considérer pour le filtrage sont 3, 5 ou 7 pixels, pour des raisons de vitesse de calcul et de préservation des angles. Les seuils bas et haut à analyser sont définis en s'inspirant de la méthode proposée par Canny [1986] pour trouver un seuil unique pour un seuillage. En considérant que le gradient lié au bruit dans l'image obéit à une distribution gaussienne, tandis que les valeurs de gradient au niveau des angles sont élevées mais rares, on peut estimer que les valeurs de gradient liées au bruits sont représentées par les 3 à 4 premiers quartiles de la distribution des gradients des pixels. On fixe donc T_h à une valeur telle que plus de 85% des pixels aient une valeur de gradient inférieure à T_h, et T_b à une valeur telle que plus de 60% des pixels aient une valeur de gradient inférieure à T_b. Cette opération est illustrée par la figure 3.17.

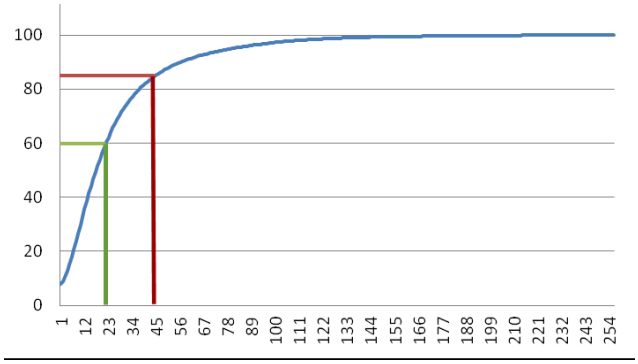


FIGURE 3.18. – Histogramme cumulé des valeurs du gradient de saturation de l'image présentée figure 3.17(b).
Le seuil bas $T_b = 21$ est fixé de façon à ce que 60% des pixels dans l'image aient une valeur inférieure à ce seuil, le seuil haut $T_h = 44$ est fixé de façon à ce que 85% des pixels aient une valeur inférieure à ce seuil.

Ces deux seuils ayant été fixés, je crée le jeu de paramètres $PS^i = \{d^i, T_b^i, T_h^i\}_{i=1..N}$ (voir procédure 3.1).

Le *pas* et le nombre de pas ont été fixés après des essais préliminaires. Il s'agit d'un compromis entre sensibilité (trouver le meilleur jeu de paramètres PS^i) et vitesse d'exécution lors de la recherche du jeu optimal.

Procédure 3.1 : definePS

Input : T_b,T_h
Output : $PS^i = \{d^i, T_b^i, T_h^i\}_{i=1..N}$

$i \leftarrow 1$;
$pas \leftarrow 0.1$;
$nbr_pas \leftarrow 5$;
pour $k = 1 \rightarrow 3$ **faire**
  pour $l = 0 \rightarrow nbr_pas - 1$ **faire**
    pour $m = 1 \rightarrow nbr_pas$ **faire**
      $d^i \leftarrow 2 * k + 1$;
      $T_h^i \leftarrow \left(1 - \left(\frac{nbr_pas}{2} + l\right) * pas\right) * T_h$;
      $T_b^i \leftarrow T_h^i * 2/m$;
      $i \leftarrow i + 1$
    fin
  fin
fin

▷ **Définition des "étalon-or potentiels" :** À partir des N (= 75 ici) combinaisons de paramètres, je crée un jeu d'images C^i obtenues avec le filtre de Canny pour chaque triplet $\{d^i, T_b^i, T_h^i\}_{i=1..N}$. Je dénombre ensuite pour chaque position de pixel le nombre de fois où celui-ci est classé comme angle dans les images C^i, appelé nombre de correspondances. Je crée alors N nouvelles images dites "étalon-or potentiel" (*potential ground truth*), PGT^j, dans lesquelles la valeur 1 est affectée à un pixel s'il est détecté dans plus de j images C^i, c'est à dire si son nombre de correspondances est supérieur à j. La valeur 0 est affectée à ce pixel dans le cas contraire. On a donc une image PGT^j par niveau de correspondance (figure 3.19).

Chaque image PGT^j est alors comparée à l'ensemble des images C^i pour obtenir quatre mesures indiquant si le seuil de correspondance j est sensible et spécifique pour détecter les angles. Plus précisément Yitzhaky et Peli [2003] définissent les mesures suivantes :

(a) (b) (c)

FIGURE 3.19. – Illustration des images PGT^j : les pixels blancs dans l'image PGT^{75} (a) ont été détectés dans les $N(=75)$ images C^i, ceux dans l'image PGT^{37} (b) l'ont été dans 50% des images, ceux dans l'image PGT^1 (c) ne l'ont été que dans une image (pas nécessairement la même pour chaque pixel).

$$\overline{TP}_{PGT^j} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{K.L}\sum_{k=1}^{K}\sum_{l=1}^{L} PGT_1^j \cap D_{i1}\right) \quad \text{le taux de "vrais positifs"}$$

$$\overline{FP}_{PGT^j} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{K.L}\sum_{k=1}^{K}\sum_{l=1}^{L} PGT_1^j \cap D_{i0}\right) \quad \text{le taux de "faux positifs"}$$

$$\overline{TN}_{PGT^j} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{K.L}\sum_{k=1}^{K}\sum_{l=1}^{L} PGT_0^j \cap D_{i0}\right) \quad \text{le taux de "vrais négatifs"}$$

$$\overline{FN}_{PGT^j} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{K.L}\sum_{k=1}^{K}\sum_{l=1}^{L} PGT_0^j \cap D_{i1}\right) \quad \text{le taux de "faux négatifs"}$$

où l'indice 1 désigne les pixels positifs (classés comme angles dans leurs images respectives) et l'indice 0 les pixels nuls. Ces mesures permettent alors de définir une sensibilité Se^j et une spécificité Sp^j pour chaque niveau de correspondance.

$$Se^j = \frac{\overline{TP}_{PGT^j}}{\overline{TP}_{PGT^j} + \overline{FN}_{PGT^j}}$$

$$Sp^j = \frac{\overline{TN}_{PGT^j}}{\overline{TN}_{PGT^j} + \overline{FP}_{PGT^j}}$$

▷ **Définition de "l'étalon-or estimé"** En adoptant une approche semblable à celle utilisée pour les tests diagnostiques, les auteurs proposent de choisir le niveau de correspondance ayant le meilleur couple $\{Se^j, Sp^j\}$ en utilisant une courbe ROC (*Receiver Operating Curve*) ou le test du χ^2. Dans l'approche basée sur la courbe ROC, le couple optimal est représenté par le point $\{Se^j, 1 - Sp^j\}$ le plus proche de

la droite reliant le point $(0,1)$ au point (P,P) avec $P = \overline{TP}_{PGT^j} + \overline{FN}_{PGT^j}$. Dans l'approche basée sur le test χ^2, on cherche à maximiser :

$$\chi^2_{PGT^j} = \frac{Se^j - Q_{PGT^j}}{1 - Q_{PGT^j}} . \frac{Sp^j - (1 - Q_{PGT^j})}{Q_{PGT^j}} \quad \text{avec} \quad Q_{PGT^j} = \overline{TP}_{PGT^j} + \overline{FP}_{PGT^j}$$

L'image correspondant à ce couple $\{Se^j, Sp^j\}$ est alors notée EGT, c'est "l'étalon-or estimé". Les couples obtenus avec les deux méthodes ayant été similaires sur les essais préliminaires (figure 3.20), je choisis d'utiliser l'approche basée sur la maximisation du χ^2, moins complexe et plus rapide.

▷ **Estimation du meilleur triplet de paramètres :** On redéfinit les mesures de $\overline{TP}$, $\overline{FP}$, $\overline{TN}$, $\overline{FN}$, avec :

$$\overline{TP}_{C^i} = \frac{1}{K.L} \sum_{k=1}^{K} \sum_{l=1}^{L} C_{i1} \cap EGT_1 \qquad \overline{FP}_{C^i} = \frac{1}{K.L} \sum_{k=1}^{K} \sum_{l=1}^{L} C_{i1} \cap EGT_0$$

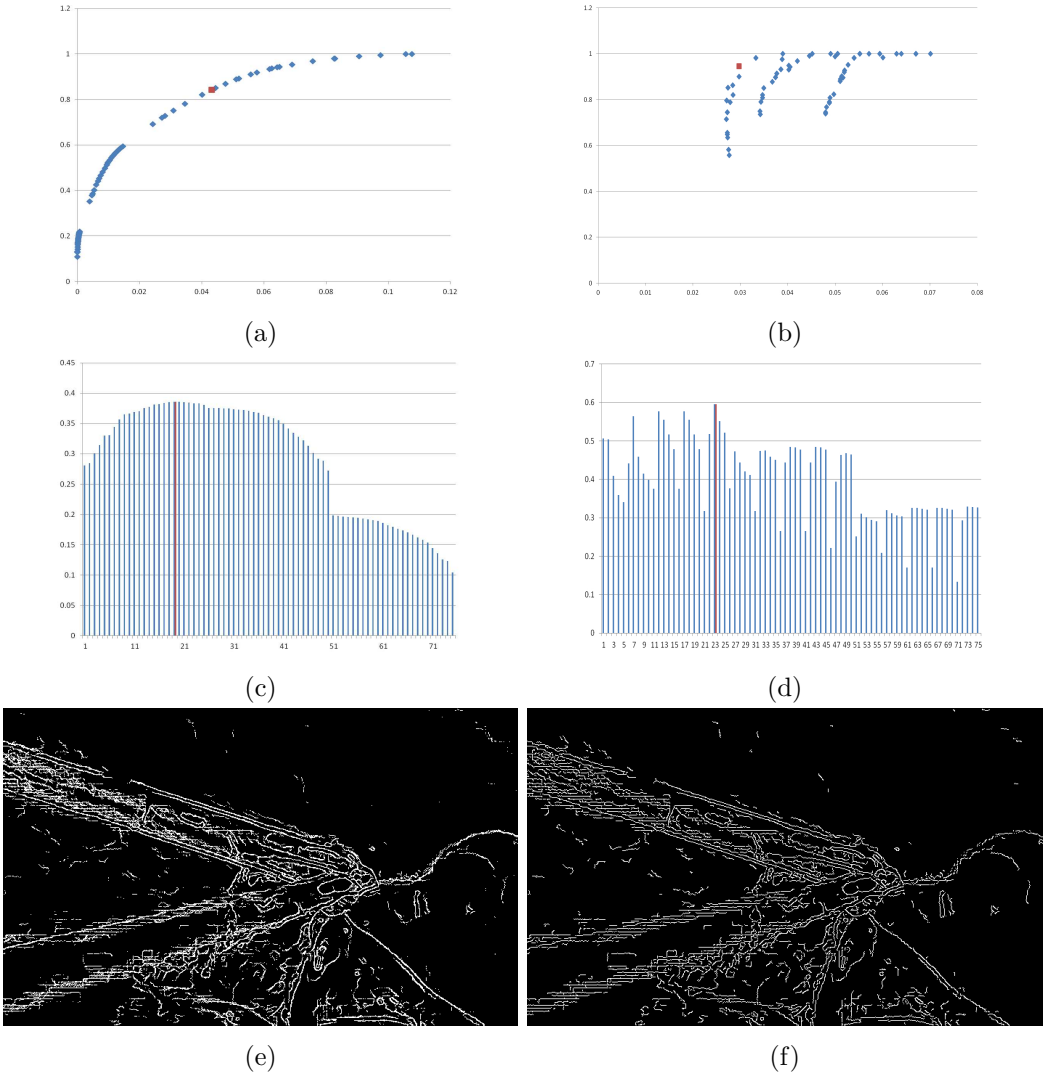
$$\overline{TN}_{C^i} = \frac{1}{K.L} \sum_{k=1}^{K} \sum_{l=1}^{L} C_{i0} \cap EGT_0 \qquad \overline{FN}_{C^i} = \frac{1}{K.L} \sum_{k=1}^{K} \sum_{l=1}^{L} C_{i0} \cap EGT_1$$

et de façon identique à ci-dessus on sélectionne le triplet $\{d, T_h^i, T_b^i\}$ permettant d'obtenir l'image C^i la plus proche possible (c'est à dire avec le moins de "faux négatifs" et de "faux positifs") de "l'étalon-or" estimé.

Ce sont ces paramètres qui serviront pour la suite de la procédure. On peut imaginer sans difficulté cependant que la scène chirurgicale change au cours de la procédure, et que les paramètres définis initialement ne soient plus optimaux. Cette méthode d'estimation des paramètres, basée sur la capture d'une image, pourrait alors s'effectuer en tâche de fond par rapport au reste de la procédure de suivi d'instrument, permettant une mise à jour régulière des paramètres (durée complète de l'estimation des paramètres pour 75 combinaisons : 15 secondes). Cette fonctionnalité n'est cependant pas encore implémentée dans le dispositif : si le chirurgien s'aperçoit que le suivi n'est plus efficace après un changement important de scène chirurgicale, il doit pour le moment relancer une sélection des paramètres de façon manuelle.

Cette limite est illustrée figure 3.21 pour deux phases différentes d'une prostatectomie radicale laparoscopique (début de procédure et milieu de procédure).

Par rapport aux méthodes existantes (sélection manuelle empirique des paramètres de seuillage dans [Oropesa et al., 2012], apprentissage d'arbres de décisions nécessitant un entraînement sur des données annotées dans [Allan et al., 2012]), l'approche choisie dans ce travail présente l'intérêt d'être objective et de ne pas



(a) (b) (c) (d) (e) (f)

FIGURE 3.20. – À gauche, processus de sélection de l'image EGT parmi les images PGT^j à l'aide de la courbe ROC (a) et du test du χ^2 (c), le couple $\{Se^j, Sp^j\}$ optimal est celui de l'image PGT^{19} (e). À droite, sélection de l'image C^i la plus proche de EGT avec les mêmes méthodes : courbe ROC (b) et χ^2 (d). L'image sélectionnée correspond au triplet $\{d = 3, T_h^i = 9, T_b^i = 6\}$ (f)

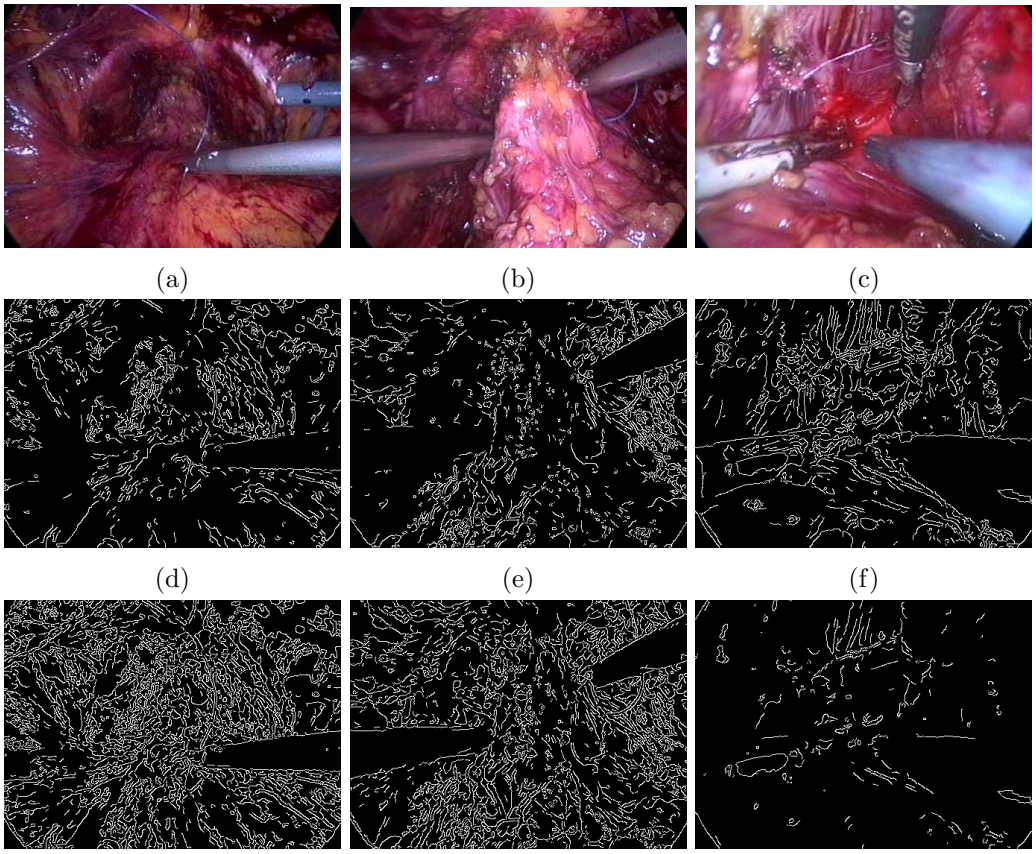


FIGURE 3.21. – À haut à gauche et au milieu, images extraites du milieu d'une procédure de prostatectomie radicale ((a) et (b)), en haut à droite image extraite du début de la même procédure (c). Ces images sont filtrées et les angles sont extraits avec leurs propres paramètres optimaux ((d) : $\{d = 5, T_h^i = 40, T_b^i = 26\}$, (e) : $\{d = 3, T_h^i = 39, T_b^i = 26\}$ et (f) : $\{d = 5, T_h^i = 20, T_b^i = 13\}$), puis avec ceux déterminés pour l'autre phase ((g) et (h) avec paramètres de (c), (i) avec paramètres de (a)). On constate que les paramètres déterminés pour une phase de la chirurgie ne sont pas satisfaisants pour l'autre, mais qu'au cours d'une phase ceux-ci restent adaptés pour plusieurs images différentes.

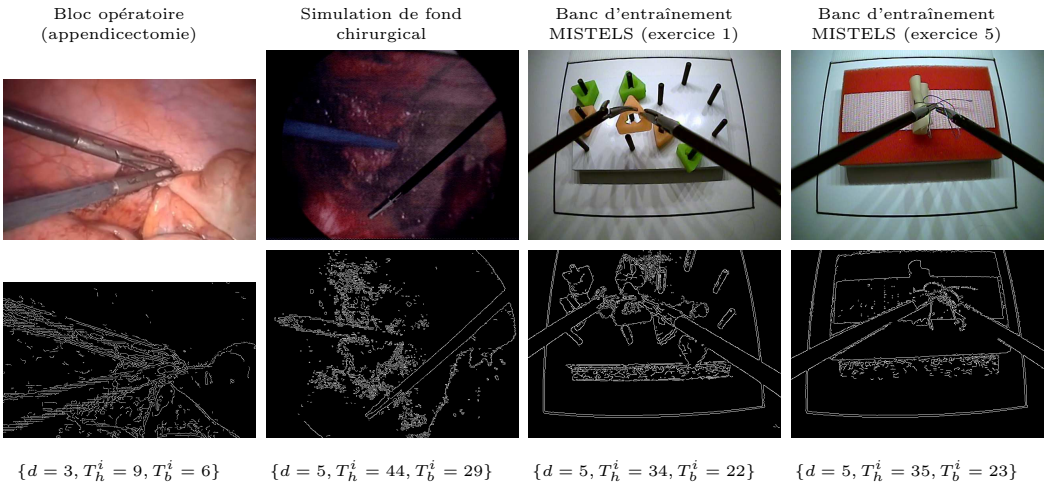


FIGURE 3.22. – Robustesse de la méthode d'optimisation des paramètres des filtres utilisés dans différents contextes d'utilisation

nécessiter de supervision lors de la phase d'apprentissage grâce à l'analyse statistique choisie. Elle peut donc être utilisée dans différents contextes sans nécessiter d'intervention de l'utilisateur, hormis la sélection d'une image "typique"(c'est à dire contenant la scène chirurgicale et un instrument), pour ajuster ses paramètres. La figure 3.22 illustre cette propriété.

De plus, cette méthode est aussi efficace pour la détection des bords de l'instrument qu'une approche basée sur un seuillage d'Otsu [Otsu, 1979], qui avait été initialement envisagée. En effet, les droites obtenues par l'algorithme de Canny implémenté ici, et celles obtenues après seuillage de l'image (avec un seuil défini à chaque image selon la méthode d'Otsu, puis calcul des gradients), étaient similaires, mais avec un temps de calcul nettement en faveur de l'approche par finalement choisie (le calcul d'un seuil unique pour le seuillage au début de la procédure aboutissant à des résultats insuffisamment robustes).

Comme on peut le voir dans la figure 3.21 par exemple, la recherche de droites dans l'image binaire obtenue après cette étape reste délicate. Une méthode brute consistant à appliquer un filtre de Hough sur l'image toute entière serait trop longue pour un suivi en temps réel, et pourrait sélectionner des droites parasites dans l'image (instrument supplémentaire, fil,...). La connaissance de la position du point d'insertion dans le repère de la caméra permet de restreindre l'espace de recherche des droites et de supprimer la détection de droites parasites [Voros et al., 2006, Doignon et al., 2006]. La section suivante présente mon approche pour déterminer cette position.

3.2.5 Détermination de la position du point d'insertion dans le repère lié à la caméra

Après avoir sélectionné les paramètres de filtrage à partir d'une image acquise au début de la procédure, je cherche à déterminer la position du point d'insertion dans $\Re_c$, le repère lié à la caméra.

Par construction, toutes les droites représentant l'axe de l'instrument dans l'espace passent par ce point (ou dans un voisinage de ce point en considérant les petits mouvements que peut faire le point d'insertion autour de sa position moyenne). Cette hypothèse est validée par Voros et al. [2006], qui montrent que durant une procédure de laparoscopie sur un spécimen anatomique frais la position du point d'insertion varie de 5mm en moyenne. Le modèle géométrique de l'instrument présenté ci-dessus permet, à partir de deux droites définissant les bords de l'instrument dans l'image (plus précisément à partir de deux paires de points définissant ces droites, éq. 3.2), de retrouver la droite en trois dimensions formant l'axe central de l'instrument.

J'isole ces droites par une transformée de Hough probabiliste [Matas et al., 2000] appliquée à l'image binarisée après filtrage par Canny. J'utilise l'implémentation de cette transformée dans OpenCV, qui permet de sélectionner les lignes candidates selon trois critères : un seuil pour la valeur associée au couple (ρ, θ) décrivant une ligne dans la matrice d'accumulation, une longueur minimale pour qu'une ligne soit retenue, et un intervalle maximal pour relier entre eux deux points formant une ligne. J'optimise ces paramètres de façon récursive pour n'obtenir que deux droites candidates. Le temps de calcul est d'environ 1 seconde par image traitée.

Dans la pratique, l'utilisateur sélectionne un ensemble d'images représentant le site opératoire, sans bouger la caméra, et avec différentes orientations de l'instrument dans ces images. La validation des deux droites candidates pour chaque image nécessite l'intervention de l'utilisateur, qui confirme ou non que ces droites correspondent bien aux bords de l'instrument. Le taux de reconnaissance est variable en fonction du contexte, proche de 100% sur fond uni ou sur un "faux fond chirurgical" (impression papier d'une capture d'écran d'une prostatectomie radicale), plus proche de 50% en conditions réelles.

Les deux segments de droites sélectionnés sont décrits par la position de leurs pixels aux extrémités. La méthode présentée ci-dessus permet donc de calculer la position de l'axe central de l'instrument dans l'espace. La répétition de cette identification pour plusieurs orientations de l'instrument permet d'obtenir un ensemble de droites $D^i : (P^i, \mathbf{e_1^i})$.
Le point d'insertion T est défini comme le point minimisant la distance

$$d = \sum_i d(T, D^i) = \sum_i \left|\left| \overrightarrow{TP^i} \otimes \mathbf{e_1^i} \right|\right|$$

En notant $\mathbf{T} = (X_T, Y_T, Z_T)^T$, $\mathbf{P^i} = (X_P^i, Y_P^i, Z_P^i)^T$ et $\mathbf{e_1^i} = (e_{1x}^i, e_{1y}^i, e_{1z}^i)^T$, mini-

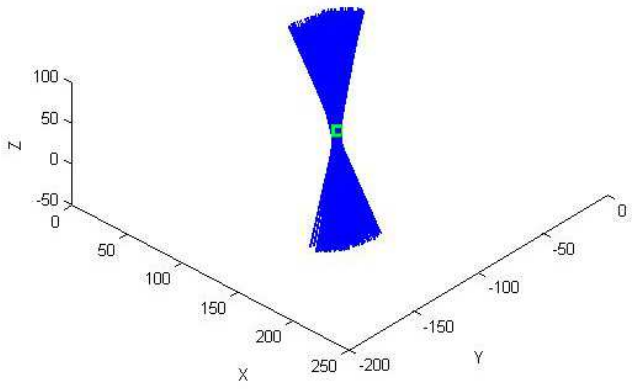


FIGURE 3.23. – Détermination du point d'insertion à l'intersection d'un ensemble de droites représentant l'axe de l'instrument détectées par Polaris (874 droites utilisées). Le point d'insertion est symbolisé par le carré vert.

miser d revient à résoudre le système linéaire :

$$A\mathbf{X} - B = 0 \tag{3.6}$$

$$\text{avec} \quad A = \begin{bmatrix} \sum_i 1 - e^i_{1x} & -\sum_i e^i_{1x}.e^i_{1y} & -\sum_i e^i_{1x}.e^i_{1z} \\ -\sum_i e^i_{1x}.e^i_{1y} & \sum_i 1 - e^i_{1y} & -\sum_i e^i_{1y}.e^i_{1z} \\ -\sum_i e^i_{1x}.e^i_{1z} & -\sum_i e^i_{1y}.e^i_{1z} & \sum_i 1 - e^i_{1z} \end{bmatrix}$$

$$\text{et} \quad B = \begin{bmatrix} \sum_i e^i_{1x}.[\mathbf{e^i} \cdot \mathbf{P^i}] - X^i_P \\ \sum_i e^i_{1y}.[\mathbf{e^i} \cdot \mathbf{P^i}] - Y^i_P \\ \sum_i e^i_{1z}.[\mathbf{e^i} \cdot \mathbf{P^i}] - Z^i_P \end{bmatrix}$$

La résolution de ce système linéaire est effectuée par une décomposition en valeurs singulières (SVD) avec la méthode proposée par OpenCV. La figure 3.23 illustre cette méthode dans des conditions optimales, c'est à dire en localisant l'axe de l'instrument avec un dispositif de suivi de type Polaris plutôt qu'avec une détection dans l'image, afin d'obtenir un très grand nombre de droites et d'éviter les mesures aberrantes.

J'utilise un algorithme de RANSAC (*RANdom SAmple Consensus*, très utilisé dans la communauté de vision par ordinateur) afin d'exclure les droites aberrantes qui ne passeraient pas par le point d'insertion, correspondant à une mauvaise qualité de détection des bords, et qui sont donc susceptibles de fausser la détermination de ce point d'insertion lors de la résolution système linéaire (3.6). Le principe de cet algorithme est expliqué en détail par Fischler et Bolles [1981], je le présente brièvement dans algorithme 3.2 avec les paramètres choisis pour la situation à laquelle je m'intéresse. La figure 3.24 illustre cette méthode dans les conditions d'utilisation. On note que la détermination est précise dans le plan xy de $\Re_c$ (mouvements laté-

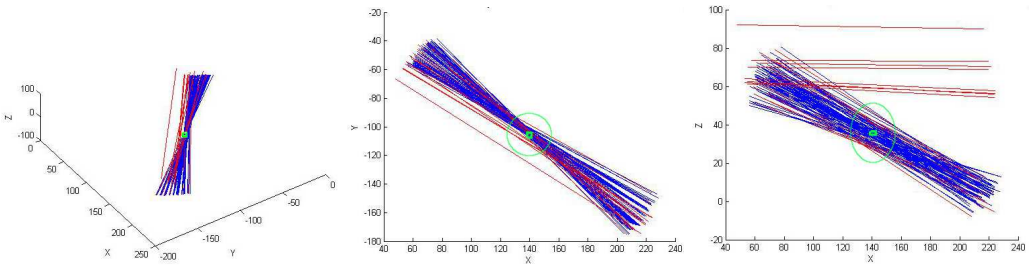


FIGURE 3.24. – Détermination du point d'insertion par analyse d'image et sélection des droites candidates par RANSAC. Les droites bleues sont les droites sélectionnées, tandis que les droites rouges sont celles rejetées par l'algorithme. Le point d'insertion évalué est le carré vert, le cercle vert correspond à la projection dans le plan d'une sphère de 15mm centrée sur ce point.

raux de l'instrument), mais que l'incertitude est plus importante selon la coordonnée z. L'effet de l'algorithme est bien visible avec l'exclusion des mesures aberrantes.

Le paramètre *taille_set* correspond au nombre de droites minimal pour initialiser le modèle (c'est à dire calculer un point d'insertion), et est donc égal à 2.

Le paramètre *tolerance* correspond à la distance seuil que l'on se fixe pour considérer qu'une droite s'intègre au modèle. Les manipulations préliminaires montrent qu'avec la méthode de sélection automatique des bords de l'instrument décrite ci-dessus, l'écart-type de la distance entre une droite et le point d'insertion (calculé en sélectionnant les meilleures droites après plusieurs itérations préalables de l'algorithme de RANSAC avec une tolérance de 10mm) est de $8,85mm$. Fischler et Bolles [1981] proposent de fixer le paramètre *tolerance* entre σ et 2σ, d'où $tolerance \approx 15mm$.

Le paramètre *max_iter* correspond au nombre d'essais maximum que l'algorithme réalise pour trouver un ensemble de droites satisfaisant les conditions du modèle. Fischler et Bolles [1981] proposent de fixer cette valeur à $E(k) + 2\sigma(k)$, où k est le nombre d'essais attendus pour trouver un jeu de n droites représentant des données correctes ($n = taille_set$). Alors $E(k) = w^{-n}$ et $\sigma(k) = \sqrt{1 - w^n}/w^n$, où w est la probabilité qu'un élément quelconque du jeu de données soit dans l'intervalle de tolérance défini plus haut. On estime $w = 0.3$ (situation défavorable avec de nombreuses mesures aberrantes), On a donc $max_iter = E(k) + \sigma(k) \approx 33$.

Enfin le paramètre *taille_min* correspond au nombre minimum de données qui doivent correspondre au modèle pour que celui-ci soit considéré comme correct. Cette taille minimale doit permettre de ne pas sélectionner un modèle erroné, qui serait basé sur un ensemble de données aberrantes mais comprises dans l'intervalle de tolérance de ce modèle. En notant y la probabilité qu'une donnée quelconque soit dans l'intervalle de tolérance d'un modèle incorrect, avec $y < w$, on cherche à

Algorithme 3.2 : Algorithme de RANSAC

Entrées : Un ensemble de N droites $\{D^i\}_{i\in[\![1,N]\!]}$ avec $N > 2$
Sorties : Le point d'insertion $iPoint$

$taille_set \leftarrow 2$
$tolerance \leftarrow 15$
$max_iter = 33$
$taille_min \leftarrow 7$
$erreur \leftarrow \infty$

pour $k = 1 \rightarrow max_iter$ **faire**
  créer un sous-ensemble S_k de $taille_set$ droites sélectionnées aléatoirement parmi $\{D^i\}_{i\in[\![1,N]\!]}$
  estimer le point d'insertion T en utilisant l'équation 3.6 pour les droites dans S_k
  pour $i/D^i \notin S_k$ **faire**
    calculer $d(T, D^i)$
    si $d(T, D^i) < tolerance$ **alors**
      ajouter D^i à S_k
    fin
  fin
  si $\left(\text{card}(S_k) > taille_min \text{ et } \frac{\sum_{i/D^i\in S_k}\left(d(T,D^i)\right)^2}{\text{card}(S_k)} < erreur\right)$ **alors**
    $erreur \leftarrow \frac{\sum_{i/D^i\in S_k}\left(d(T,D^i)\right)^2}{\text{card}(S_k)}$
    $iPoint \leftarrow T$
  fin
fin
retourner $iPoint$

rendre la probabilité que $taille_min - taille_set$ droites supplémentaires s'ajustent au modèle la plus faible possible ($y^{(taille_min - taille_set)} < 0,01$). Une valeur de 7 pour $taille_min$ remplit ce critère.

En cas de droites parasites impossibles à supprimer (droite des supports plastique dans le cas d'un banc d'entraînement par exemple) ou dans le cas de performances non satisfaisantes de la méthode de sélection automatique des bords de l'instrument, il est possible de définir les droites formant ces derniers à la main. Cette solution est cependant difficilement compatible avec les conditions du bloc opératoire.

3.2.6 Représentation spatiale de l'orientation de l'instrument

i) Représentation discrète de l'orientation sur une sphère

Comme illustré sur la figure 3.5, je cherche, après avoir déterminé la position du point d'insertion dans le repère de la caméra, à représenter de façon discrète l'espace des orientations possibles de l'instrument autour de ce point (figure 3.25).

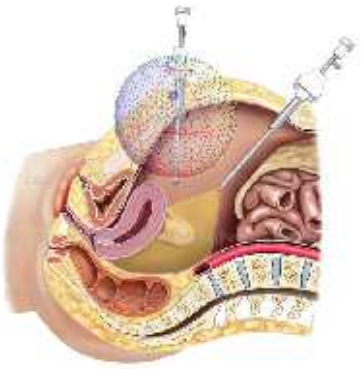
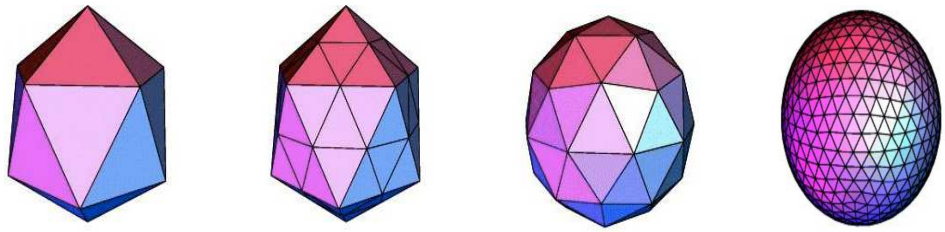


FIGURE 3.25. – Orientations possibles de l'instrument autour du point d'insertion, représentées par la grille géodésique sphérique présentée dans cette section.

Le système de coordonnées le plus naturel pour remplir cette tâche est une représentation des orientations par le couple d'angle (θ, λ) représentant la longitude et la latitude dans un système de coordonnées sphériques ayant pour origine le point d'insertion. Cette représentation n'est cependant pas adéquate car elle ne satisfait pas à plusieurs propriétés que l'on souhaite respecter [Heikes et Randall, 1995] :

l'homogénéité : la distance entre deux points voisins sur la sphère ne doit pas varier en fonction de la position de ces points. Or, si la distance latitudinale (distance entre deux points voisins situés à la même longitude) peut être



(a) (b) (c) (d)

FIGURE 3.26. – Construction itérative de la sphère géodésique : à partir d'un icosaèdre inscrit dans une sphère de rayon unité (a), chaque face triangulaire est subdivisée en quatre nouveaux triangles (b), et les nouveaux vertex sont projetés sur la surface de la sphère unité (c). Cette étape est répétée de façon itérative jusqu'à atteindre la résolution souhaitée. À droite la sphère géodésique après 3 itérations (d). Illustration extraite de [Ringler et al., 1999]

constante, la distance longitudinale est une fonction de la latitude, c'est à dire qu'elle décroît lorsque la latitude de deux points voisins augmente.

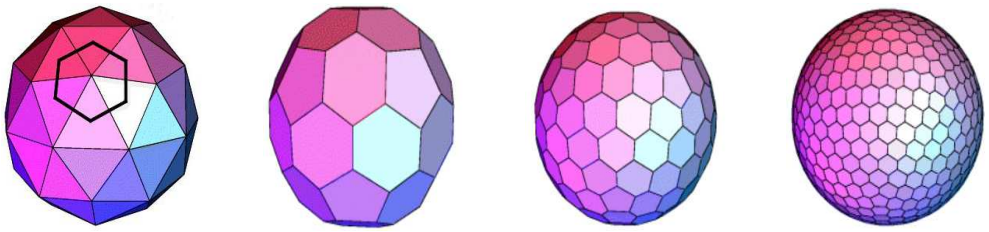
l'isotropisme : la distance entre deux points voisins ne doit pas dépendre de la direction entre ces deux points. Dans le système de coordonnées sphériques, la résolution horizontale n'est pas identique à la résolution verticale (celle-ci peut être fixe, la résolution horizontale varie avec la latitude).

la différentiabilité en tout point : la longitude présente une discontinuité aux pôles, ce qui rend le calcul de distances d'un point à l'autre plus complexe, bien que possible mais au prix d'une complexification importante de l'implémentation.

Pour pallier ces limitations liées au système de coordonnées sphériques, Heikes et Randall [1995] proposent de discrétiser la surface de la sphère à l'aide d'une grille géodésique sphérique pour étudier des modèles météorologiques. Ce type de représentation présente l'intérêt d'être quasi-homogène et quasi-isotrope pour la résolution, en plus d'être facilement adaptable à de multiples résolutions.

Cette grille géodésique sphérique est construite à partir d'un icosaèdre centré sur le point d'insertion, en subdivisant de façon itérative chacune de ses faces triangulaires en quatre nouveaux triangles, puis en projetant les nouveaux vertex sur une sphère (figure 3.26).

Chaque vertex V_i est alors considéré comme une orientation discrète possible de l'instrument, centre d'une cellule de Voronoi de forme hexagonale (cellule comprenant tous les points de la sphère plus proches de V_i que de tout autre vertex, figure 3.27a). Seules les cellules situées aux vertex originaux de l'icosaèdre sont différentes dans la mesure où il s'agit de pentagones. La grille géodésique obtenue est illustrée



(a) (b) (c) (d)

FIGURE 3.27. – Pavage de la grille sphérique géodésique avec des cellules de Voronoi. (a) : cellule élémentaire centrée sur un vertex ; (b-d) : grilles d'ordre 1, 2 et 3. Illustration extraite de [Ringler et al., 1999]

figure 3.27b-3.27d pour des nombres d'itérations variables. Je définis l'ordre de la grille par le nombre d'itérations nécessaires à sa construction (la grille constituée de 12 pentagones à partir de l'icosaèdre est d'ordre 0).

Le nombre de cellules total dans la grille est $N_c = 5(2^{2q+1}) + 2$ avec q le nombre d'itérations.

ii) Structure des données

Cette représentation des orientations de l'instrument autour du point d'insertion présente l'intérêt de permettre une implémentation aisée sous forme de matrices rectangulaires pour manipuler les données. En effet la grille géodésique peut être partitionnée, quel que soit son ordre, en 12 triangles dont les extrémités sont les vertex de l'icosaèdre original (figure 3.28a). En groupant ces triangles, on obtient cinq panneaux (figure 3.28c), qui peuvent être déformés pour se présenter sous forme de rectangles, dont les cellules sont disposées en lignes et en colonnes (figure 3.28d). Le cas des pôles, qui n'appartiennent initialement à aucun rectangle, peut facilement être traité en ajoutant une ligne et une colonne à chaque rectangle, les autres éléments de ces lignes et colonnes ne contenant pas de données (figure 3.28b).

À chaque orientation possible de l'instrument autour du point d'insertion peut donc être associée une cellule, repérée par les indices $\{i, j\}$, d'un panneau rectangulaire p de dimension $(2^q + 1) * (2^{q+1} + 1)$ (figure 3.28e).

Je crée pour chacun des panneaux p :

– une matrice de *correspondance*, C^p, dont chaque élément C^p_{ij} contient un vecteur représentant l'orientation de l'instrument passant par la cellule associée sur la grille géodésique. Ce vecteur est décrit en coordonnées cartésiennes dans le repère de la caméra $\Re_c$. Cette matrice de *correspondance* permet de simplifier la représentation des orientations (chaque orientation possible sur la grille géodésique est associée à un unique triplet $\{p, i, j\}$ avec p le numéro du panneau et $i \in [\![1, 2^q + 1]\!], j \in [\![1, 2^{q+1} + 1]\!]$ les indices de la cellule dans ce panneau) tout en

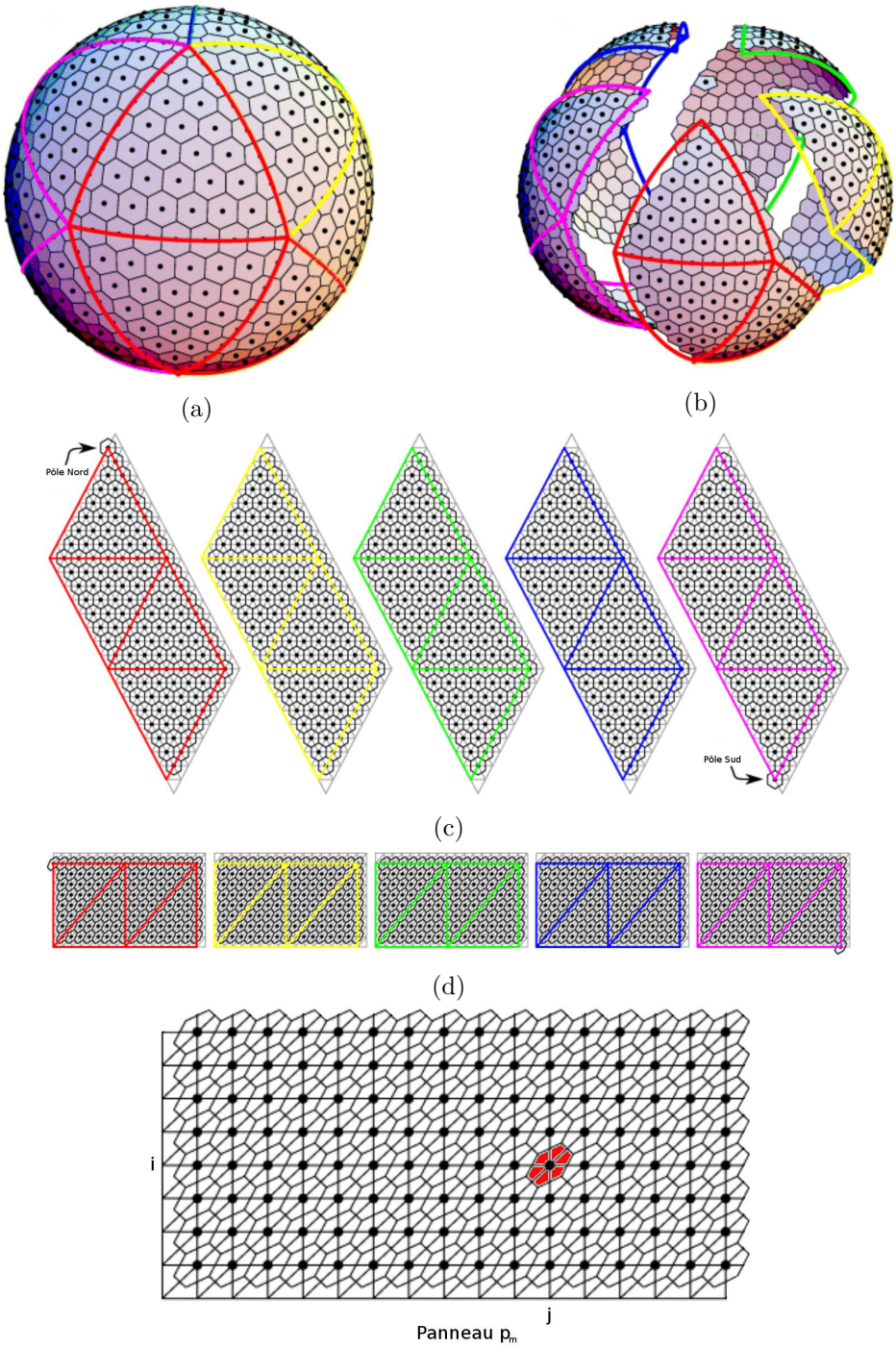


FIGURE 3.28. – Représentation de la partition de la grille géodésique sphérique d'ordre 3 en panneaux de cellules formant des matrices de données. Illustration extraite de [Ringler et al., 1999]

Profondeur d'insertion (en mm)	Ordre de la grille						
	3	4	5	6	7	8	9
100	15.03	7.53	3.76	1.88	0.94	0.47	0.24
150	22.55	11.29	5.65	2.82	1.41	0.71	0.35
200	30.07	15.05	7.53	3.76	1.88	0.94	0.47
250	37.58	18.81	9.41	4.71	2.35	1.18	0.59
300	45.10	22.58	11.29	5.65	2.82	1.41	0.71

Tab. 3.1. – Distance (en mm) entre deux points de l'espace correspondant à une profondeur donnée de l'instrument, dont l'orientation est représentée par deux cellules voisines de la grille géodésique, en fonction du nombre d'itérations utilisées pour sa construction.

gardant un accès rapide au vecteur d'orientation $\mathbf{e_1}$ dans $\Re_c$ stocké dans C_{ij}^p. Dans la suite, j'adopte l'équivalence $\mathbf{e_1} \Leftrightarrow \{p, i, j\}$ (équivalence stricte, chaque pôle n'étant représenté que dans un panneau).

– une matrice *mesure*, Me^p, dont les éléments Me_{ij}^p contiennent la valeur de la mesure de probabilité que l'instrument soit dans l'orientation associée à C_{ij}^p (mesure décrite section 3.2.7) ;
– une matrice *masque*, Ma^p, permettant d'exclure ($Ma_{ij}^p = -1$) de l'analyse les orientations pour lesquelles la demi-droite décrite par $(T, \mathbf{e_1})$ (où T est le point d'insertion) ne possède pas d'intersection avec le champ de vision de la caméra. Cette matrice permet de restreindre l'espace de recherche des orientations et d'accélérer le suivi d'instrument dans la suite. Le masque est mis à jour à chaque déplacement de la caméra (c'est à dire pour chaque déplacement du point d'insertion dans $\Re_c$) par une sous-procédure spécifique (temps de calcul dépendant de la résolution de la grille).

Cette modélisation des orientations possibles de l'instrument autour de son point d'insertion pourrait permettre de mesurer pour chaque cellule $\{p, i, j\}$ telle que $Ma_{ij}^p \neq -1$ une probabilité que l'instrument suive l'orientation associée. Cette approche est cependant sensible aux fausses détections et nécessite de réaliser un nombre toujours trop important de mesures.

En effet, le nombre de cellules de la grille géodésique conditionne la qualité de la mesure. La résolution entre deux positions voisines de l'instrument, pour une profondeur d'insertion donnée, est dépendante de cette dernière et de l'ordre de la grille (tableau 3.1).

En considérant que le degré de précision recherché correspond environ à un rayon d'instrument (2.5mm) au niveau de la pointe de celui-ci, on note que pour différencier deux orientations voisines de l'instrument lorsque celui-ci est enfoncé au maximum (300mm), la grille géodésique doit être au minimum d'ordre 7 (soit 163842 cellules). Malgré la restriction de l'espace de recherche aux orientations possédant

une intersection avec le champ de vision de la caméra, le nombre de cellules à explorer reste trop important pour réaliser une mesure brute.

Je propose pour sélectionner les cellules à explorer l'utilisation d'un algorithme de CONDENSATION, dont j'explique le principe et l'implémentation dans la section suivante.

3.2.7 Algorithme de CONDENSATION

i) Principe général

Comme suggéré initialement par Duchateau et al. [2007], la réduction de l'espace de recherche est basée sur l'algorithme de CONDENSATION, proposé par Isard et Blake [1998a]. Cet algorithme de propagation de densité de probabilité utilise un échantillonnage pondéré des orientations. Chaque orientation échantillonnée au temps t est représentée par une particule $\mathbf{s}_t^{(n)}$, affectée d'un poids $\pi_t^{(n)}$ correspondant à la mesure de probabilité que l'instrument suive cette orientation. L'ensemble des particules $\{\mathbf{s}_t^{(n)}\}_{n\in[\![1,N]\!]}$, où N est le nombre de particules utilisées, décrit donc une densité de probabilité de l'orientation de l'instrument au temps t.

L'ensemble des particules $\{\mathbf{s}_t^{(n)}\}_{n\in[\![1,N]\!]}$ évolue en fonction du temps selon un modèle dynamique, et les poids $\pi_t^{(n)}$ sont actualisés à chaque étape en fonction des observations.

Plus précisément, à partir de la distribution $\{\mathbf{s}_t^{(n)}, \pi_t^{(n)}\}_{n\in[\![1,N]\!]}$ au temps t, on sélectionne N nouvelles particules $\mathbf{s}_{t+1}^{(n)}$ parmi $\{\mathbf{s}_t^{(n)}\}_{n\in[\![1,N]\!]}$ avec une probabilité $\frac{\pi_t^{(n)}}{\sum_n^N \pi_t^{(n)}}$ de tirer la particule $\mathbf{s}_t^{(n)}$. Ainsi une particule $\mathbf{s}_t^{(n)}$ avec un poids $\pi_t^{(n)}$ important pourra être représentée plusieurs fois dans le nouvel ensemble $\{\mathbf{s}_{t+1}^{(n)}\}_{n\in[\![1,N]\!]}$. Les probabilités a priori $\pi_{t+1}^{(n)}$ sont toutes égales à 1.

Les particules $\mathbf{s}_{t+1}^{(n)}$ se propagent alors dans l'espace d'évolution admissible en obéissant à un modèle dynamique comportant une composante déterministe et une composante stochastique. Une nouvelle mesure est effectuée pour les nouvelles positions des particules, et les poids $\pi_{t+1}^{(n)}$ sont mis à jour en fonction de cette mesure (probabilité a posteriori).

La figure 3.29 résume le principe général de l'algorithme de CONDENSATION dans le cas général.

Cet algorithme présente l'avantage de pouvoir représenter plusieurs hypothèses concurrentes alternatives, c'est à dire que même si la densité de probabilité au temps t n'est pas maximale pour la particule correspondant effectivement à l'instrument (sans être nulle), cette particule ne sera pas éliminée de la suite de l'analyse, et pourra être sélectionnée pour former de nouvelles particules au temps $t+1$. Le filtrage de la trajectoire, en temps réel [Isard et Blake, 1998b] ou a posteriori dans

ce travail (voir chapitre 4), permet alors d'exclure les mesures aberrantes et de relier les mesures correctes par interpolation.

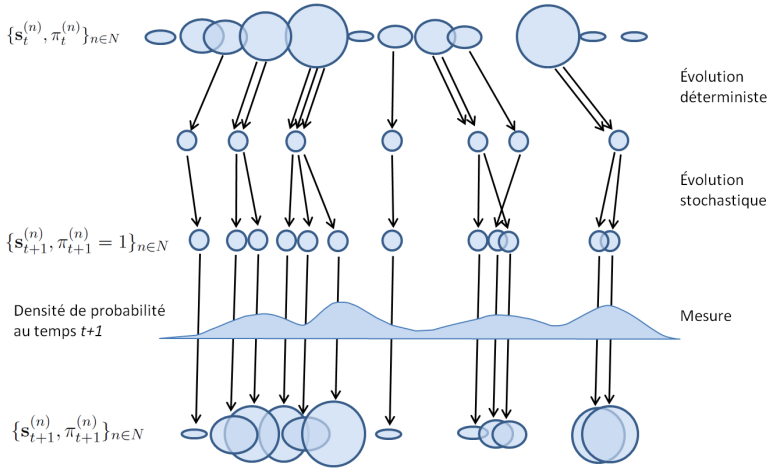


FIGURE 3.29. – Cas général de l'algorithme de CONDENSATION. Ensemble de particules décrivant une position dans un espace à une dimension. Illustration inspirée par [Isard et Blake, 1998a].

ii) Modèle particulaire de l'orientation de l'instrument

Dans mon modèle, chaque particule $\mathbf{s}_t^{(n)}$ est définie au temps t par cinq variables : $\mathbf{s}_t^{(n)} = \{x_{t-1}^{(n)}, y_{t-1}^{(n)}, x_t^{(n)}, y_t^{(n)}, p_t^{(n)}\}$, où $x_{t-1}^{(n)}$ et $y_{t-1}^{(n)}$ sont les positions au temps $t-1$ de la particule dans le repère lié au panneau $p_t^{(n)}$, $x_t^{(n)}$ et $y_t^{(n)}$ sont les positions courantes de la particule dans le repère lié panneau $p_t^{(n)}$.
On note $\mathbf{X}_t^{(n)} = [x_{t-1}^{(n)} \quad y_{t-1}^{(n)} \quad x_t^{(n)} \quad y_t^{(n)}]^T$.

Le modèle d'évolution est le suivant :

évolution déterministe : on considère que le mouvement au cours du geste chirurgical correspond à un mouvement dont la vitesse est de la forme $\mathbf{v_{t+1}} = \delta * \mathbf{v_t}$ (c'est à dire que la vitesse est exponentiellement décroissante avec le temps). Cette hypothèse correspond bien à un geste pour lequel le chirurgien s'approche d'une cible en ralentissant, et présente l'avantage de ne pas nécessiter la connaissance d'états antérieurs à $\mathbf{s}_t^{(n)}$ pour calculer $\mathbf{s}_{t+1}^{(n)}$:

$$\mathbf{X}_{t+1}^{(n)} = A\mathbf{X}_t^{(n)} \quad \text{avec} \quad A = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -\delta & 0 & 1+\delta & 0 \\ 0 & -\delta & 0 & 1+\delta \end{bmatrix}$$

Les essais préliminaires montrent que ce modèle permet de suivre correctement l'instrument avec une valeur de δ fixée à 0.3.

évolution stochastique : le processus de diffusion stochastique doit avoir un comportement isotrope sur la surface de la sphère (pas de direction du mouvement privilégiée). On modélise ce comportement isotrope par une loi normale multi-dimensionnelle sur les trois directions (i, j, k) de la grille hexagonale d'un panneau avant déformation (figure 3.30 à gauche). On a donc la densité de probabilité jointe de (i, j, k) :

$$p(i,j,k) = \frac{1}{(2\pi)^{3/2}} \mathrm{e}^{-\frac{1}{2}(i^2+j^2+k^2)}$$

En notant qu'après déformation des panneaux pour obtenir des rectangles, on a $i = x$, $j = x + y$ et $k = y$ (voir figure 3.30b), on peut ré-écrire la densité de probabilité jointe en fonction de (x, y) :

$$p(x,y) = \frac{1}{(2\pi)^{3/2}} \mathrm{e}^{-\frac{1}{2}(x^2+(x+y)^2+y^2)}$$

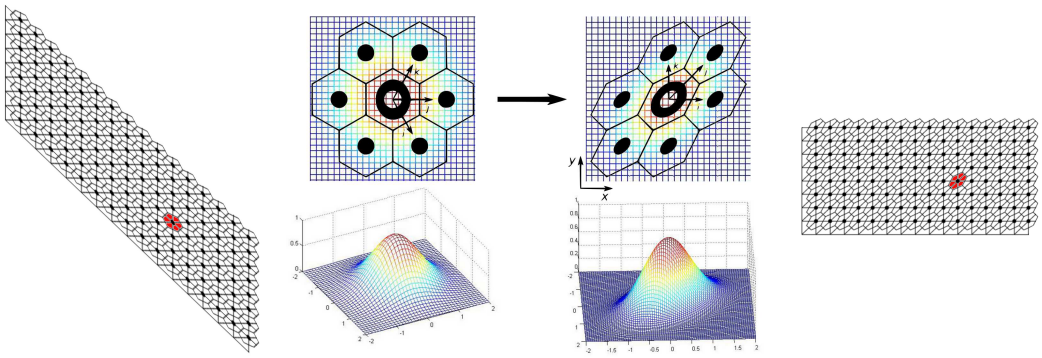
$$\text{soit} \quad p(\mathbf{w}) = \alpha \frac{1}{(2\pi)^{N/2}\sqrt{|\Sigma|}} \mathrm{e}^{-\frac{1}{2}(\mathbf{w}^T \Sigma^{-1} \mathbf{w})}$$

où $p(\mathbf{w})$ est la forme générale de la densité de probabilité d'une loi normale multi-dimensionnelle avec $\mathbf{w}$ le vecteur de variables aléatoires de dimension N, dont la distribution a une moyenne nulle et une matrice de variance-covariance Σ. Dans le cas étudié ici on a donc :

$$\mathbf{w} = [x \quad y]^T \quad \text{avec } N = 2,$$
$$\Sigma = \begin{bmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{bmatrix},$$
$$|\Sigma| = 1/3 \quad \text{le déterminant de la matrice } \Sigma$$
$$\text{et} \quad \alpha = \sqrt{6\pi} \quad \text{un coefficient de normalisation}$$

À chaque temps t, on construit à partir du vecteur $\mathbf{w_t} = [X_1 \quad X_2]^T$, où les variables aléatoires X_1 et X_2 suivent une loi normale centrée réduite, le vecteur $B\mathbf{w_t}$ qui suit la loi de probabilité $p(\mathbf{w_t})$ recherchée, avec B tel que $B^T B = \Sigma^{-1}$ soit $B = \begin{bmatrix} 1 & (1-\sqrt{3})/2 \\ 1 & (1+\sqrt{3})/2 \end{bmatrix}$. La composante stochastique du mouvement à un instant donné est donc $\beta B \mathbf{w_t}$ où β est un coefficient décrivant le poids relatif de la composante stochastique par rapport à la composante déterministe, mis à jour au cours du suivi en fonction de la stabilité de la détection (augmenté en cas de vitesse angulaire importante de l'axe de l'instrument, diminué en cas d'axe constant).

Pour s'adapter à la dimension du vecteur $\mathbf{X}_t^{(n)}$, on note dorénavant $B \leftarrow \begin{bmatrix} 0 & 0 \\ 0 & B \end{bmatrix}$ et $\mathbf{w_t} \leftarrow [0 \quad 0 \quad \mathbf{w_t}^T]^T$.



(a) (b) (c)

FIGURE 3.30. – Passage du pavage hexagonal original (a) au pavage utilisé dans les panneaux rectangulaires (b). La densité de probabilité est modifiée par le changement de système de coordonnées (c) pour que le mouvement stochastique conserve sa caractéristique d'isotropisme à la surface de la sphère modélisée par la grille malgré la déformation de cette dernière.

L'évolution dynamique des particules est donc décrite par l'équation

$$X_{t+1}^{(n)} = A\mathbf{X}_t^{(n)} + \beta B\mathbf{w_t} \tag{3.7}$$

à chaque itération de l'algorithme.

À chaque itération, $x_t^{(n)}$ et $y_t^{(n)}$ sont ramenées aux valeurs entières les plus proches pour correspondre aux colonnes et rangs des matrices définies plus haut.
Les changements de panneaux (lorsque $x_t^{(n)} \notin [\![1, (2^{q+1}+1)]\!]$ ou $y_t^{(n)} \notin [\![1, (2^q+1)]\!]$) sont contrôlés par une sous-procédure qui effectue les changements de repère nécessaires pour replacer la particule sur le bon panneau. De même une autre sous-procédure empêche les particules d'évoluer vers des orientations qui sont exclues dans la matrice *masque*.

iii) Mise à jour des poids des particules

À chaque temps t j'ai donc un ensemble $\{\mathbf{s}_t^{(n)}\}_{n\in[\![1,N]\!]}$, dont les particules ont été sélectionnées parmi $\{\mathbf{s}_{t-1}^{(n)}\}_{n\in[\![1,N]\!]}$ en fonction des poids $\{\pi_{t-1}^{(n)}\}_{n\in[\![1,N]\!]}$, et qui ont évolué selon le modèle dynamique décrit ci-dessus. La probabilité a priori que l'orientation réelle de l'instrument corresponde à l'orientation décrite par une particule $\mathbf{s}_t^{(n)}$ est égale à $1/N$. L'observation permettant de calculer la probabilité a posteriori est effectuée grâce l'image binaire C obtenue après filtrage bilatéral et

détection des angles selon l'algorithme de Canny ainsi qu'aux images de Sobel S_x et S_y ayant permis l'application de l'algorithme de Canny.

Pour chaque particule $\mathbf{s}_t^{(n)}$ le vecteur $\mathbf{e_1}^{(n)}$ est extrait de la matrice *correspondance*, et les droites $\Gamma_1^{(n)}$ et $\Gamma_2^{(n)}$ décrites par l'équation (3.3) de la section 3.2.3 sont définies avec $\gamma \in [0, 400]$ (on décrit les bords d'un instrument de 400mm de long au maximum). Les droites $D_1^{(n)}$ et $D_2^{(n)}$ correspondant à la projection de ces droites sur le plan image (équation (3.1)) correspondent aux bords attendus de l'instrument décrit par la particule $\mathbf{s}^{(n)}$; les points de ces droites dans C, S_x et S_y sont alors explorés, en partant des points proximaux par rapport au point d'insertion pour aller vers les points distaux.

Pour chaque bord supposé de l'instrument, on calcule le score $\Delta_k^{(n)}$ à partir des images C, S_x et S_y, puis on met à jour le poids $\pi_t^{(n)}$ de la particule $\mathbf{s}_t^{(n)}$ avec la probabilité a posteriori de cette dernière (algorithme 3.3).

Algorithme 3.3 : Calcul du poids $\pi_t^{(n)}$ de la particule $\mathbf{s}_t^{(n)}$

Données : les images C, S_x et S_y
Entrées : les droites $D_1^{(n)}$ et $D_2^{(n)}$
Sorties : le poids $\pi_t^{(n)}$

pour $k = 1 \rightarrow 2$ **faire**
  $\Delta_k^{(n)} \leftarrow 0$ $continuity \leftarrow 0$
  $\mathbf{a} \leftarrow$ vecteur normal à $D_k^{(n)}$
  $iterateur[] \leftarrow$ tableau contenant les coordonnées des points de $D_k^{(n)}$
  pour $i = 1 \rightarrow taille(iterateur)$ **faire**
    si $C_{iterateur[i]} \neq 0$ **alors**
      $poids_local \leftarrow \left| \frac{\mathbf{a} \cdot [S_{x,iterateur[i]}, S_{y,iterateur[i]}]^T}{\left|\left|[S_{x,iterateur[i]}, S_{y,iterateur[i]}]^T\right|\right|} \right|$
      $continuity \leftarrow continuity + poids_local$
      $\Delta_k^{(n)} \leftarrow \Delta_k^{(n)} + continuity$
    sinon
      $continuity \leftarrow 0$
    fin
  fin
fin
$\pi_t^{(n)} = \Delta_1^{(n)} * \Delta_2^{(n)}$
retourner $\pi_t^{(n)}$

On note que l'intensité du gradient n'est pas prise en compte dans ce calcul, seule son orientation (par rapport à la normale au bord attendu de l'instrument dans

l'image) entrant en jeu. Elle intervient en fait de façon détournée dans la sélection des point pour lesquels j'étudie l'orientation du gradient, qui sont ceux correspondant d'après l'algorithme de Canny à des contours significatifs dans l'image.
L'introduction de la variable temporaire *continuity* permet de favoriser les droites constituées de points successifs.

Si plusieurs particules partagent la même position sur la grille géodésique, le score n'est calculé qu'une seule fois afin d'améliorer le temps de calcul.

On obtient finalement un ensemble de particules $\{\mathbf{s}_t^{(n)}\}_{n\in[\![1,N]\!]}$, affectées des poids $\{\pi_t^{(n)}\}_{n\in[\![1,N]\!]}$, qui représentent les orientations supposées de l'instrument autour du point d'insertion. Je choisis comme orientation mesurée de l'instrument l'orientation associée à la particule affectée du score le plus élevé. Les expériences préliminaires ont en effet montré que le choix de l'espérance des orientations $\sum_n^N \pi_t^{(n)}\mathbf{e_1}^{(n)}$ comme orientation mesurée, tel que suggéré par Isard et Blake [1998a], ne permettait pas d'obtenir un suivi de qualité satisfaisante, certaines droites parasites de l'image engendrant une dérive systématique de la mesure. Le choix que je fais diminue la stabilité de la méthode de suivi (sensibilité aux fausses détections), mais cet effet est modeste et peut être contrôlé par l'application d'un filtre sur la séquence des orientations pour en éliminer les mesures aberrantes.

3.2.8 Détermination profondeur d'insertion

i) Partition en classes par l'algorithme d'Otsu

L'orientation de l'instrument ayant été déterminée dans l'espace, je cherche à mesurer le dernier degré de liberté accessible par ma méthode, la valeur de la profondeur d'insertion de celui-ci le long de son axe. Cette mesure est actuellement effectuée en analysant la distribution de caractéristiques de pixels appartenant à des segments de la projection de l'axe central de l'instrument dans l'image.
L'axe central de l'instrument dans le repère cartésien lié à la caméra s'écrit

$$\Gamma_c : \overrightarrow{CT} + \gamma\mathbf{e_1}, \gamma \in [0, 400]$$

Je considère N points uniformément répartis sur la projection D_c de Γ_c dans l'image, et construis N segments de D_c centrés sur ces points, de longueur $2k$.
Pour chaque segment, j'applique un algorithme d'Otsu [Otsu, 1979] sur la valeur des pixels dans l'espace colorimétrique TSV afin de les classer en deux groupes distincts ; je cherche le segment pour lequel il existe la meilleure séparabilité en deux classes de pixels, ce qui correspond en théorie à l'interface instrument/scène chirurgicale. Je définis alors la position de la pointe dans l'image comme le milieu du segment ayant la variance inter-classe maximale parmi les segments explorés.

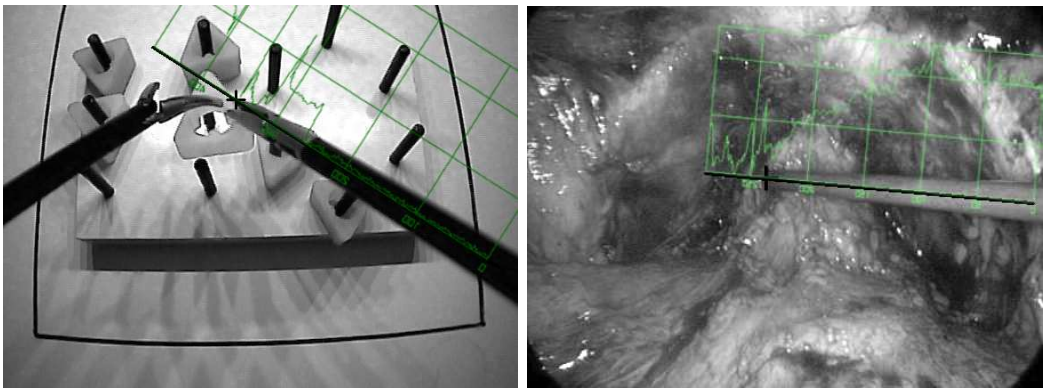
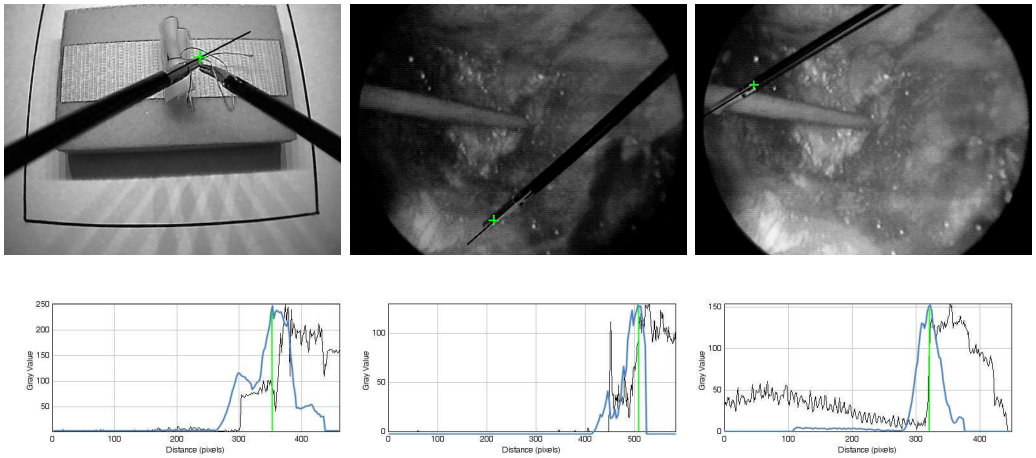


FIGURE 3.31. – Distribution des valeurs des pixels le long de la projection de l'axe central de l'instrument. La croix signale la position estimée de la pointe.

Les figures 3.31 à 3.34 illustrent les résultats obtenus avec cette méthode dans différents contextes.

On note que l'interface détectée par cette approche est dépendante de ce contexte : sur un banc d'essai avec un fond uni très lumineux (figure 3.32a), c'est l'extrémité de la partie métallique qui correspond au segment avec la meilleure variance inter-classe ; dans le contexte d'un fond plus sombre (sur une impression de scène chirurgicale par exemple), c'est l'interface entre le manche et la partie métallique qui sera sélectionnée (figures 3.32b et3.32c). Cependant, au sein d'un contexte d'utilisation, l'interface détectée ne varie pas d'une image à l'autre. La détection est robuste dans ces deux contextes, même en présence de reflets spéculaires (faible taux de fausses détections, qu'il est possible de filtrer en même temps que les fausses détections de l'orientation). La limite de cette approche est la situation où deux instruments se croisent au voisinage de leur pointe, l'interface entre le manche du second instrument et l'image de fond pouvant être confondu avec l'interface recherchée (figure 3.33).

Dans le contexte du bloc opératoire (figure 3.34), la détection de la position de la pointe devient moins robuste. De nombreux types d'instruments sont utilisés, avec autant d'interfaces différentes entre le manche et la pointe, et des pointes plus variées que celles utilisées sur banc d'entraînement (pince fenêtrée, pince diathermique,...). La limitation liée à l'intersection de deux instruments au voisinage de leur pointe reste valable (figure 3.34f). On note cependant que, malgré des imprécisions, la pointe identifiée reste dans le voisinage de la pointe réelle dans le plan de l'image, quel que soit le type de manche utilisé (noir dans les figures 3.34a-3.34c, argenté dans les figures 3.34d et3.34e ou bicolore avec de l'écriture dans la figure 3.34f).



(a) (b) (c)

FIGURE 3.32.

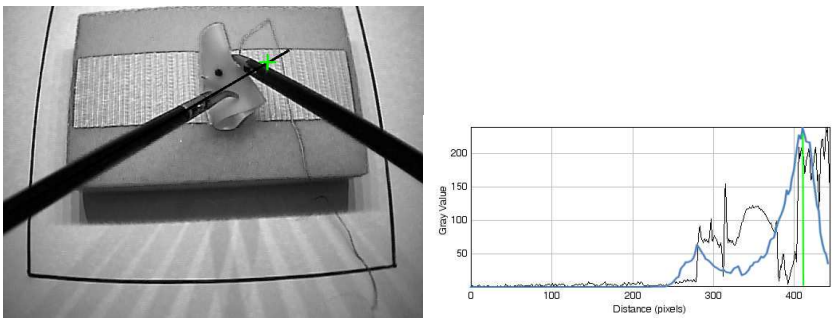
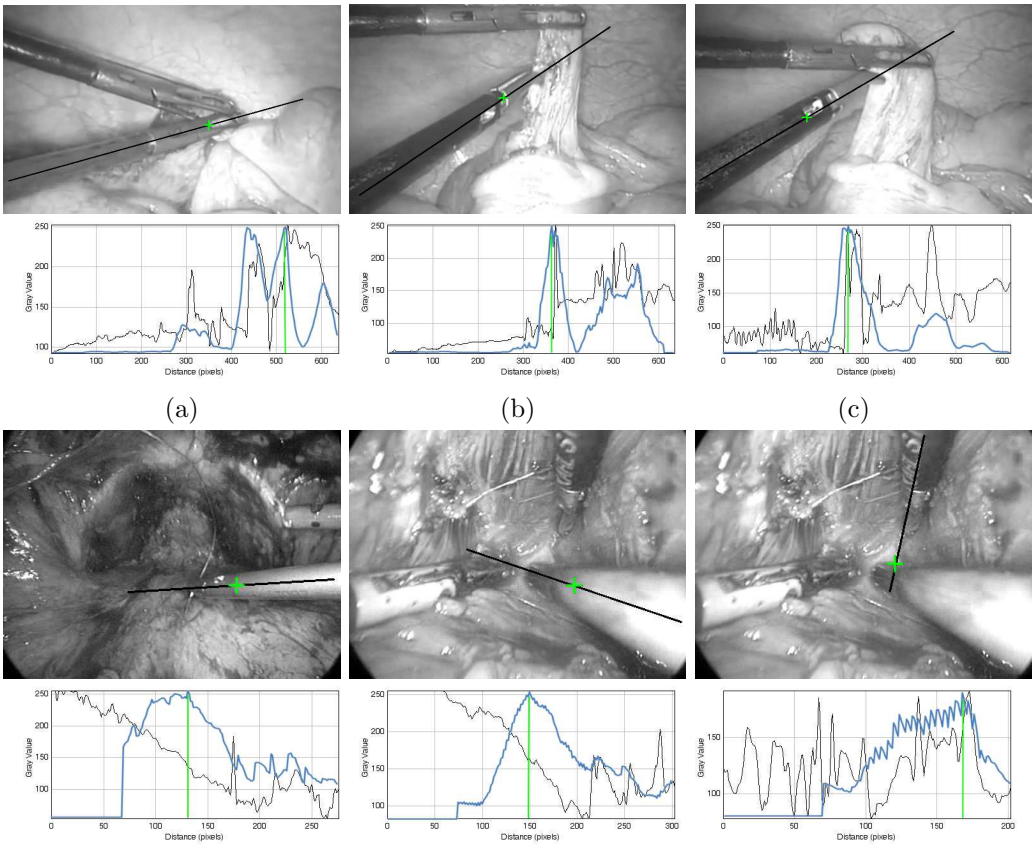


FIGURE 3.33.



(a) (b) (c)

(d) (e) (f)

FIGURE 3.34.

ii) Autres méthodes explorées

Si la méthode présentée plus haut a montré des résultats satisfaisants dans des conditions de banc d'entraînement après filtrage pour éliminer les valeurs aberrantes, elle n'est pas suffisamment robuste pour permettre un suivi dans des conditions réelles de chirurgie. Je présente ici les différentes pistes explorées pour améliorer cette mesure.

▷ **Algorithme de CONDENSATION pour décrire la profondeur d'insertion :** Dans l'optique de réduire les fausses détections, j'ai implémenté un algorithme de CONDENSATION semblable à celui décrit plus haut pour décrire la position de la pointe de l'instrument. La profondeur était mesurée dans l'espace (profondeur d'insertion réelle). Cette approche n'a pas abouti en raison d'une trop forte instabilité de la profondeur détectée. En effet, la moindre fausse détection (en particulier selon z dans $\Re_c$) de l'orientation de l'instrument faisait diverger l'ensemble des particules décrivant la pointe, et le temps de convergence vers la pointe après récupération d'une orientation correcte était trop important.

Une approche pour accroître la stabilité de la mesure serait une description de la position des particules dans le plan image, puis une évaluation de la profondeur correspondante dans l'espace.

▷ **Analyse de la direction des gradients des bords de l'instrument :** En me basant sur les travaux de Climent et Marés [2004], j'ai également évalué la profondeur d'insertion en étudiant la répartition de la direction des gradients le long des bords de l'instrument. Cette approche ne donne pas de résultats satisfaisants, en particulier si les manches des instruments se croisent ou si un instrument s'appuie sur une partie de ses bords sur une structure de même valeur ou saturation (selon l'espace colorimétrique utilisé). De plus la robustesse des résultats est très dépendante de la précision de la détection des bords de l'instrument.

▷ **Analyse de la direction et de la valeur des gradients à l'extrémité de l'instrument :** Une approche consistant à étudier la valeur et l'orientation du gradient dans la direction de l'axe de l'instrument a montré des résultats satisfaisants en considérant des nuages de pixels répartis autour de la position supposée de la pointe. Cette approche est cependant trop coûteuse en temps de calcul pour être utilisée ici.

▷ **Correspondance avec un modèle standard :** Une approche basée sur la méthode de correspondance avec un modèle standard ("*template matching*"), et étudiant la corrélation pixel par pixel entre un modèle de la pointe et un rectangle

se translatant le long de l'axe de l'instrument n'a pas non plus montré de résultats satisfaisants. En effet la pointe de l'instrument étant métallique, sa valeur et sa saturation sont plus variables que celles du manche en fonction des conditions d'éclairage, changeantes au cours de l'intervention. De plus la forme de la pointe est variable en fonction de la rotation de l'instrument sur son propre axe et du degré d'inclinaison de celui-ci par rapport au plan image. Une mise à jour du modèle à chaque image, de façon similaire à [Loukas et al., 2013], permettrait probablement d'améliorer la performance de cette approche.

3.2.9 Améliorations apportées à la méthode

▷ **Déplacement du point d'insertion** L'approche proposée se base sur la position du point d'insertion dans le repère de la caméra. Il est évident qu'au cours d'une intervention chirurgicale, en raison de l'élasticité des tissus et des contraintes appliquées à la paroi par le chirurgien, le point d'insertion est en réalité mobile autour d'une position moyenne. Cette variabilité est prise en compte en ajoutant, pour chaque particule, à la position du point d'insertion un bruit gaussien dans l'espace. Ce bruit est un vecteur 3x1 pour lequel chaque composante suit une loi normale $\mathcal{N}(0, 5)$, l'écart-type des déplacements étant choisi à partir des travaux de Voros et al. [2006]. A chaque itération le point d'insertion est mis à jour avec le point d'insertion modifié de la particule de plus haut poids. J'applique un seuil empêchant le point d'insertion de dériver à plus de 15mm du point d'insertion initialement déterminé (limite choisie après discussion avec des chirurgiens), pour contrôler le cas où l'instrument ne change pas d'orientation pendant plusieurs itérations et où le point d'insertion est donc susceptible de se décaler fortement sur l'axe.

▷ **Prise en compte de plusieurs instruments** Cette méthode permet de suivre de façon simultanée plusieurs instruments sans les confondre, grâce à la contrainte du point d'insertion. Chaque instrument est représenté par un jeu de particules spécifiques, qui évoluent sur des géodes différentes (figure 3.35). Le temps de calcul n'est que modérément augmenté, toute la partie axée sur le traitement de l'image étant commune. Comme discuté dans la section 3.2.8, le suivi des deux instruments rencontre une limite lorsque les axes des instruments se croisent au voisinage des pointes, sans superposition des manches : la détection de la pointe n'est alors plus fiable.

3.3 Résultats bruts

La méthode présentée ci-dessus a d'abord été évaluée dans des conditions de banc d'entraînement, sur des fonds unis ou sur des "fausses" scènes chirurgicales (impression d'images extraites de procédures laparoscopiques). Le calibrage de la caméra

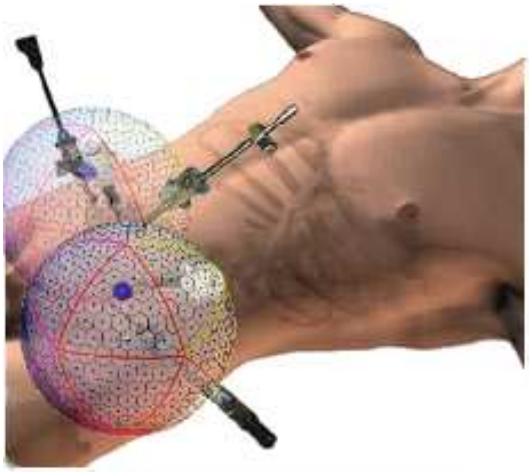


FIGURE 3.35. – Suivi simultané de deux instruments en utilisant une géode centrée sur chaque point d'insertion

est réalisé par la méthode de Zhang [2000] implémentée dans OpenCV (figure 3.36). Le point d'insertion est déterminé en utilisant la méthode décrite section 3.2.5, plusieurs déterminations successives avec des images différentes montrent une bonne reproductibilité de la mesure, aussi bien avec la détection automatique des droites qu'avec leur sélection manuelle (figure 3.37).

La méthode de suivi est implémentée en C++ (Visual C++ 2008 Express Edition, Microsoft) , et toutes les vidéos ont été traitées en temps différé sur un PC Intel Core i7, 3,4GHz avec 3.2GB RAM sous Windows XP Pro (Microsoft). Les données de position à chaque instant sont stockées au format .csv avant traitement et affichage par le logiciel Matlab (The MathWorks, Inc.).

Visuellement, le suivi est de bonne qualité, avec une détermination de la position de la pointe dans l'image qui correspond à la pointe effectivement observée (figure 3.38). En revanche, la reconstruction de la trajectoire en trois dimensions laisse apparaître un signal fortement bruité par rapport à une trajectoire de référence déterminée à l'aide d'un dispositif de marqueurs optiques (figure 3.39). Les erreurs de mesure avec cette trajectoire de référence sont présentés pour deux contextes d'utilisation dans le tableau 3.2. On note que les erreurs sont ponctuellement très importantes (jusqu'à plusieurs dizaines de centimètres), et qu'elles sont dépendantes du nombre de particules choisies pour décrire l'orientation de l'instrument. Ces erreurs sont trop importantes pour permettre d'analyser la performance de la méthode de suivi et sa dépendance aux paramètres choisis sans traitement préalable. Je présente dans le chapitre suivant les sources d'erreurs de la méthode, le traitement effectué sur la trajectoire afin de réduire le bruit, ainsi que les résultats obtenus dans différents contextes d'utilisation.

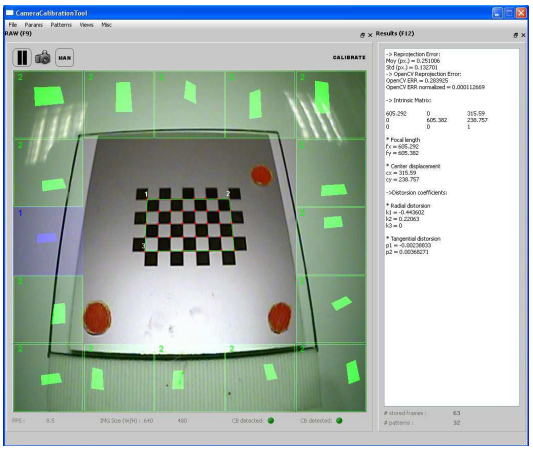


FIGURE 3.36. – Étape de calibrage de la caméra avec la procédure de Zhang [2000]

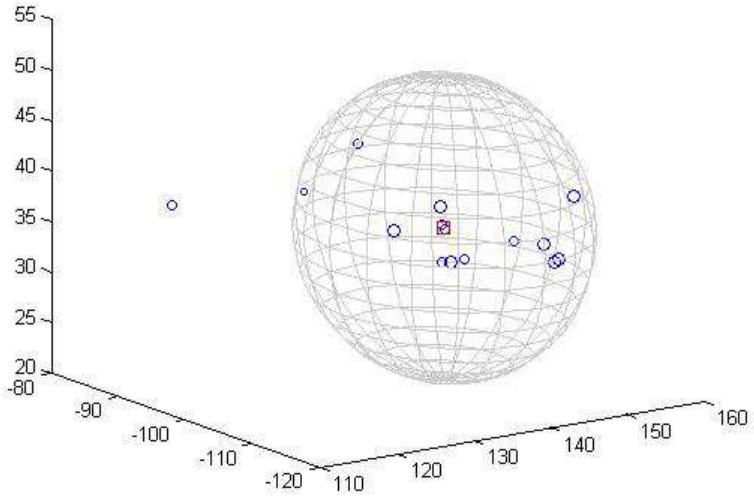


FIGURE 3.37. – Position des points d'insertion avec 15 déterminations itératives, avec une sélection manuelle des droites ou une sélection automatique. Le point rouge correspond à la position moyenne des points d'insertion, les points bleus à chaque point d'insertion. La taille de chaque point bleu correspond à l'inverse de l'erreur quadratique moyenne du modèle utilisé pour déterminer le point d'insertion associé. La sphère représentée a un rayon de 15mm, tous les points d'insertion sauf un sont donc dans un voisinage proche du point d'insertion moyen. Le point le plus éloigné des autres a une erreur quadratique moyenne importante, ce qui peut amener l'opérateur à l'exclure et à recommencer sa mesure.

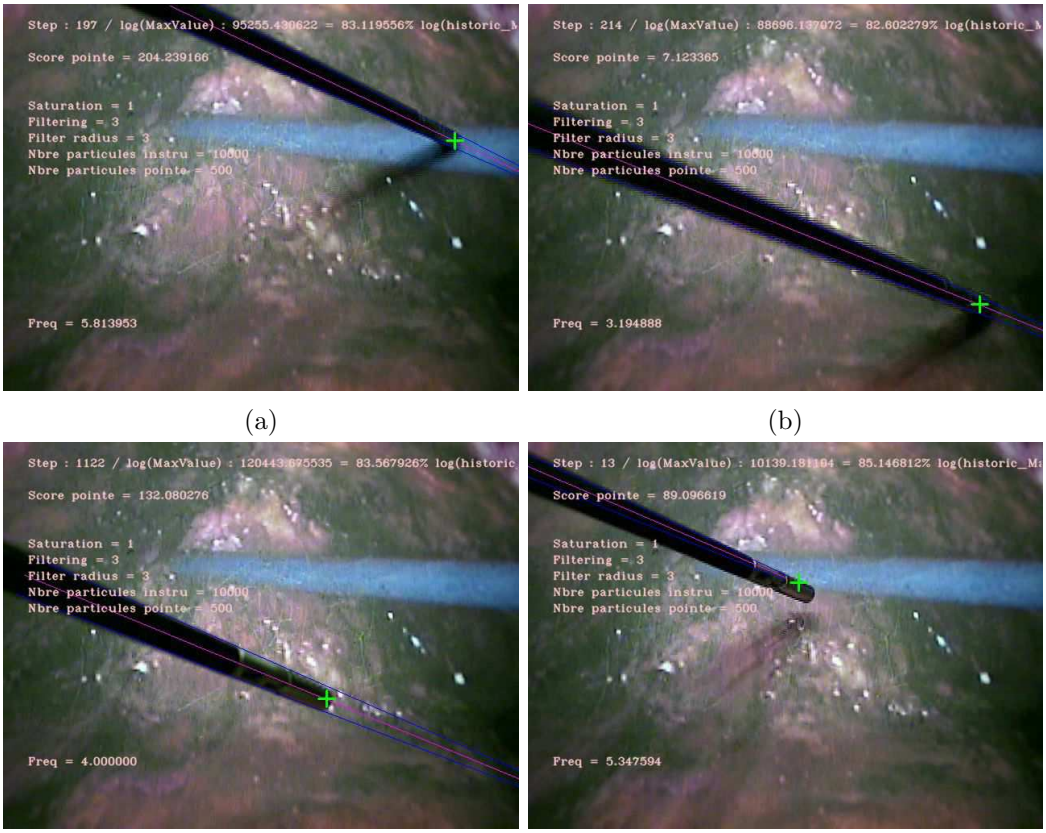


FIGURE 3.38. – Exemples d'image de suivi sur un fond de prostatectomie radicale laparoscopique. On note que la pointe marquée en vert est toujours dans le voisinage de la pointe observée dans l'image, même lorsque les bords calculés ne sont pas correctement ajustés sur les bords observés de l'instrument dans l'image (d). Le suivi est sensiblement de la même qualité pour l'ensemble de la séquence (1164 images consécutives).

Nombre de particules	Carré sur fond uni		Prostatectomie	
	Dist. moyenne	Dist. maximale	Dist. moyenne	Dist. maximale
10000	62.6 mm	230 mm	63.7 mm	320 mm
2000	61.4 mm	340 mm	75.7 mm	410 mm
1000	75.7 mm	370 mm	76.2 mm	430 mm

Tab. 3.2. – Précision de la méthode avant traitement des trajectoires, en suivant l'instrument sur un fond uni et sur un fond simulant une prostatectomie, en fonction du nombre de particules utilisées pour décrire son orientation. Précision mesurée par les distances moyenne et maximale en 3D entre les points de la trajectoire obtenue par la méthode de suivi développée et les points de la trajectoire obtenue par suivi de marqueurs optiques.

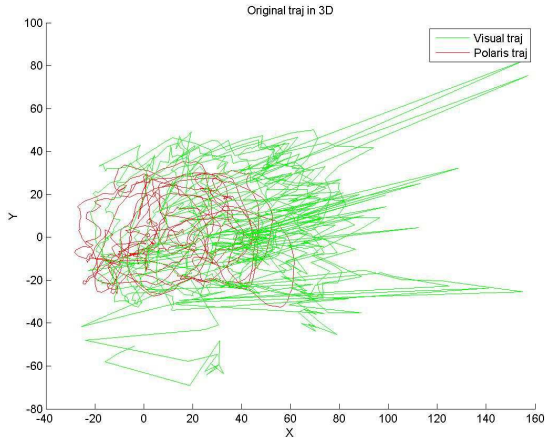


FIGURE 3.39. – Représentation de la trajectoire de la pointe dans l'espace, obtenue à partir des images de la figure 3.38, avant filtrage. On constate que la trajectoire est très bruitée avec de nombreuses mesures aberrantes.

Références

Eric Acosta et Bharti Temkin : Haptic laparoscopic skills trainer with practical user evaluation metrics. Studies in health technology and informatics, 111:8–11, 2005. PMID : 15718689. 74

Thomas E Ahlering, Douglas Skarecky, David Lee et Ralph V Clayman : Successful transfer of open surgical skills to a laparoscopic environment using a robotic interface : initial experience with laparoscopic radical prostatectomy. The Journal of urology, 170(5):1738–1741, novembre 2003. PMID : 14532766. 73

M. Allan, S. Ourselin, S. Thompson, D. Hawkes, J. Kelly et D. Stoyanov : Towards detection and localization of instruments in minimally invasive surgery. Biomedical Engineering, IEEE Transactions on, PP(99):1–1, 2012. 79, 101

Brian F Allen, Florian Kasper, Gabriele Nataneli, Erik Dutson et Petros Faloutsos : Visual tracking of laparoscopic instruments in standard training environments. Studies in health technology and informatics, 163:11–17, 2011. PMID : 21335750. 80

Brice Anselme et Sébastien Gadal : UVED - suivi de l'environnement par télédétection. http ://e-cours.univ-paris1.fr/modules/uved/envcal/html/compositions-colorees/representations-couleur/modeles-perceptuels/tsl.html, 2008. 90

Christophe Boschet : Laparoscopie Répartie. Thèse de doctorat, Université de Grenoble, 2010. 78

John Canny : A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-8(6):679–698, 1986. 93, 96, 98

Alicia M. Cano, Francisco Gayá, Pablo Lamata, Patricia Sánchez-González et Enrique J. Gómez : Laparoscopic tool tracking method for augmented reality surgical applications. In Fernando Bello et P. J. Eddie Edwards, éditeurs : Biomedical Simulation, numéro 5104 in Lecture Notes in Computer Science, pages 191–196. Springer Berlin Heidelberg, janvier 2008. 80

Alicia M. Cano, Pablo Lamata, Francisco Gayá et Enrique J. Gómez : New methods for Video-Based tracking of laparoscopic tools. In Matthias Harders et Gábor Székely, éditeurs : Biomedical Simulation, numéro 4072 in Lecture Notes in Computer Science, pages 142–149. Springer Berlin Heidelberg, janvier 2006. 80

Fabrice Chassat et Stephane Lavallée Ph.D : Experimental protocol of accuracy evaluation of 6-D localizers for computer-integrated surgery : Application to four optical localizers. In William M. Wells, Alan Colchester et Scott Delp, éditeurs : Medical Image Computing and Computer-Assisted Interventation — MICCAI'98, numéro 1496 in Lecture Notes in Computer Science, pages 277–284. Springer Berlin Heidelberg, janvier 1998. 75

M.K. Chmarra, N.H. Bakker, C.A. Grimbergen et J. Dankelman : TrEndo, a device for tracking minimally invasive surgical instruments in training setups. Sensors and Actuators A : Physical, 126 (2):328–334, 2006. 76

MK Chmarra, CA Grimbergen et J. Dankelman : Systems for tracking minimally invasive surgical instruments. Minimally Invasive Therapy & Allied Technologies, 16(6):328–340, 2007. 72

Joan Climent et Roberto A. Hexsel : Particle filtering in the hough space for instrument tracking. Computers in Biology and Medicine, 42(5):614–623, 2012. 80

Joan Climent et Pere Marés : Automatic instrument localization in laparoscopic surgery. Electronic Letters on Computer Vision and Image Analysis, 4(1):21–31, 2004. 79, 80, 123

Jean Pierre Cocquerez et Sylvie Philipp-Foliguet : Analyse d'images : filtrage et segmentation. Enseignement de la physique. MASSON, octobre 1995. 78

Stephane Cotin, Nicholas Stylopoulos, Mark Ottensmeyer, Paul Neumann, David Rattner, Steven Dawson, Takeyoshi Dohi et Ron Kikinis : Metrics for laparoscopic skills trainers : The weakest link ! Medical Image Computing and Computer-Assisted Intervention — MICCAI 2002, pages 35–43, janvier 2002. 76

S.M. Cristancho, A.J. Hodgson, O.N.M. Panton, A. Meneghetti, G. Warnock et K. Qayumi : Intraoperative monitoring of laparoscopic skill development based on quantitative measures. Surgical endoscopy, 23(10):2181–2190, 2009. 76, 77

S. DiMaio et C. Hasser : The da vinci research interface. In MICCAI Workshop on Systems and Architecture for Computer Assisted Interventions. Midas Journal, 2008. 74

C Doignon, P Graebling et M de Mathelin : Real-time segmentation of surgical instruments inside the abdominal cavity using a joint hue saturation color feature. Real-Time Imaging, 11(5-6):429–442, décembre 2005. 79, 88, 89, 92, 95

C. Doignon, F. Nageotte et M. De Mathelin : Detection of grey regions in color images : application to the segmentation of a surgical instrument in robotized laparoscopy. In 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2004. (IROS 2004). Proceedings, volume 4, pages 3394–3399 vol.4, septembre 2004. 89

C. Doignon, F. Nageotte et M. de Mathelin : The role of insertion points in the detection and positioning of instruments in laparoscopy for robotic tasks. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006, page 527–534, 2006. 81, 104

C. Doignon, F. Nageotte et M. de Mathelin : Segmentation and guidance of multiple rigid objects for intra-operative endoscopic vision. Dynamical Vision, page 314–327, 2007. 79

J. Duchateau, Sandrine Voros et P. Cinquin : Development and precision evaluation of an Image-Analysis based endoscopic tool tracking system. In Proceedings of Surgetica, pages 49–52, Chambéry, France, septembre 2007. 82, 114

Richard O. Duda et Peter E. Hart : Use of the hough transformation to detect lines and curves in pictures. Communications of the ACM, 15(1):11–15, janvier 1972. 79

Martin A. Fischler et Robert C. Bolles : Random sample consensus : a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM, 24(6):381–395, 1981. 106, 107

Yihong Gong, G. Proietti et C. Faloutsos : Image indexing and retrieval based on human perceptual color clustering. In 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1998. Proceedings, pages 578–583, 1998. 78

R.M. Haralick, K. Shanmugam et Its'Hak Dinstein : Textural features for image classification. IEEE Transactions on Systems, Man and Cybernetics, SMC-3(6):610–621, 1973. 78

R. Hartley et A. Zisserman : Multiple view geometry in computer vision. Cambridge Univ Pr, 2003. 83, 85

Ross Heikes et David A Randall : Numerical integration of the shallow-water equations on a twisted icosahedral grid. part i : Basic design and results of tests. Monthly Weather Review, 123(6):1862–1880, 1995. 109, 110

S Duke Herrell et Jr Smith, Joseph A : Robotic-assisted laparoscopic prostatectomy : what is the learning curve ? Urology, 66(5 Suppl):105–107, novembre 2005. PMID : 16194715. 73

Radu Horaud et Fadi Dornaika : Hand-Eye calibration. The International Journal of Robotics Research, 14(3):195–210, juin 1995. 82

Ming-Kuei Hu : Visual pattern recognition by moment invariants. IRE Transactions on Information Theory, 8(2):179–187, 1962. 78

M. Isard et A. Blake : Condensation—conditional density propagation for visual tracking. International journal of computer vision, 29(1):5–28, 1998a. 80, 114, 115, 119

M. Isard et A. Blake : A smoothing filter for condensation. Lecture Notes in Computer Science, 1406: 767–781, 1998b. 114

Alexandre Krupa, J. Gangloff, C. Doignon, M.F. de Mathelin, G. Morel, J. Leroy, L. Soler et J. Marescaux : Autonomous 3-D positioning of surgical instruments in robotized laparoscopic surgery using visual servoing. IEEE Transactions on Robotics and Automation, 19(5):842–853, octobre 2003. 82

Stéphane Lavallée, Philippe Cinquin et Jocelyne Troccaz : Computer integrated surgery and therapy : State of the art. STUDIES IN HEALTH TECHNOLOGY AND INFORMATICS, pages 239–309, 1997. 75

Jong-Sen Lee : Digital image smoothing and the sigma filter. Computer Vision, Graphics, and Image Processing, 24(2):255–269, 1983. 92

Constantinos Loukas, Vasileios Lahanas et Evangelos Georgiou : An integrated approach to endoscopic instrument tracking for augmented reality applications in surgical simulation training. The international journal of medical robotics + computer assisted surgery : MRCAS, janvier 2013. PMID : 23355307. 80, 81, 124

Alasdair I.M Macmillan et Alfred Cuschieri : Assessment of innate ability and skills for endoscopic manipulations by the advanced dundee endoscopic psychomotor tester : predictive and concurrent validity. The American Journal of Surgery, 177(3):274–277, mars 1999. 76

J. Matas, C. Galambos et J. Kittler : Robust detection of lines using the progressive probabilistic hough transform. Computer Vision and Image Understanding, 78(1):119–137, 2000. 105

Ignacio Oropesa, Magdalena K Chmarra, Patricia Sánchez-González, Pablo Lamata, Sharon P Rodrigues, Silvia Enciso, Francisco M Sánchez-Margallo, Frank-Willem Jansen, Jenny Dankelman et Enrique J Gómez : Relevance of Motion-Related assessment metrics in laparoscopic surgery. Surgical innovation, septembre 2012. PMID : 22983805. 80, 81, 101

Nobuyuki Otsu : A threshold selection method from Gray-Level histograms. IEEE Transactions on Systems, Man and Cybernetics, 9(1):62–66, 1979. 104, 119

Carol E. Reiley, Henry C. Lin, David D. Yuh et Gregory D. Hager : Review of methods for objective surgical skill evaluation. Surgical Endoscopy, 25(2):356–366, février 2011. 78

Austin Reiter, Peter K. Allen et Tao Zhao : Feature classification for tracking articulated surgical tools. In Nicholas Ayache, Hervé Delingette, Polina Golland et Kensaku Mori, éditeurs : Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012, numéro 7511 in Lecture Notes in Computer Science, pages 592–600. Springer Berlin Heidelberg, janvier 2012. 73

Todd D. Ringler, Ross P. Heikes et David A Randall : Spherical geodesic grids : A new approach to modeling the climate. http ://kiwi.atmos.colostate.edu/BUGS/geodesic/, 1999. 110, 111, 112

Jacob Rosen, Jeffrey D Brown, Lily Chang, Mika N Sinanan et Blake Hannaford : Generalized approach for modeling minimally invasive surgery as a stochastic process using a discrete markov model. IEEE Trans Biomed Eng, 53(3):399–413, mars 2006. 76, 77

Patricia Sánchez-González, Alicia M. Cano, Ignacio Oropesa, Francisco M. Sánchez-Margallo, Francisco Del Pozo, Pablo Lamata et Enrique J. Gómez : Laparoscopic video analysis for training and image-guided surgery. Minimally Invasive Therapy & Allied Technologies, 20(6):311–320, décembre 2011. 80

Alvy Ray Smith : Color gamut transform pairs. volume 12, pages 12–19. ACM, 1978. 89

C. Sokollik, J. Gross et G. Buess : New model for skills assessment and training progress in minimally invasive surgery. Surgical Endoscopy And Other Interventional Techniques, 18(3):495–500, mars 2004. 76, 77

Stefanie Speidel, Michael Delles, Carsten Gutt et Rüdiger Dillmann : Tracking of instruments in minimally invasive surgery for surgical skill analysis. In Guang-Zhong Yang, TianZi Jiang, Dinggang Shen, Lixu Gu et Jie Yang, éditeurs : Medical Imaging and Augmented Reality, numéro 4091 in Lecture Notes in Computer Science, pages 148–155. Springer Berlin Heidelberg, janvier 2006. 79, 80

Stefanie Speidel, Gunther Sudra, Julien Senemaud, Maximilian Drentschew, Beat Peter Müller-Stich, Carsten Gutt et Rüdiger Dillmann : Recognition of risk situations based on endoscopic instrument tracking and knowledge based situation modeling. pages 69180X–69180X–8, mars 2008. 79

F. Tatar, J.R. Mollinger, R. C. Den Dulk, W. A. Van Duyl, J. F L Goosen et A. Bossche : Ultrasonic sensor system for measuring position and orientation of laproscopic instruments in minimal invasive surgery. In Microtechnologies in Medicine amp ; Biology 2nd Annual International IEEE-EMB Special Topic Conference on, pages 301–304, 2002. 76

Carlo Tomasi et Roberto Manduchi : Bilateral filtering for gray and color images. pages 839–846. IEEE, 1998. 92

K. R. Van Sickle, D. A. McClusky Iii, A. G. Gallagher et C. D. Smith : Construct validation of the ProMIS simulator using a novel laparoscopic suturing task. Surgical Endoscopy And Other Interventional Techniques, 19(9):1227–1231, septembre 2005. 78

S. Voros, J.A. Long et P. Cinquin : Automatic detection of instruments in laparoscopic images : A first step towards high-level command of robotic endoscopic holders. International Journal of Robotics Research, 26(11-12):1173–1190, 2007. 79, 80

Sandrine Voros, Jean-Alexandre Long et Philippe Cinquin : Automatic localization of laparoscopic instruments for the visual servoing of an endoscopic camera holder. In Rasmus Larsen, Mads Nielsen et Jon Sporring, éditeurs : Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006, numéro 4190 in Lecture Notes in Computer Science, pages 535–542. Springer Berlin Heidelberg, janvier 2006. 104, 105, 124

H.W. Wehn et P.R. Belanger : Ultrasound-based robot position estimation. IEEE Transactions on Robotics and Automation, 13(5):682–692, 1997. 76

Rémi Wolf, Josselin Duchateau, Philippe Cinquin et Sandrine Voros : 3D tracking of laparoscopic instruments using statistical and geometric modeling. In Gabor Fichtinger, Anne Martel et Terry Peters, éditeurs : Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011, volume 6891 de Lecture Notes in Computer Science, page 203–210. Springer, 2011. 80

G Wyszecki et WS Stiles : Color science : concepts and methods, quantitative data and formulae. Wiley classics library edition, 2000. 88

Yitzhak Yitzhaky et Eli Peli : A method for objective edge detection evaluation and detector parameter selection. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 25(8):1027–1033, 2003. 97, 99

ZY Zhang : A flexible new technique for camera calibration. Ieee Transactions On Pattern Analysis and Machine Intelligence, 22(11):1330–1334, novembre 2000. 12, 85, 125, 126

CHAPITRE 4

FILTRAGE DES TRAJECTOIRES ET ÉVALUATION DE LA PRÉCISION DE LA MÉTHODE DÉVELOPPÉE

Sommaire

4.1 Filtrage des trajectoires 134
4.1.1 Nature du bruit enregistré 134
4.1.2 Filtrage des mesures aberrantes 135
4.1.3 Lissage de la trajectoire obtenue 137
4.2 Précision de la mesure et influence des paramètres 139
4.2.1 Dispositif d'enregistrement des trajectoires d'évaluation . . . 139
4.2.2 Définition des trajectoires réelles 140
4.2.3 Mise en évidence d'une erreur systématique dans l'estimation de la position de l'instrument dans l'espace 141
4.2.4 Erreur dans l'estimation de la position de l'instrument dans le plan image 146
4.2.5 Analyse de l'influence du choix des paramètres 146
4.2.6 Évaluation sur spécimen anatomique 153
4.3 Discussion et conclusion pour le suivi d'instruments 154
4.3.1 Limites actuelles de la méthode proposée 154
4.3.2 Apports principaux de la méthode proposée 156
4.3.3 Optimisations envisagées 157
4.3.4 Conclusion 159

Comme je le montre à la fin du chapitre précédent, la trajectoire des instruments dans l'espace reste bruitée malgré l'optimisation des différentes étapes de leur détection, y compris dans des conditions favorables (banc d'entraînement, c'est-à-dire avec un arrière plan peu variable). Je présente dans ce chapitre la méthode développée pour diminuer le bruit du signal obtenu après enregistrement de l'ensemble

FIGURE 4.1. – Erreur importante dans la détermination de l'orientation de l'instrument, liée à un échec de la détection d'un des bords de l'instrument (image sélectionnée parmi les premières images de la séquence, avant que l'algorithme de CONDENSATION ne converge).

de la trajectoire, puis évalue la précision du suivi d'instrument en fonction des paramètres choisis.

4.1 Filtrage des trajectoires

4.1.1 Nature du bruit enregistré

La trajectoire obtenue avec la méthode présentée au chapitre 3 présente deux types de perturbations du signal, distincts l'un de l'autre.

mesures aberrantes : les bords supposés de l'instrument ne sont pas dans le voisinage des bords réels, ou la pointe détectée n'est pas dans le voisinage de la pointe réelle. Cette situation survient majoritairement lorsqu'aucun instrument n'est visible dans l'image, ou lorsque la détection de la pointe échoue lors du croisement de l'axe de deux instruments dans le voisinage de celle-ci. Elle peut également survenir de façon plus rare en cas de mouvement très brutal de l'instrument (par exemple lorsqu'une tension est appliquée sur un tissu et que la préhension de ce tissu est libérée involontairement). Ces situations sont caractérisées d'une part par un poids faible pour la particule représentant l'orientation de l'instrument la plus probable (si c'est la détection des bords qui échoue), d'autre part par une position de la pointe très distante des points voisins temporellement.

faibles perturbations autour de l'orientation et de la profondeur : l'incertitude principale pour la précision de la mesure se situe dans le plan perpendiculaire au plan image. Une faible erreur dans l'estimation des bords peut aboutir à une

erreur de plusieurs millimètres pour la position de la pointe dans ce plan image, indépendamment de la précision de la détection de la pointe elle-même. La figure 4.2 illustre cette sensibilité de la mesure aux erreurs de détection des bords. Pour chaque angle que fait l'instrument par rapport au plan image dans la situation schématisée dans la ligne supérieure de la figure, c'est à dire avec un instrument situé à une distance de la caméra similaire à celle rencontrée sur le banc d'essai, j'évalue l'erreur dans l'estimation de la position de la pointe en cas de mauvaise estimation de la position des bords. On note qu'une erreur de deux pixels entraîne un décalage de la pointe pouvant atteindre 20mm dans la direction z (normale au plan image). Si la taille de l'instrument augmente dans l'image (si la caméra se rapproche de la zone de travail, situation schématisée dans la ligne inférieure de la figure), ce qui est le cas dans les situations de chirurgie réelle, cette erreur diminue, pour atteindre 2.5mm au maximum si la section d'entrée de l'instrument occupe toute la hauteur de l'image. Cet effet est également illustré figure 4.3, où pour deux images successives l'orientation estimée de l'instrument change (effet visible en observant la position du point de fuite des deux droites formant les bords de l'instrument).

4.1.2 Filtrage des mesures aberrantes

En cas d'échec de la localisation de la position de l'instrument, il peut exister plusieurs données successives fausses. L'utilisation simple d'un filtre médian, tel celui proposé par Oropesa et al. [2012] n'est donc pas fiable, car la taille du filtre devra varier : un filtre prenant en compte un faible nombre de données ne filtrera pas une mauvaise détection se prolongeant trop longtemps, un filtre de grande taille risque de faire disparaître des informations pertinentes pour la suite de ce travail.

Je m'appuie pour l'exclusion des valeurs aberrantes sur la méthode proposée par He et Tian [1998] qui recherchent les valeurs aberrantes pour la position de marqueurs de Polaris. En résumé, la méthode consiste à rechercher les blocs de données de position encadrées par des valeurs aberrantes et opposées de la vitesse (fortement supérieures à l'écart-type de la vitesse). Pour l'appliquer au contexte de cette thèse, j'apporte les modifications suivantes :

exclusion des valeurs associées à des scores de détection faibles : soient $S_o(t)$ et $S_p(t)$ les scores associés respectivement à la détermination de l'orientation de l'instrument et à la détermination de la profondeur d'insertion, à l'instant t. Si $\log S_o(t) < 0.5 \log\left(\max_t S_o(t)\right)$ ou $\log S_p(t) < 0.5 \log\left(\max_t S_p(t)\right)$ la position $X(t)$ est marquée comme aberrante.

interpolation des données manquantes et segmentation de la trajectoire : si plus de ω positions successives sont considérées comme aberrantes, je divise la trajectoire en deux sous-trajectoires, considérant que l'instrument est sorti du champ de vision de la caméra durant cet intervalle. Je ne retiens que les sous-trajectoires contenant plus de ω positions. Si le segment de données aberrantes

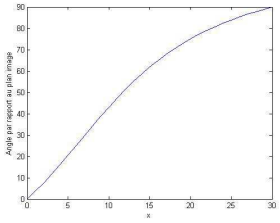
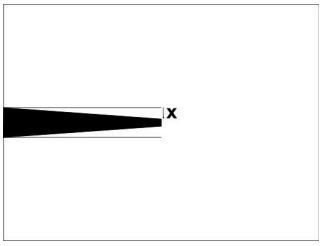
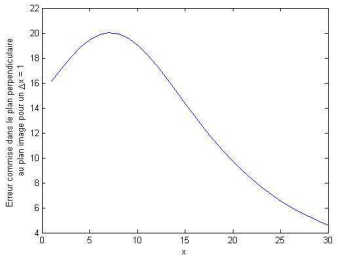
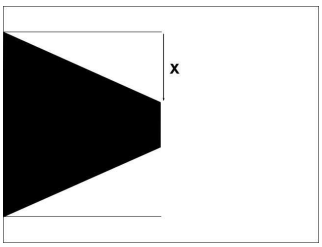
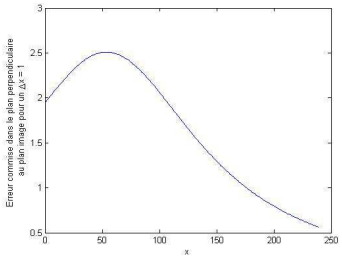
Angle de l'instrument par rapport au plan image, en fonction de x	Instrument dont on fait varier l'orientation dans un plan perpendiculaire au plan image	Distance dans l'espace entre deux positions de la pointe pour $\Delta x = 1$, avec une profondeur d'insertion de 250mm
		
		

FIGURE 4.2. – Erreur de mesure dans l'espace en fonction de l'angle de l'instrument par rapport au plan image, en cas d'erreur sur la détection des bords de 2 pixels. Les figures de la première ligne correspondent à une configuration de banc d'entraînement (instrument distant de la camére), la seconde ligne à une configuration plus proche d'une chirurgie réelle (instrument proche de la caméra).

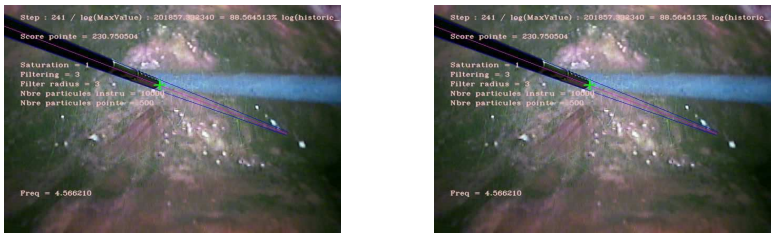


FIGURE 4.3. – Variation fine de l'orientation de l'instrument entre deux images successives.

est plus court que ω positions successives, j'interpole les données manquantes de façon linéaire (je considère que le mouvement est linéaire sur des intervalles de temps inférieurs à la seconde, avec $\omega = 25$). Une interpolation polynomiale cubique ou par splines n'améliore pas la qualité du suivi (expériences préliminaires).

recherche des variations brutales de la vitesses : j'étudie la distribution des vitesses $V^i(t)$ de la pointe de l'instrument selon les directions x, y et z du repère cartésien lié à la caméra, avec $i \in \{x, y, z\}$, pour l'ensemble de la trajectoire. Soient μ_V^i et σ_V^i la moyenne et l'écart-type de la distribution de la vitesse selon chaque direction. Soit k la tolérance admise pour considérer qu'une valeur de la vitesse n'est pas aberrante. En effet, comme le notent He et Tian [1998], les mesures aberrantes sont distantes spatialement des mesures les encadrant, et sont donc caractérisées par des pics de vitesse : $V^i(t) \notin [\mu_V^i - k\sigma_V^i, \mu_V^i + k\sigma_V^i]$ si $X(t-1)$ est une mesure pertinente et $X(t)$ est une valeur aberrante, ou si $X(t-1)$ est une mesure aberrante et $X(t)$ est une valeur pertinente. Je recherche donc sur l'ensemble de la trajectoire toutes les positions pour lesquelles la vitesse n'est pas comprise dans l'intervalle $[\mu_V^i - k\sigma_V^i, \mu_V^i + k\sigma_V^i]$.

recherche des couples de vitesses encadrant des mesures aberrantes : un ensemble de mesures de position aberrantes sera donc encadré par deux valeurs de la vitesse n'appartenant pas à l'intervalle de tolérance, et de directions opposées spatialement. Je considère que toutes les positions comprises entre t et $t + \Delta t$ telles que $V^i(t) \notin [\mu_V^i - k\sigma_V^i, \mu_V^i + k\sigma_V^i]$, $V^i(t + \Delta t) \notin [\mu_V^i - k\sigma_V^i, \mu_V^i + k\sigma_V^i]$ et $V^i(t) * V^i(t + \Delta t) < 0$ sont des mesures aberrantes, avec l'intervalle Δt le plus faible possible, et limité à ω. J'interpole les données manquantes selon les mêmes critères qu'à l'étape précédente. Les expériences préliminaires permettent de fixer $k = 1.5$.

La figure 4.4 illustre le résultat de cette étape du filtrage.

4.1.3 Lissage de la trajectoire obtenue

Afin de limiter l'influence des bruits de mesure (tels que décrits plus haut), je filtre la trajectoire obtenue à l'étape précédente avec des splines de lissage cubiques [1]. La documentation associée au programme utilisé pour calculer ces splines suggère de choisir le paramètre de lissage dans un voisinage de $p = \frac{1}{1+\frac{h^3}{6.10^n}}$ où h est la distance moyenne séparant deux points consécutifs de la trajectoire, et n varie de 1 (très faible lissage) à -1 (fort lissage). Ici l'écart moyen entre deux points est comprise entre 5 et 10mm, le paramètre de lissage est choisi constant, $p = \frac{1}{1+\frac{10^4}{6}}$ après exploration de différentes valeurs de h et de n. Cette valeur de p permet de lisser suffisamment la trajectoire sans faire disparaître pour autant les mouvements parasites d'intensité

1. Toolbox csaps de Matlab

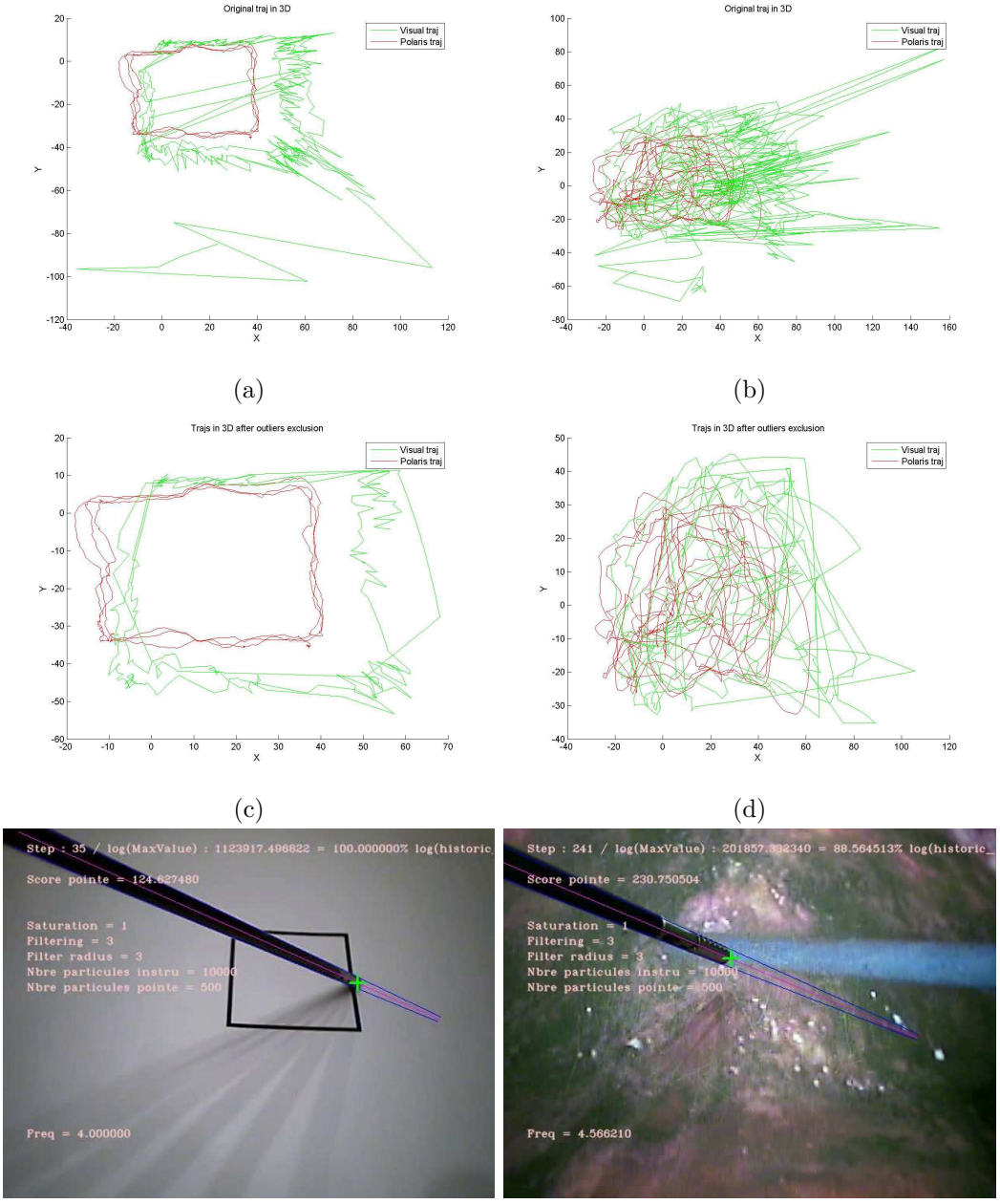


FIGURE 4.4. – Exemples de trajectoires avant ((a) et (b)) et après ((c) et (d)) exclusion des mesures aberrantes. La première trajectoire représente un carré parcouru à plusieurs reprises (e), la seconde consiste en des mouvements typiques de chirurgie (boucles, aller-retour latéraux) devant une image de fond de prostatectomie radicale (f).

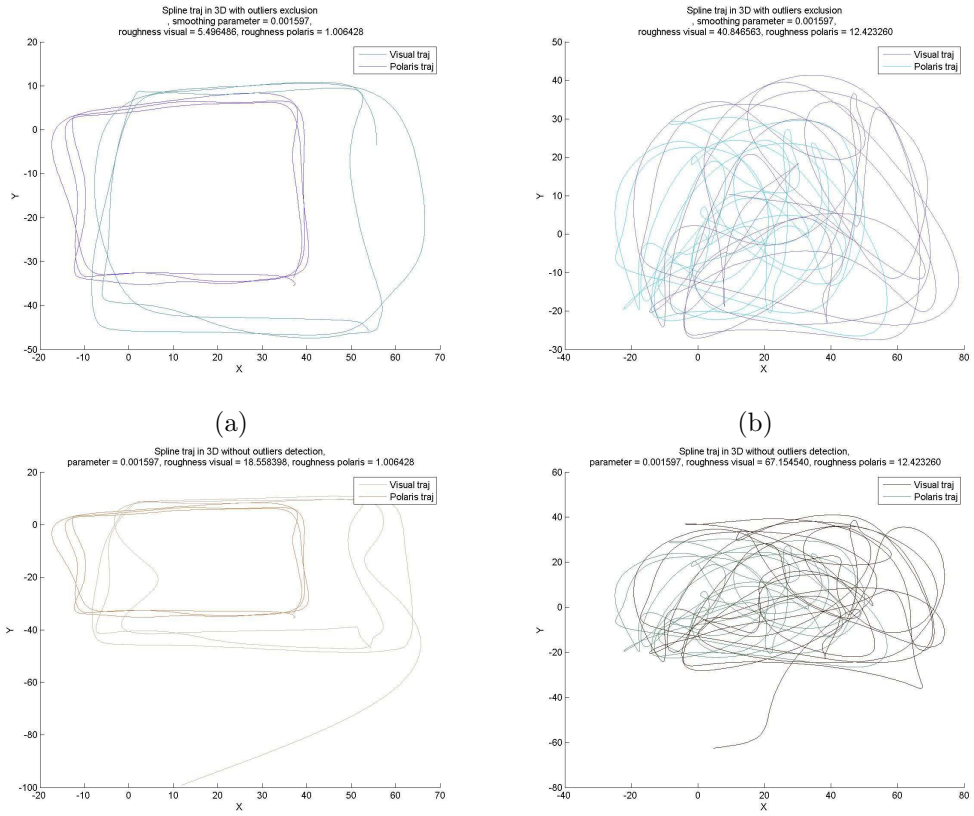


FIGURE 4.5. – Exemples de trajectoires après lissage par une spline avec un paramètre de lissage $p = \frac{1}{1+\frac{10^4}{16}}$. Résultats avec l'étape intermédiaire d'exclusion des mesures aberrantes ((a) et (b)) et sans cette étape ((c) et (d)).

suffisante pour être liés à l'opérateur. La figure 4.5 illustre l'effet de cette étape du filtrage sur les mêmes trajectoires qu'à l'étape précédente.

On observe sur cette figure l'intérêt de l'étape d'exclusion des mesures aberrantes, qui permet aux splines d'ignorer ces points sans modifier le paramètre de lissage.

4.2 Précision de la mesure et influence des paramètres

4.2.1 Dispositif d'enregistrement des trajectoires d'évaluation

Les trajectoires utilisées pour évaluer la précision de la mesure et l'influence des paramètres sur cette précision ont été acquises sur un banc d'entraînement (décrit en détail section 5.1), qui est équipé d'une caméra CCD intégrée offrant une résolution de $640 * 480$ pixels ayant une sortie RCA. Le signal analogique est transmis, via

un boîtier d'acquisition (DFG/USB2pro, The Imaging Source) qui le convertit en signal numérique, vers un port USB de l'ordinateur.

4.2.2 Définition des trajectoires réelles

Afin d'évaluer la précision de la détermination de la trajectoire en trois dimensions avec la méthode développée dans cette thèse, j'ai enregistré, à l'aide d'un système de suivi par marqueurs optiques actifs (accuTrack 250, Atracsys, Suisse), pour chaque image acquise, la position de la pointe de l'instrument ainsi que l'orientation de son axe (figure 4.6). Un calibrage préalable me permet de connaître la matrice de calibrage extrinsèque de la caméra dans le repère lié au dispositif de suivi par marqueurs optiques, et donc d'appliquer les translations et rotations nécessaires pour recaler les trajectoires obtenues dans deux repères différents dans un unique repère. Le calibrage entre les deux repères s'effectue par la localisation de trois points de géométrie connue, identifiables par la caméra, dans le repère lié au dispositif de suivi.
Il existe des sources d'erreur potentielles dans la détermination de cette trajectoire de référence. Ces erreurs peuvent être liées :

- à la précision intrinsèque du dispositif de suivi des marqueurs optiques (l'erreur quadratique moyenne annoncée par le fabricant est de 0.2mm, mais peut être supérieure en fonction des conditions d'utilisation),
- aux imprécisions du recalage entre le repère dans lequel les marqueurs optiques sont détectés et la caméra (erreur dans le calibrage extrinsèque du système),
- à la stabilité de la relation entre la position des marqueurs optiques, fixés à l'extrémité utilisateur de l'instrument, et la position de la pointe de celui-ci, le manche étant légèrement flexible.

L'acquisition des images envoyées par la caméra et des positions des marqueurs optiques est synchronisée, une position étant enregistrée pour chaque image reçue depuis la caméra. Les images sont enregistrées individuellement au format bitmap avant d'être fusionnées pour former une vidéo (compression sans perte de type HuffYUV effectuée par VirtualDub [2]).

La trajectoire obtenue avec le suivi par marqueurs optiques est filtrée uniquement par une spline, avec le même paramètre de lissage que celui utilisé plus haut.

J'étudie, pour chaque image, plusieurs variables afin de comparer les trajectoires obtenues par les deux méthodes :

- l'erreur angulaire en trois dimensions entre les axes
- l'erreur en distance en trois dimensions entre les pointes
- l'erreur angulaire dans le plan image entre les projections des axes
- l'erreur en distance dans le plan image entre les pointes
- la vitesse de calcul.

2. http ://www.virtualdub.org/

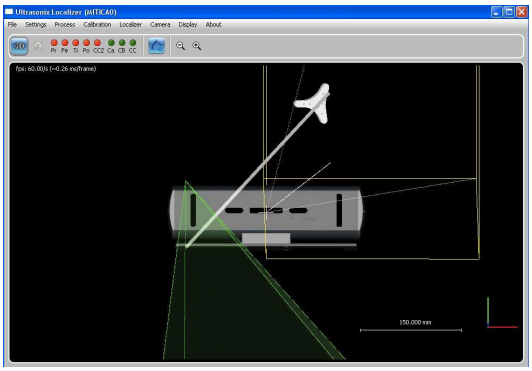


FIGURE 4.6. – Suivi de la position de l'instrument à l'aide du dispositif accuTrack 250.

4.2.3 Mise en évidence d'une erreur systématique dans l'estimation de la position de l'instrument dans l'espace

Comme on peut le noter sur les figures précédentes, il existe une erreur systématique sur la détermination de la position de la pointe de l'instrument dans l'espace, même en se plaçant dans les conditions optimales suivantes : 10000 particules pour représenter l'orientation de l'instrument, 500 sous-segments explorés le long de l'axe pour localiser la pointe, résolution de la géode de 9 (soit plus de 2.10^6 cellules). Ces paramètres permettent cependant d'obtenir un suivi visuellement satisfaisant en deux dimensions. En décomposant l'erreur sur les deux étapes principales de la méthode de détection de la position de l'instrument (détermination de l'orientation de l'instrument et détermination de la profondeur d'insertion), on observe :

- que l'erreur angulaire moyenne dans l'estimation de l'orientation est de 5 degrés en moyenne (figure 4.7), ce qui entraîne, pour une profondeur d'insertion de 250mm, une erreur systématique d'environ 21mm dans l'espace.
- que l'erreur observée dans l'espace est supérieure à cette erreur attendue, de l'ordre de 60mm (figure 4.8, avec les mêmes trajectoires que pour la figure 4.7), mettant en évidence une erreur systématique dans l'estimation de la profondeur.

On note également que les images pour lesquelles l'erreur est maximale ne sont pas nécessairement celles pour lesquelles les scores de détection étaient les plus bas, sans être inférieurs au seuil défini plus haut (figure 4.8). En observant les images associées à ces fausses détections supposées, on remarque que la détection des bords et de la pointe est visuellement de bonne qualité (figure 4.9).

De plus, la trajectoire déterminée avec mon approche est systématiquement décalée spatialement de la trajectoire déterminée avec le suivi par marqueurs optiques, avec un aspect d'homothétie entre les deux trajectoires (figures 4.7a-4.7d et 4.10). Il est probable que le calibrage, à la fois extrinsèque et intrinsèque, comporte des

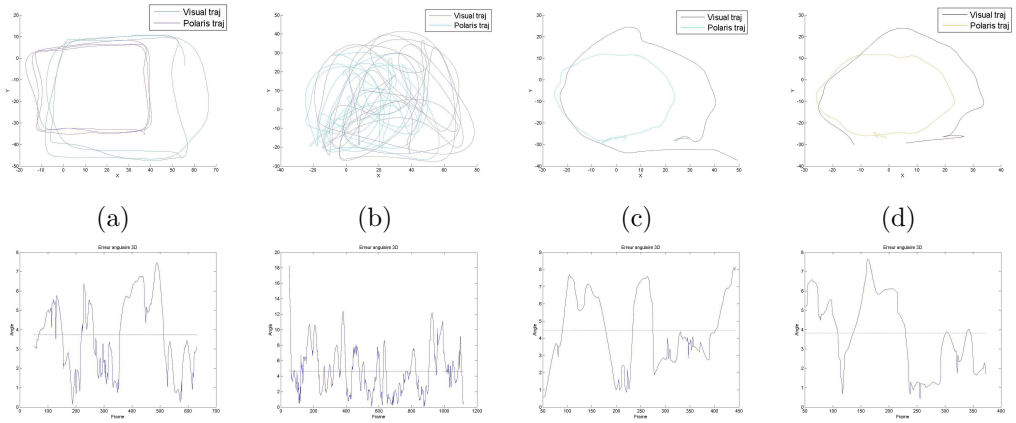


FIGURE 4.7. – Trajectoires après filtrage des mesures aberrantes et lissage par une spline (ligne du haut) et erreur angulaire systématique dans l'estimation de l'orientation de l'instrument associée à ces trajectoire (ligne du bas). Les lignes horizontales correspondent aux erreurs moyennes.

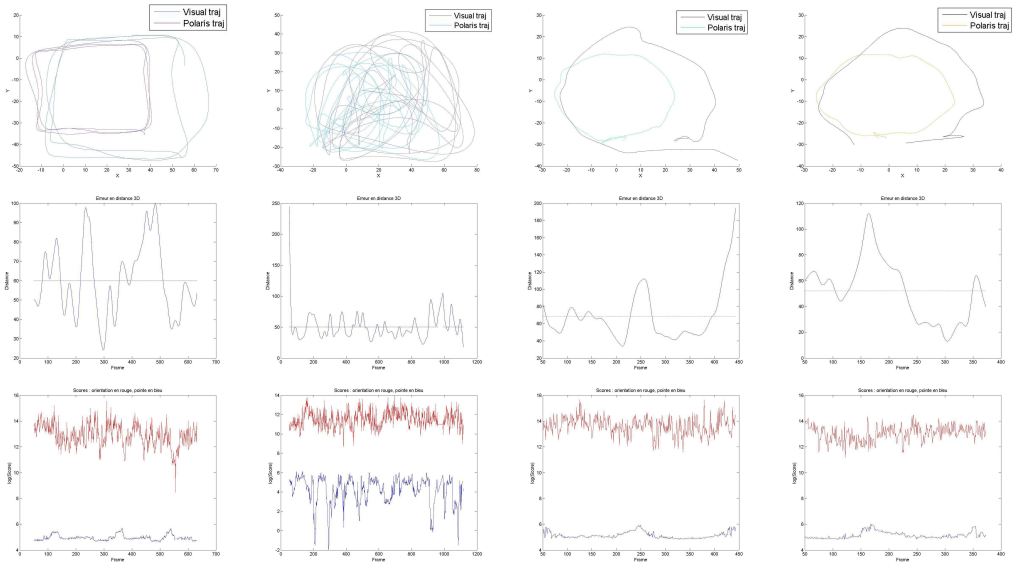


FIGURE 4.8. – Trajectoires (ligne du haut, identiques à celles de la figure 4.7), erreurs systématiques dans l'estimation de la position dans l'espace de l'instrument (ligne du milieu), et scores associés aux détections de l'orientation et de la pointe (ligne du bas).

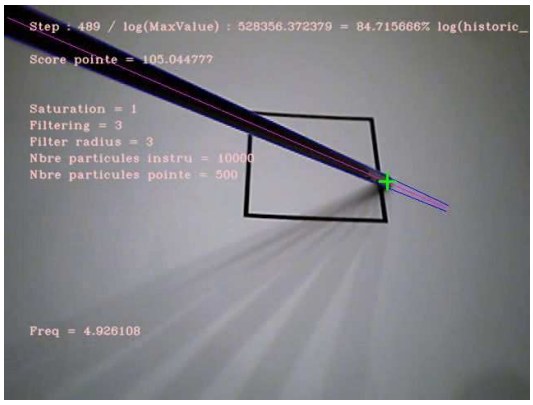


FIGURE 4.9. – Image correspondant à une détection de bonne qualité visuellement mais associée à une distance à la position de référence (marqueurs optiques) importante (100mm).

erreurs minimes (erreur de re-projection de 0,3 pixel en moyenne pour le calibrage intrinsèque) qui s'additionnent et aboutissent à une erreur dans le recalage des deux trajectoires dans un repère commun (figure 4.11). Le modèle théorique choisi pour représenter la caméra (modèle du sténopé) n'est par ailleurs probablement pas strictement exact, et ne prend pas en compte l'ensemble des déformations subies par les rayons lumineux avant d'atteindre le capteur. La matrice de calibrage intrinsèque n'est par conséquent pas forcément valable sur l'ensemble du champ de vision de la caméra.

On observe également sur la figure 4.10 que l'erreur est principalement composée d'une erreur dans le plan Z par rapport au plan X. L'erreur dans le plan Y est similaire à celle observée dans le plan X (données non représentées pour ce dernier plan). Ce constat est en accord avec les incertitudes attendues concernant l'inclinaison de l'instrument par rapport au plan image (figure 4.2).

Afin de limiter l'influence de cette erreur de recalage sur l'analyse de l'influence des paramètres, j'applique à la trajectoire déterminée par ma méthode une transformation rigide (translation et rotation) afin de minimiser l'écart quadratique moyen avec la trajectoire de référence [Fitzpatrick et al., 2000]. La figure 4.12 présente les trajectoires obtenues après cette étape de recalage rigide, ainsi que les erreurs en distance en en angle, en trois dimensions ; la figure 4.13 présente les trajectoires en fonction du temps selon les orientations X et Z, on note que l'erreur résiduelle reste principalement contenue dans l'estimation de la position en Z.

On constate que l'erreur moyenne est fortement diminuée après ce recalage (résultat attendu car le recalage consiste à minimiser l'erreur quadratique moyenne entre les deux trajectoires), les erreurs angulaires ne sont pas modifiées.

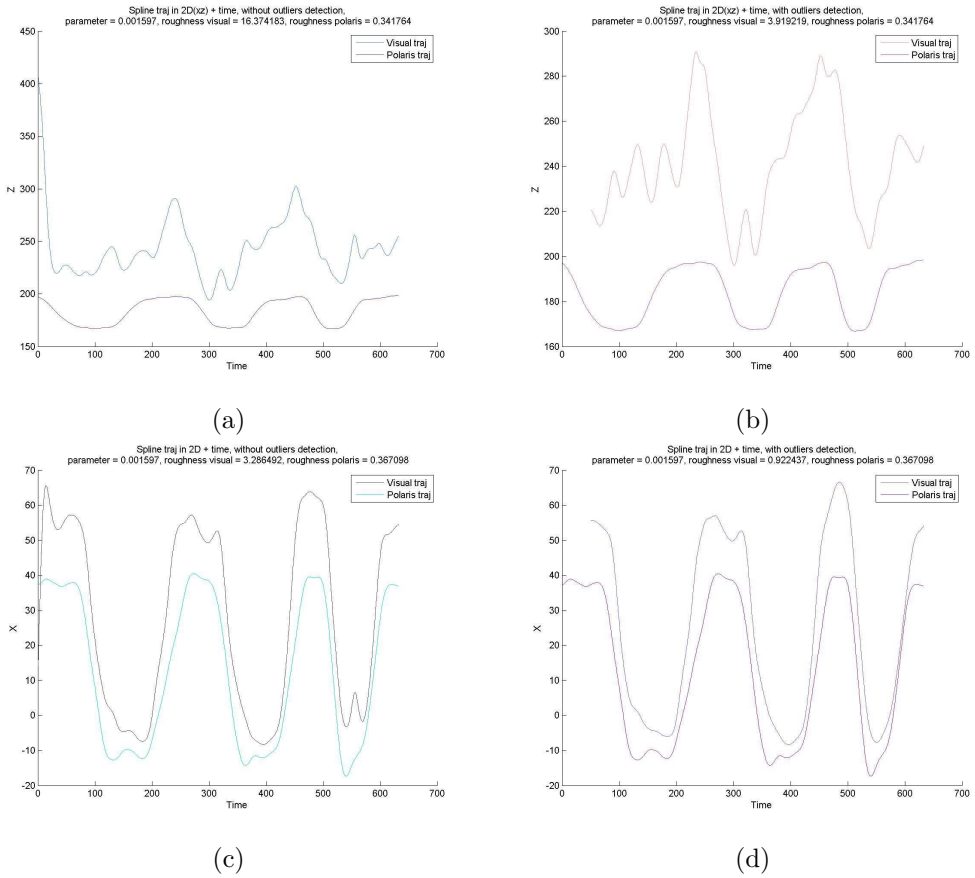
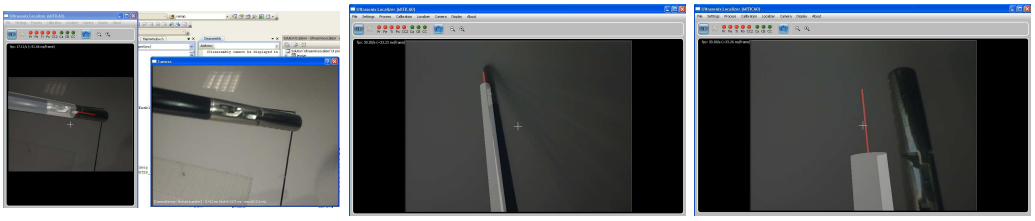
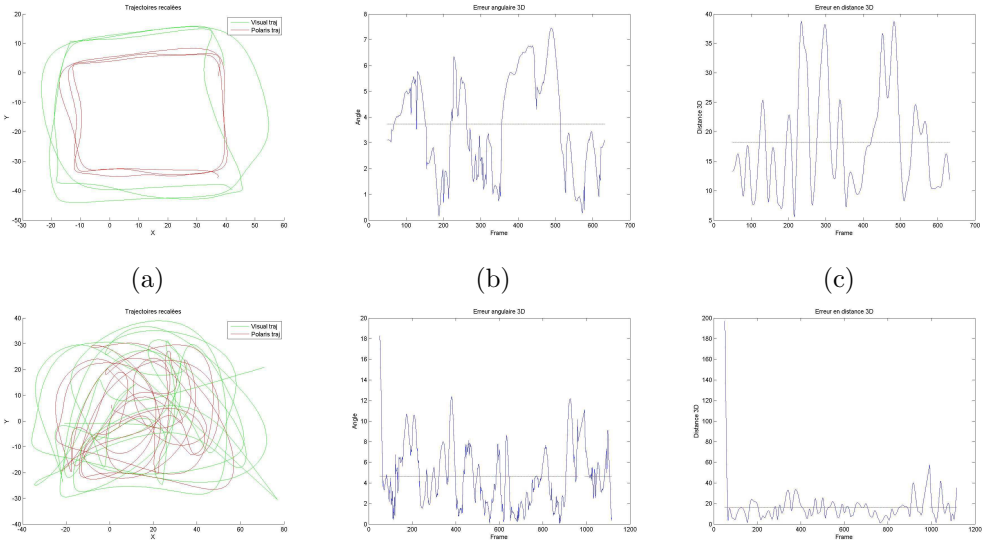


FIGURE 4.10. – Projection de la position de l'instrument sur les axes X et Z, mettant en évidence un aspect d'homothétie se superposant aux oscillations correspondant aux erreurs uniquement liées à la méthode proposée. Trajectoires obtenues sans exclusion des mesures aberrantes à gauche, avec exclusion de ces mesures à droite.



(a) (b) (c)

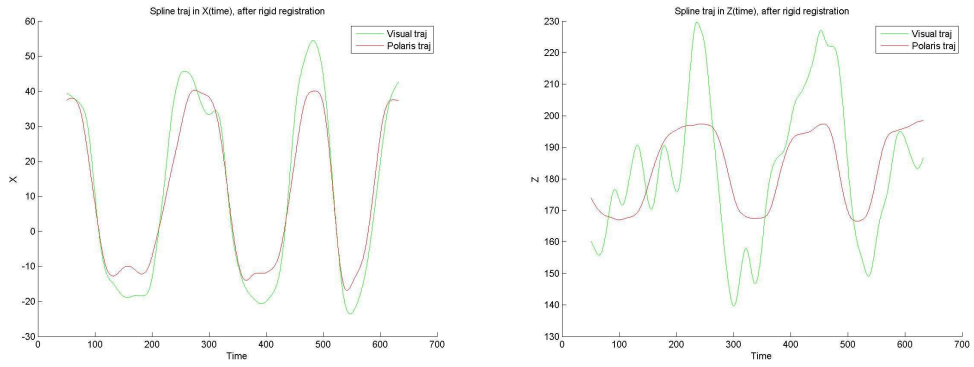
FIGURE 4.11. – Erreurs de re-projection liées aux incertitudes des calibrages intrinsèque et extrinsèque. On note que la re-projection du cylindre associé à l'instrument, dont la position est déterminée par le suivi par marqueurs optiques, n'est pas toujours superposable à l'image de l'instrument vu par la caméra (erreurs de re-projection croissantes de la gauche vers la droite).



(a) (b) (c)

(d) (e) (f)

FIGURE 4.12. – Trajectoires obtenues après recalage rigide ((a) et (d)), avec les erreurs, angulaires ((b) et (e)) et en distance ((c) et (f)), associées.



(a) Projection sur l'axe X (b) Projection sur l'axe Z

FIGURE 4.13. – Projection de la position de l'instrument sur les axes X et Z, après recalage rigide.

4.2.4 Erreur dans l'estimation de la position de l'instrument dans le plan image

La reprojection de l'axe directeur et de la pointe du cylindre modélisant l'instrument dans le dispositif de suivi par marqueurs optiques, sur le plan image de la caméra, permet d'estimer pour chaque image une erreur angulaire et une erreur de distance (en pixels) en deux dimensions. Cette erreur correspond à l'évaluation visuelle de la qualité du suivi. La figure 4.14 présente quelques exemples de ces erreurs pour les trajectoires étudiées plus haut, après recalage rigide dans l'espace. On observe sur les trajectoires projetées sur le plan image le même effet d'homothétie que discuté ci-dessus, qui n'est pas visible lors de l'observation du suivi. Il existe donc une erreur systématique dans le recalage entre le repère de la caméra et le repère du dispositif de suivi des marqueurs optiques, que je n'ai pas pu réduire malgré le soin apporté à l'étape de calibrage.

Malgré l'existence de cette erreur j'analyse dans la section suivante l'influence des paramètres sur la qualité du suivi.

4.2.5 Analyse de l'influence du choix des paramètres

Les différents paramètres étudiés pour évaluer la qualité du suivi sont le nombre de particules décrivant l'orientation de l'instrument, le nombre de segments explorés pour rechercher la pointe de l'instrument le long de son axe et la résolution de la grille géodésique.
J'étudierai également l'influence du type de fond et du choix du filtre.

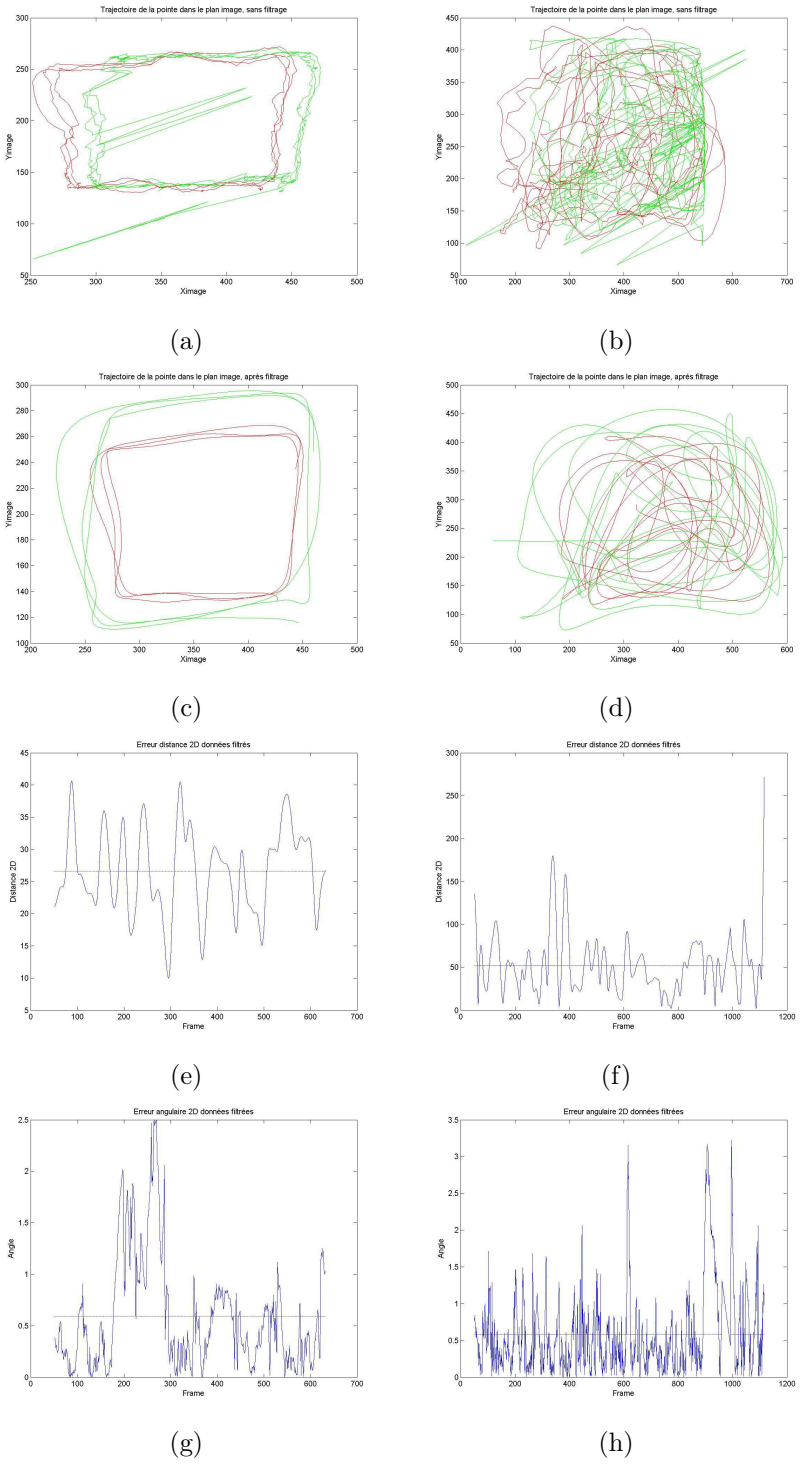


FIGURE 4.14. – Erreur sur la détermination de l'orientation et de la position de la pointe de l'instrument dans le plan image : (a) et (b) sont les projections des trajectoires non filtrées, (c) et (d) sont les projections des trajectoires après élimination des mesures aberrantes et lissage, (e) et (f) sont les erreurs résiduelles dans le plan image (en nombre de pixels), (g) et (h) sont les erreurs angulaires entre les projections des axes dans le plan image.

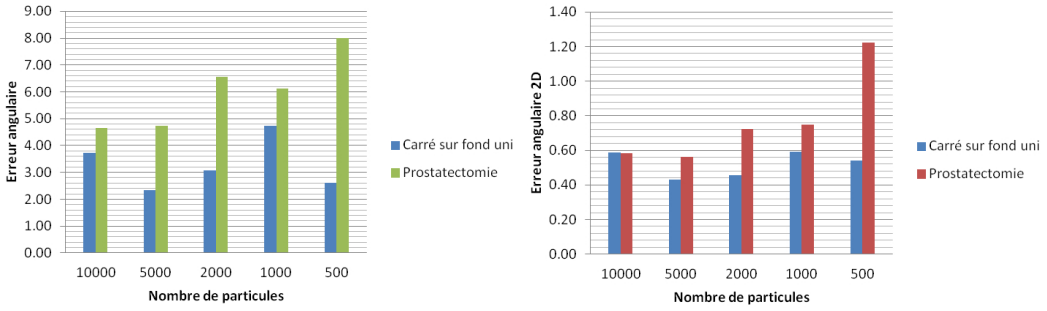
i) Influence du nombre de particules pour l'orientation de l'instrument

Le choix du nombre de particules représentant l'orientation de l'instrument dans l'espace est crucial pour la précision de la méthode, mais également pour sa vitesse d'exécution. Un grand nombre de particules réduit la probabilité d'absence de particule sur la cellule qui correspondrait à la meilleure observation, et réduit donc le nombre de mesures aberrantes (qui surviennent en cas d'absence de particules sur la cellule représentant l'orientation réelle).

De plus, la convergence vers l'orientation correcte au lancement du suivi est plus rapide, la sphère étant couverte avec une meilleure résolution. Ce critère est important en particulier pour des géodes ayant une haute résolution, pour lesquelles la convergence peut s'avérer longue si peu de particules sont présentes. Ainsi, pour une géode de résolution 9, dans une situation favorable (fond uni), l'orientation de l'instrument est déterminée de façon correcte (visuellement) au bout de 10 images si elle est décrite par 10000 particules, au bout de 25 images si elle est décrite par 500 particules. Cet effet est majoré si l'on utilise un fond plus complexe (prostatectomie par exemple, convergence vers l'orientation correcte après 60 et 125 images respectivement).

Enfin, en cas d'échec de la mesure à un instant t, un grand nombre de particules permet de converger à nouveau vers l'orientation correcte plus rapidement.

Sur les trajectoires présentées précédemment, l'effet du nombre de particules sur la précision de la mesure (mesurée par l'erreur angulaire) est minime si le fond est uni, mais est plus marqué si le fond est complexe (prostatectomie radicale par exemple). Cet effet est illustré par la figure 4.15.



(a) (b)

FIGURE 4.15. – Influence du nombre de particules sur la précision de la détermination de l'orientation de l'instrument, dans l'espace (a), et dans le plan image (b).

Par ailleurs, la figure 4.16 montre que la fréquence de calcul est fortement corrélée au nombre de particules utilisées, même si l'algorithme de CONDENSATION permet de concentrer celles-ci sur un nombre limité de cellules, et que la mesure dans l'image n'est effectuée qu'une fois par orientation possible. En fonction de l'utilisa-

tion envisagée de la méthode de suivi développée, il faudra donc faire un compromis entre vitesse de calcul et précision de la mesure.

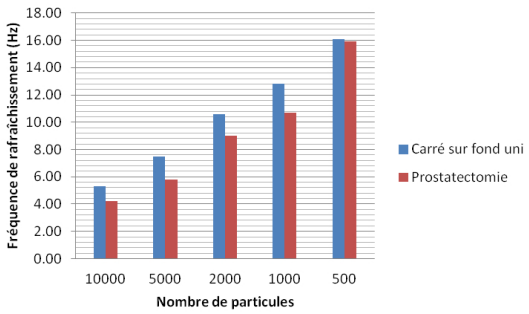


FIGURE 4.16. – Effet sur la vitesse de calcul du nombre de particules utilisées pour déterminer l'orientation de l'instrument.

ii) Influence du nombre de segments explorés pour identifier la pointe de l'instrument

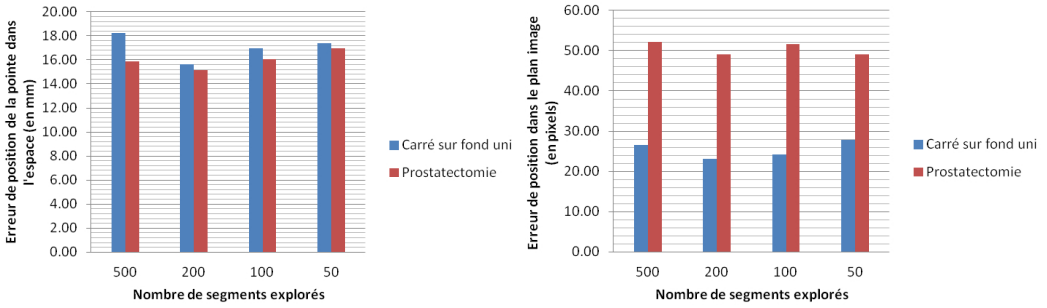
Pour un nombre de particules donné, le nombre de segments explorés pour déterminer la profondeur d'insertion de l'instrument n'a que peu d'influence sur la précision de la mesure évaluée en comparant aux données du suivi par marqueurs optiques, dans le domaine que j'ai exploré (100, 200 et 500 segments). On note cependant visuellement un discret effet avec de nombreuses trémulations autour de la position de la pointe si l'on utilise seulement 100 segments. Ces trémulations sont probablement lissées par l'application de splines, ce qui explique l'absence de différence constatée sur la figure 4.17.

Le nombre de segments explorés, dans le domaine étudié ici, n'a aucune influence sur la vitesse de calcul de la méthode.

iii) Influence de la résolution de la grille géodésique

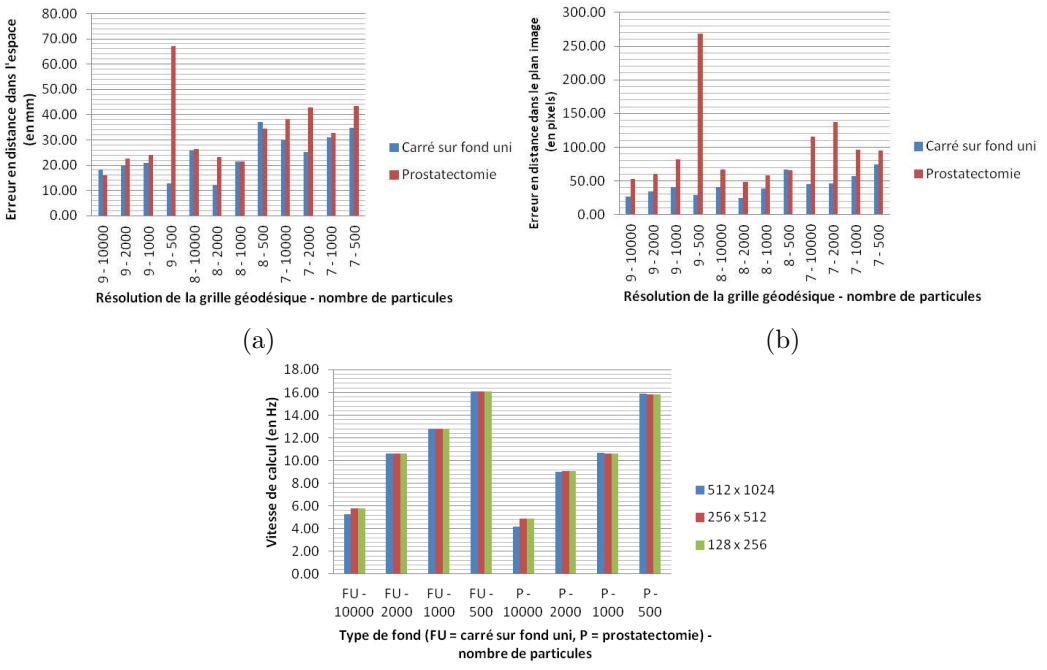
Comme illustré dans le chapitre précédent (tableau 3.1), la résolution de la grille géodésique, dont les cellules représentent les orientations discrètes de l'instrument, a une importance majeure pour la précision du suivi. Je compare dans la figure 4.18 les résultats obtenus avec des grilles géodésiques d'ordre 7, 8 et 9, en reprenant la notation définie dans la section 3.2.6 (soit des panneaux de 128×256, 256×512 et 512×1024 cellules). Je compare également dans cette figure les vitesses de calcul pour chaque résolution.

On constate que l'erreur augmente discrètement lorsque on diminue la résolution de la géode, en particulier lors du passage de l'ordre 8 à l'ordre 7, ce qui correspond



(a) (b)

FIGURE 4.17. – Influence du nombre de segments explorés le long de l'axe de l'instrument sur la précision de la détermination de la profondeur d'insertion de celui-ci, dans l'espace (a), et dans le plan image (b). On utilise ici 10000 particules pour décrire l'orientation de l'instrument, l'erreur angulaire est constante (= 3.77° et 4.57° en fonction du fond utilisé), quel que soit le nombre de segments explorés.



(a) (b)

(c)

FIGURE 4.18. – Influence de la résolution de la grille géodésique sur la précision de la détermination de l'orientation de l'instrument, dans l'espace (a) et dans le plan image (b). La résolution de la géode n'influe pas sur la vitesse de calcul de la méthode (c).

au phénomène attendu. On note surtout que l'erreur est très importante dans la situation où un faible nombre de particules est utilisé avec une grille hautement résolue, dans le cas d'un fond complexe. L'analyse de l'enregistrement vidéo de la procédure de suivi met en évidence la très grande instabilité de la détection : le temps de convergence vers l'orientation correcte de l'instrument est long, et le faible nombre de particules utilisé limite l'existence de plusieurs pics de densité de probabilité qui pourraient permettre, en cas d'échec ponctuel de la mesure, de converger à nouveau rapidement vers l'orientation correcte.

Cet exemple illustre l'importance du choix conjoint du nombre de particules et de la résolution de la géode, en fonction de la précision recherchée : il est contre-productif de vouloir compenser un faible nombre de particules par une résolution élevée de la géode, de même qu'il est inutile de multiplier le nombre de particules en cas d'utilisation d'une géode peu résolue (dans une optique de suivi en temps réel par exemple).

On remarque enfin que la vitesse de calcul ne dépend pas de la résolution de la géode dans la situation que j'étudie ici (caméra fixe). En revanche, la durée de l'initialisation des matrices *correspondance* et *masque*, définies dans la section 3.2.6, est dépendante de la résolution choisie (taille des matrices en 2^{2q+1} où q est l'ordre de la grille). Si l'on s'oriente vers une utilisation avec une caméra mobile, la matrice *masque* doit être mise à jour à chaque déplacement de la caméra (durée approximative de $0.03s$ pour une grille d'ordre 7, de $0.64s$ pour une matrice d'ordre 9). La matrice *correspondance*, pour des petits déplacements itératifs de la caméra tels que ceux effectués par un robot porte-endoscope, n'a pas besoin d'être mise à jour car sa rotation est compensée par les déplacements stochastiques des particules à la surface de la grille géodésique. Ce délai de mise à jour de la matrice *masque* doit donc être pris en compte lors de l'élaboration de dispositifs de suivi utilisant une caméra mobile.

iv) Influence du type de fond

Comme on peut le noter sur les figures précédentes, le type de fond utilisé pour évaluer la méthode de suivi d'instruments a une influence à la fois sur la précision et sur la vitesse de calcul de celle-ci. En effet, l'utilisation d'un fond uni permet d'une part de faciliter le traitement de l'image (seuils de Canny facilement déterminés avec la méthode décrite section iv)), d'autre part de ne pas avoir de segments de droites parasites pouvant être confondus avec un bord. L'ensemble des particules converge donc sur un très faible nombre de cellules de la grille géodésique, limitant ainsi le nombre de mesures à effectuer dans l'image. À l'inverse, une image plus complexe (fond de prostatectomie par exemple) fait apparaître des segments de droite n'appartenant pas à l'instrument, mais qui pourraient appartenir à la projection d'un des bords de celui-ci dans une orientation spécifique (faux-positif). Ces segments de droite concentrent autour de leur position une certaine fraction des particules, ce qui multiplie le nombre de mesures à réaliser dans l'image, et ralentit par conséquent

	Carré sur fond uni	Prostatectomie
Filtre Sigma modifié	4.9 ips	4.0 ips
Filtre bilatéral	5.3 ips	4.2 ips

Tab. 4.1. – Vitesses de calcul de la méthode de suivi en fonction du filtrage appliqué à l'image de saturation.

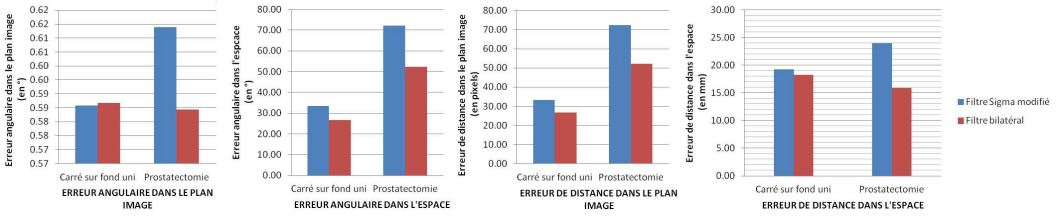


FIGURE 4.19. – Effet des filtres utilisés pour réduire le bruit dans l'image sur la précision de la détection de l'instrument.

la vitesse de calcul. De plus, le voisinage des bords de l'instrument étant variable en fonction de l'orientation de celui-ci dans l'image, l'image binaire obtenue par le filtre de Canny ne laisse pas apparaître aussi nettement les droites que dans le cas d'un fond uni, pouvant conduire à une évaluation moins précise de la position des bords dans l'image, et par conséquent à une diminution de la précision de la méthode.

v) Influence du choix du filtre

Comme expliqué à la fin de la section ii), l'implémentation optimisée du filtre Sigma modifié par Doignon et al. [2005] permet d'obtenir des vitesses de calcul sensiblement identiques à celles du filtre bilatéral pour les diamètres de filtre utilisés (tableau 4.1). Il existe une discrète différence en faveur du filtre bilatéral sur les exemples étudiés dans cette partie (figure 4.19), qui ne se retrouve cependant pas systématiquement dans l'ensemble des enregistrements avec d'autres fonds. A contrario, il n'existe pas, pour aucun fond utilisé, de différence en faveur du filtre Sigma. Je continue donc à utiliser le filtre bilatéral dans la suite de ce travail.

vi) Influence du nombre d'instruments suivis

Un des intérêts de la méthode développée au cours de cette thèse est la possibilité de suivre simultanément plusieurs instruments chirurgicaux dans la cavité abdominale, chaque instrument étant associé à un point d'insertion qui lui est propre. Le nombre d'instruments suivis n'a aucune influence sur la précision de la méthode, le suivi de chaque instrument correspondant à un processus distinct (seule la phase de traitement de l'image est commune aux différents processus). En revanche la vitesse de calcul est fortement corrélée au nombre d'instruments suivis : elle est

Nombre de particules	Suivi d'un instrument	Suivi de deux instruments
10000	5.8 ips	4.0 ips
2000	12.8 ips	8.0 ips
1000	15.1 ips	10.6 ips

Tab. 4.2. – Vitesses de calcul de la méthode de suivi en fonction du nombre d'instruments suivis. Vitesses établies avec affichage en temps réel de l'image originale avec les droites correspondant aux bords supposés de l'instrument, de l'image de Canny, et enregistrement du flux vidéo et des positions détectées de l'instrument. Ces vitesses peuvent donc être améliorées en choisissant les informations pertinentes à afficher ou récupérer selon l'utilisation envisagée.

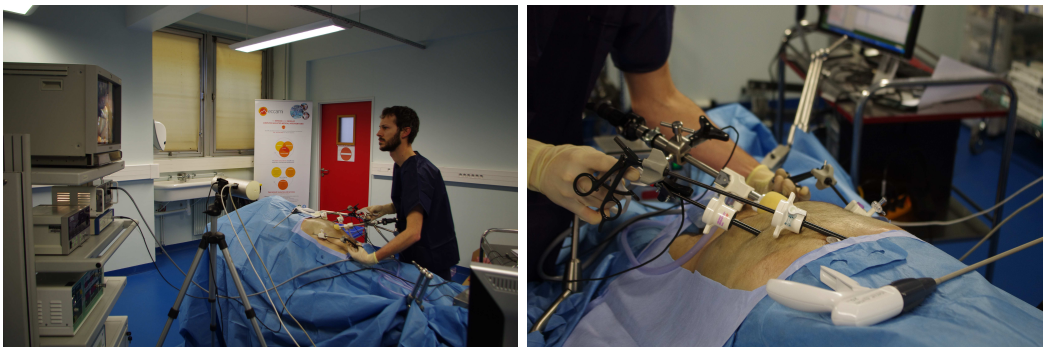
diminuée environ d'un tiers pour le suivi de deux instruments par rapport au suivi d'un unique instrument (figure 4.2), quel que soit le nombre de particules utilisées pour la détermination de l'orientation de l'instrument.

4.2.6 Évaluation sur spécimen anatomique

La méthode ayant montré des résultats encourageants sur banc d'essai, j'ai voulu l'évaluer dans le contexte d'utilisation pour lequel elle a été initialement développée, c'est à dire dans des conditions de laparoscopie chez l'homme.
Les manipulations ont été effectuées au Laboratoire d'Anatomie Des Alpes Françaises de la faculté de médecine de Grenoble, avec l'aide d'un chirurgien pédiatre (Dr. Yohann Robert). Les images étaient enregistrées grâce à une colonne d'endoscopie Karl Storz, avant d'être envoyées vers l'ordinateur via le boîtier d'acquisition, avec une résolution de 768×576. Des marqueurs optiques étaient rigidement liés à l'endoscope et à l'instrument principal du chirurgien, qui avait comme consigne d'éviter au maximum les occultations entre les marqueurs optiques et la caméra accuTrack 250 (figure 4.20). Le calibrage était réalisé de façon similaire à celui décrit plus haut.

Les interventions enregistrées étaient des cholécystectomies effectuées chez des spécimens anatomiques formolés.

Les données acquises au cours de ces manipulations n'ont malheureusement pas pu être exploitées, d'une part en raison de problèmes techniques propres à la chirurgie (manipulation très difficile des tissus après fixation par le formol, pneumo-péritoine de très faible volume, brides,...), d'autre part en raison d'une trop mauvaise qualité des images enregistrées par rapport aux images présentées dans le chapitre 3. Cette mauvaise qualité des images est probablement liée à un mauvais paramétrage du boîtier d'acquisition, ainsi qu'à l'ancienneté du matériel disponible au moment de la manipulation au laboratoire d'anatomie.



(a) (b)

FIGURE 4.20. – Acquisition de données en conditions réelles au laboratoire d'anatomie.

4.3 Discussion et conclusion pour le suivi d'instruments

J'ai présenté et évalué dans les deux chapitres précédents une méthode originale permettant de déterminer la position dans l'espace de plusieurs instruments de laparoscopie, à partir d'images endoscopiques. Cette approche s'appuie uniquement sur la connaissance a priori de la géométrie de l'instrument, et peut donc être aisément adaptée pour différentes configurations de chirurgie laparoscopique.

4.3.1 Limites actuelles de la méthode proposée

Les résultats obtenus jusqu'à présent montrent un suivi de bonne qualité visuelle, avec cependant des erreurs dans la détermination de la position de la pointe dans l'espace. Ces erreurs ont plusieurs origines, outre l'erreur de calibrage extrinsèque, que je résume ici.

▷ **Erreur dans l'étape de calibrage intrinsèque de la caméra :** cette erreur est liée, d'une part, à l'approximation de la caméra par un modèle de sténopé qui, même s'il est largement utilisé dans la communauté de vision par ordinateur, comporte des limites (correction incomplète de la distorsion radiale, [Hartley et Zisserman, 2003]) et, d'autre part, à la procédure de calibrage à l'aide d'une mire plane : la caméra comporte en réalité une lentille, et possède donc une profondeur de champ limitée. Les images des points de la mire ne se situant pas dans l'intervalle de profondeur adéquat sont floues, et limitent donc la précision de la phase de calibrage. La projection des points de l'espace sur le plan image à l'aide des éléments de la matrice de calibrage intrinsèque ne sera donc pas exacte.

▷ **Erreur dans la détection des bords :** comme illustré par la figure 4.2, la détermination de l'orientation de l'axe dans l'espace est très sensible aux erreurs dans

l'estimation des bords de l'instrument dans l'image. En particulier, si les bords de l'instrument apparaissent flous (taux de rafraîchissement de la caméra trop faible ou profondeur de champ inadaptée), il est possible d'avoir un décalage d'un ou plusieurs pixels pour un des quatre points représentant les extrémités des droites définissant les bords. Cette erreur est majorée dans le contexte d'utilisation de la méthode sur banc d'entraînement, dans lequel l'instrument est situé à une distance plus importante de la caméra que dans une situation de chirurgie réelle. Cette sensibilité n'avait pas, à ma connaissance, été mise en évidence dans les publications antérieures se rapportant au suivi d'instruments dans l'espace. Cette erreur pourrait être limitée grâce à l'utilisation d'une caméra haute-définition pour l'acquisition des images, avec un taux de rafraîchissement élevé, similaire aux caméras que l'on peut trouver à l'heure actuelle dans les blocs chirurgicaux. Cependant l'augmentation de la taille de l'image est susceptible de ralentir considérablement la vitesse du suivi. L'effet d'un sous-échantillonnage de ces images sur la précision de la méthode de suivi doit alors être évalué.

▷ **Erreur dans la détection de la position de la pointe :** la détermination de la profondeur d'insertion de l'instrument le long de son axe directeur est à l'heure actuelle l'étape la moins robuste de la méthode proposée. Si cette détermination paraît simple théoriquement (détection d'une région frontière entre deux domaines), je n'ai pas réussi dans le cadre de ce travail à identifier une approche suffisamment robuste pour considérer cette étape comme résolue. Dans les conditions favorables étudiées plus haut, le taux d'échec de localisation reste faible, mais augmente lorsque deux instruments sont présents dans l'image et que leurs axes se croisent à proximité de leurs extrémités.

Les principaux paramètres influençant le niveau d'erreur et la vitesse de calcul de la méthode proposée ont été décrits plus haut, il s'agit du nombre de particules décrivant l'orientation de l'instrument et de la résolution de la géode, ces deux paramètres devant varier ensemble afin de permettre une convergence suffisamment rapide des particules vers l'orientation de l'instrument représentée par une cellule sur la grille géodésique. Le choix de ces paramètres représente un compromis entre la précision du suivi dans l'espace et la vitesse de calcul : une grille géodésique d'ordre 7, associée à 500 particules, permet un suivi de la position de la pointe de l'instrument dans l'image de bonne qualité visuelle, avec une fréquence de calcul proche de la fréquence d'acquisition de la caméra, mais convergera plus lentement vers l'orientation correcte de l'instrument dans l'espace et sera plus sensible aux mesures aberrantes ; a contrario une grille géodésique d'ordre 9, associée à 10000 particules, permet un suivi de la position de la pointe de l'instrument dans l'espace de qualité satisfaisante, avec une vitesse de convergence élevée vers l'orientation correcte et peu de mesures aberrantes, au prix d'une vitesse de calcul plus lente (environ égale au cinquième de la fréquence d'acquisition des images par la caméra).

4.3.2 Apports principaux de la méthode proposée

Malgré ces limites, mon approche présente un certain nombre d'apports par rapport aux méthodes existantes.

▷ **Suivi simultané de plusieurs instruments sur un fond complexe :** la contrainte imposée par la localisation du point d'insertion dans l'espace permet d'exclure de l'analyse des zones de l'image pouvant être confondues avec des bords de l'instrument, mais dont l'orientation n'est pas compatible avec les orientations accessibles pour ce dernier. De plus, plusieurs instruments peuvent être suivis simultanément, chacun d'entre eux étant identifié par la position de son point d'insertion.

▷ **Représentation simplifiée des orientations dans l'espace :** l'utilisation d'une grille géodésique sphérique représentant les orientations possibles de l'instrument, centrée sur le point d'insertion de celui-ci, permet de manipuler directement les orientations de l'instrument dans l'espace. Les propriétés de quasi-homogénéité et de quasi-isotropie de cette grille permettent par ailleurs de modéliser les variations d'orientation de l'instrument de façon simple dans un repère plan orthogonal grâce au changement de repère présenté dans la section 3.2.6. Enfin la connaissance de la position du point d'insertion dans le repère de la caméra permet de réduire l'espace d'observation des orientations dans la grille, certaines orientations n'étant pas associées à une projection de l'instrument dans l'image.

▷ **Utilisation d'un filtre préservant les angles dans l'image :** cette technique avait été utilisée par Doignon et al. [2006], mais avec un filtre différent de celui utilisé dans ce travail. Une comparaison des performances obtenues avec les deux types de filtre ne montre pas d'effet marqué sur la qualité du suivi de l'un par rapport à l'autre. En revanche l'utilisation de l'un ou l'autre de ces filtres par rapport à l'absence de filtrage réduit fortement la quantité d'angles détectés dans l'image, et par conséquent le risque d'obtenir des bords ne correspondant pas aux bords de l'instrument dans l'image.

▷ **Détermination automatique des seuils pour l'algorithme de Canny :** cette étape correspond à une étape d'apprentissage de l'information contenue dans l'image pour ne conserver que l'information pertinente d'un point de vue statistique. Ce travail est, à ma connaissance, la première utilisation de cette approche dans le domaine du suivi d'instruments en chirurgie. La détermination automatique des paramètres de l'algorithme est primordiale si l'on souhaite utiliser ce type d'approche dans le contexte du bloc opératoire, où l'on observe une grande variabilité des conditions d'illumination, et où le chirurgien ne doit pas avoir d'action spécifique à réaliser en plus de son geste au cours de celui-ci.

▷ **Utilisation de l'algorithme de CONDENSATION :** cet algorithme présente l'intérêt de permettre le suivi de propagation de densités de probabilité pour modéliser les variations de l'orientation de l'instrument. Dans le domaine du suivi d'instruments laparoscopiques, cet algorithme avait déjà été utilisé, mais les particules ne modélisaient pas l'orientation de l'instrument dans l'espace (position de la pointe dans le plan image dans [Speidel et al., 2006], orientation des bords de l'instrument dans le plan image dans [Climent et Hexsel, 2012]). L'utilisation de ce type d'algorithme permet de réduire considérablement le nombre d'observations à effectuer dans l'image, celles-ci n'étant réalisées que si la probabilité a priori que l'instrument soit dans les orientations associées est non nulle. Ainsi 1000 particules suffisent à localiser l'instrument sur une grille géodésique d'ordre 9 (comportant donc $\approx 2.10^6$ cellules).

4.3.3 Optimisations envisagées

Un certain nombre d'optimisations doivent être envisagées afin d'améliorer les résultats obtenus jusqu'à présent.

▷ **Amélioration de la détection de la position du point d'insertion :** la détermination de la position du point d'insertion de l'instrument dans la cavité abdominale, dans le repère de la caméra, est actuellement réalisable de façon automatique supervisée ou manuellement. La détection automatique des bords de l'instrument dans l'image, à l'aide de la transformée de Hough, a cependant un taux d'échec proche de 50% dans les conditions d'utilisation du bloc opératoire. Il s'agit donc avant tout de réduire le nombre d'images à acquérir avant d'obtenir une paire de droites correspondant aux bords de l'instrument dans l'image. Les pistes envisagées sont :

- d'une part de prendre en compte l'orientation du gradient dans l'image par rapport à l'orientation de la droite paramétrée par les coordonnées de Hough (on s'attend à ce que la direction du gradient soit orthogonale à la direction de la droite), dans une approche similaire à O'Gorman et Clowes [1976] ;
- d'autre part d'utiliser des méthodes heuristiques afin de ne sélectionner que les paires de droites pour lesquelles l'orientation relative et la distance entre les deux droites est compatible avec la projection des bords de l'instrument dans l'image.

▷ **Réduction du bruit dans la détermination de l'orientation :** actuellement la mesure est fortement bruitée par l'existence de mesures aberrantes et de variations fines autour de l'orientation exacte de l'instrument dans l'espace. Je traite pour le moment ce bruit en temps différé, avec les méthodes présentées plus haut. Il serait intéressant d'être capable de contrôler ce bruit en temps réel. Les mesures aberrantes en terme de vitesse angulaire sont déjà filtrées en temps réel (avec un seuil fixé manuellement), mais celles concernant la vitesse dans le repère cartésien

(correspondant à une variation très rapide de la profondeur d'insertion) ne le sont pas pour le moment. Il est techniquement envisageable de déterminer les distributions de ces vitesses au cours d'une phase d'apprentissage au début de la chirurgie, et de détecter ensuite les mesures aberrantes avec la même approche que précédemment. Cependant on envisagera plutôt l'implémentation d'un filtre prédictif de Kalman [Kalman, 1960], qui permet d'obtenir une estimation de la position actuelle à partir :

- de la position à l'instant précédent,
- de l'observation (c'est à dire la position estimée par l'algorithme de CONDENSATION dans ce contexte),
- du modèle d'évolution dynamique du système (identique à celui de l'algorithme de CONDENSATION),
- et du modèle de l'observation (comprenant en particulier la matrice de covariance du bruit de l'observation).

Cette approche présente l'avantage de mettre à jour, à chaque itération, la matrice de covariance de la composante stochastique du modèle dynamique, de façon plus fine que la méthode utilisée actuellement (où je fais varier le coefficient multiplicateur de cette matrice en fonction de la vitesse angulaire de l'instrument à l'itération précédente) ; elle permet également de prendre en compte le bruit de l'observation dans l'estimation de la position à l'instant t, en mettant à jour sa matrice de covariance à chaque itération.

▷ **Robustesse de la détermination de la profondeur d'insertion :** la principale étape limitante actuellement pour obtenir un suivi robuste est la détermination de la position de la pointe de l'instrument le long de son axe. L'optimisation de cette étape en cours, grâce à un travail en collaboration avec Raphaël Sznitman (École Polytechnique Fédérale de Lausanne, Suisse) visant à fusionner l'approche développée ici et l'approche qu'il présente dans [Sznitman et al., 2013] dans une seule méthode. Mon approche permettra de détecter l'orientation de l'instrument dans l'espace et de localiser grossièrement la position de la pointe dans l'image, afin de réduire le domaine d'exploration des classifieurs (arbres de décision) utilisés par son approche, qui permettent une localisation fine de la position de la pointe, indépendamment de son orientation.

▷ **Autres pistes de réflexion :** plusieurs autres optimisations sont envisageables à moyen terme pour améliorer la qualité du suivi :

- L'incertitude concernant la détermination de l'orientation de l'instrument dans un plan perpendiculaire au plan image pourrait être réduite en utilisant une caméra stéréoscopique, constituée d'une paire de caméras C_1 et C_2 rigidement liées, et dont la matrice de calibrage extrinsèque pour passer du repère $\Re_{c1}$ au repère $\Re_{c2}$ est connue. La visualisation de la scène chirurgicale sous deux angles différents permet d'effectuer deux mesures distinctes pour une orientation pos-

sible de l'instrument, ce qui devrait accroître la précision de la méthode. Une autre utilisation possible de ce type de caméra serait d'utiliser deux processus distincts (c'est à dire modéliser deux instruments passant par le même point d'insertion), un pour chaque caméra. Si les poids des meilleures observations sont similaires, et que les orientations associées à ces observations sont proches sur la géode, on peut alors considérer que l'orientation réelle de l'instrument est l'intersection des plans contenant chacun le centre optique d'une caméra et la droite représentant l'axe de l'instrument détecté par cette caméra. En cas d'échec d'une des deux mesures (poids de la meilleure observation faible), l'orientation de l'instrument est donnée par l'autre processus.

- De façon assez similaire, l'utilisation de mini-caméras supplémentaires, réparties dans la cavité abdominale selon l'approche proposée par Boschet [2010], pourrait permettre d'augmenter de façon très notable la précision de la méthode, en identifiant la position de l'instrument selon deux points de vue très différents.
- Certaines des améliorations suggérées ci-dessus sont susceptibles de ralentir la vitesse de calcul pour le suivi d'instrument. De plus, j'ai montré dans la section 4.2.5 qu'il est nécessaire d'utiliser un nombre élevé de particules pour obtenir un suivi de qualité satisfaisante dans l'espace. La vitesse de calcul n'est dans cette situation pas compatible avec un suivi de la position en temps réel. Or la structure des données utilisée dans cette approche se prête parfaitement à une implémentation parallèle de la méthode de suivi sur une architecture GPU (*Graphic Processing Unit*), à la fois pour le traitement de l'image et pour l'évolution et l'observation des particules représentant l'orientation de l'instrument. L'approche de Sznitman et al. [2013] pour la détection fine de la pointe de l'instrument est également parallélisable.
- Enfin il sera nécessaire d'implémenter une méthode de mise à jour des valeurs des seuils pour l'algorithme de Canny, qui soit capable d'opérer en tâche de fond (*multi-threading*), pour s'adapter aux différentes phases de la chirurgie.

4.3.4 Conclusion

La méthode de suivi d'instruments présentée ici est encore imparfaite. En effet, on note que l'évaluation de la position de la pointe de l'instrument dans l'espace ne suit pas exactement la trajectoire enregistrée par le dispositif de marqueurs optiques, malgré l'élimination des mesures aberrantes et le lissage par des splines. Si on peut discuter la précision du calibrage extrinsèque permettant de recaler les trajectoires obtenues avec les deux méthodes d'acquisitions dans un même repère (aspect d'homothétie persistant, aboutissant à une erreur systématique), il faut constater que les formes des trajectoires ne sont pas strictement superposables, avec en particulier des fluctuations dans le plan perpendiculaire au plan image pour la trajectoire déterminée avec ma méthode.

Les erreurs dans l'estimation de la position dans l'espace ne permettent donc pas pour le moment d'envisager l'utilisation de cette méthode pour mesurer des distances par rapport à des structures cibles, identifiées par d'autres modalités (fluorescence et échographie trans-rectale par exemple pour la prostatectomie [Voros et al., 2013]) et recalées avec le repère de la caméra, ou directement segmentées dans l'image endoscopique [Sánchez-González et al., 2011, Yip et al., 2012]. Il n'est pas non plus possible de déclencher des alertes en cas de trop forte proximité avec des structures sensibles, comme suggéré par Speidel et al. [2008], ce qui était un des objectifs initiaux de ce travail.

Une autre limite rencontrée dans cette partie de mon travail a été la difficulté à acquérir des images de chirurgie de qualité suffisante avec une caméra fixe pour évaluer la robustesse de la méthode développée dans des conditions d'utilisation proches de la réalité. Le traitement de l'image présenté ici a été développé à l'aide d'images extraites d'enregistrements de procédures chirurgicales, et devrait donc fonctionner dans les conditions d'une chirurgie réelle, mais je n'ai pas pu évaluer la partie dynamique de l'approche (suivi des particules) dans ces conditions.

Actuellement, la détermination de la position de la pointe de l'instrument dans le plan image est suffisamment rapide et précise pour permettre l'asservissement des mouvements du robot porte-endoscope à la position de la pointe. Cet asservissement fait partie des travaux de la thèse d'Anthony Agustinos. En choisissant d'utiliser 2000 particules se déplaçant sur une grille géodésique d'ordre 8, il parvient à contrôler les déplacements du robot ViKY® afin que la pointe reste en permanence au centre d'une image imitant un fond chirurgical. Un calibrage main-œil lui permet, à partir des commandes passées au robot, de mettre à jour pour chaque mouvement effectué la position du point d'insertion dans le repère lié à la caméra, et ainsi de poursuivre le suivi de l'instrument au cours du temps.

Malgré les limites explicitées ci-dessus, les trajectoires de la pointe de l'instrument dans l'espace, obtenues sur des fonds chirurgicaux, sont suffisamment proches de celles déterminées par le dispositif de marqueurs optiques pour analyser la qualité du geste chirurgical à partir de métriques dérivées de ces trajectoires, dans le contexte d'un banc d'entraînement. Cette analyse est l'objet de la partie suivante de ce manuscrit.

Références

Christophe Boschet : Laparoscopie Répartie. Thèse de doctorat, Université de Grenoble, 2010. 159

Joan Climent et Roberto A. Hexsel : Particle filtering in the hough space for instrument tracking. Computers in Biology and Medicine, 42(5):614–623, 2012. 157

C Doignon, P Graebling et M de Mathelin : Real-time segmentation of surgical instruments inside

the abdominal cavity using a joint hue saturation color feature. Real-Time Imaging, 11(5-6):429–442, décembre 2005. 152

C. Doignon, F. Nageotte et M. de Mathelin : The role of insertion points in the detection and positioning of instruments in laparoscopy for robotic tasks. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006, page 527–534, 2006. 156

J Michael Fitzpatrick, Derek LG Hill, Calvin R Maurer Jr et Milan Sonka : Image registration. In Handbook of medical imaging : medical image processing and analysis, volume 2, pages 447–513. SPIE-International Society for Optical Engineering, 2000. 143

R. Hartley et A. Zisserman : Multiple view geometry in computer vision. Cambridge Univ Pr, 2003. 154

Jiping He et Chun-Xiang Tian : A statistical smoothness measure to eliminate outliers in motion trajectory tracking. Human movement science, 17(2):189–200, 1998. 135, 137

RE Kalman : A new approach to linear filtering and prediction theory. Trans. ASME J. Basic Eng, 82:35–45, 1960. 158

Frank O'Gorman et MB Clowes : Finding picture edges through collinearity of feature points. Computers, IEEE Transactions on, 100(4):449–456, 1976. 157

Ignacio Oropesa, Magdalena K Chmarra, Patricia Sánchez-González, Pablo Lamata, Sharon P Rodrigues, Silvia Enciso, Francisco M Sánchez-Margallo, Frank-Willem Jansen, Jenny Dankelman et Enrique J Gómez : Relevance of Motion-Related assessment metrics in laparoscopic surgery. Surgical innovation, septembre 2012. PMID : 22983805. 135

Patricia Sánchez-González, Alicia M. Cano, Ignacio Oropesa, Francisco M. Sánchez-Margallo, Francisco Del Pozo, Pablo Lamata et Enrique J. Gómez : Laparoscopic video analysis for training and image-guided surgery. Minimally Invasive Therapy & Allied Technologies, 20(6):311–320, décembre 2011. 160

Stefanie Speidel, Michael Delles, Carsten Gutt et Rüdiger Dillmann : Tracking of instruments in minimally invasive surgery for surgical skill analysis. In Guang-Zhong Yang, TianZi Jiang, Dinggang Shen, Lixu Gu et Jie Yang, éditeurs : Medical Imaging and Augmented Reality, numéro 4091 in Lecture Notes in Computer Science, pages 148–155. Springer Berlin Heidelberg, janvier 2006. 157

Stefanie Speidel, Gunther Sudra, Julien Senemaud, Maximilian Drentschew, Beat Peter Müller-Stich, Carsten Gutt et Rüdiger Dillmann : Recognition of risk situations based on endoscopic instrument tracking and knowledge based situation modeling. pages 69180X–69180X–8, mars 2008. 160

Raphael Sznitman, Rogerio Richa, Russell H. Taylor, Bruno Jedynak et Gregory D. Hager : Unified detection and tracking of instruments during retinal microsurgery. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(5):1263–1273, 2013. 158, 159

S. Voros, A. Moreau-Gaudry, B. Tamadazte, G. Custillon, R. Heus, M.-P. Montmasson, F. Giroud, O. Gaiffe, C. Pieralli, G. Fiard, J.-A. Long, J.-L. Descotes, C. Vidal, A. Nguyen-Dinh et P. Cinquin : Devices and systems targeted towards augmented robotic radical prostatectomy. IRBM, 2013. 160

M. C. Yip, D. G. Lowe, S. E. Salcudean, R. N. Rohling et C. Y. Nguan : Tissue tracking and registration for Image-Guided surgery. IEEE Transactions on Medical Imaging, 31(11):2169–2182, 2012. 160

Troisième partie

Analyse de la qualité de gestes chirurgicaux à partir de connaissances a priori

CHAPITRE 5

ACQUISITION DE GESTES CHIRURGICAUX SUR UN BANC D'ENTRAÎNEMENT MISTELS

Sommaire

5.1 Banc d'entraînement MISTELS et score FLS 166
5.2 Évaluation de la qualité du geste à l'aide du score GOALS 169
5.3 Données acquises 170
5.3.1 Volontaires 170
5.3.2 Chaîne d'acquisition des données 171
5.3.3 Exemple de trajectoires obtenues 172
5.4 Métriques explorées 173

La méthode développée et présentée dans la partie précédente de ce manuscrit permet d'obtenir les trajectoires de la pointe d'instruments de laparoscopie dans l'espace de travail du chirurgien, dans les conditions favorables du banc d'entraînement. Je me concentre donc dans la suite de ce travail sur l'acquisition de gestes dans ce contexte afin d'analyser leur qualité.

Ma méthode impose de connaître la position du point d'insertion dans le repère de la caméra à tous les instants de la chirurgie, soit en utilisant une caméra fixe, soit en utilisant un robot porte-endoscope. Cette contrainte ne permet pas encore d'utiliser cette méthode dans des blocs chirurgicaux variés, condition nécessaire pour récolter suffisamment de données afin d'être significatif au moment de l'analyse.

D'autre part, l'utilisation d'un banc d'entraînement permet d'évaluer les gestes de chirurgiens à des stades plus variés de leur progression, allant du novice (externe) au chirurgien expert, facilitant l'identification de paramètres pertinents pour évaluer la qualité de réalisation de tâches standardisées. Il sera nécessaire dans un second temps de contrôler la portabilité des métriques identifiées sur banc d'entraînement

vers le bloc opératoire, et d'évaluer leur sensibilité pour différencier un geste sub-optimal d'un geste optimal.

Je cherche dans ce chapitre à identifier les métriques dérivées des trajectoires permettant de prédire la qualité du geste. L'objectif principal est de prédire la qualité du geste de manière globale ; l'objectif secondaire est d'apporter une information supplémentaire au chirurgien en lui précisant les domaines de compétence technique mis en jeu au cours de la chirurgie (dextérité bi-manuelle, perception de la profondeur,...) dans lesquels il n'est pas optimal.

Les métriques identifiées et les modèles prédictifs seront spécifiques à différentes tâches représentatives des tâches effectuées au cours d'une procédure chirurgicale réelle, les compétences recherchées pouvant varier d'une tâche à l'autre.
Le chapitre suivant explorera la possibilité d'appliquer ces métriques pour analyser des gestes réalisés en situation d'entraînement.

5.1 Banc d'entraînement MISTELS et score FLS

J'utilise le banc d'entraînement à la chirurgie laparoscopique MISTELS, développé par le groupe de Fried à Toronto, et déjà présenté dans la section 2.2.2.i). Il permet la réalisation de tâches standardisées, inspirées par les tâches nécessaires à la réalisation de cholécystectomie, d'appendicectomie, de cure de hernie inguinale et de fundoplicature de Nissen par voie laparoscopique [Fried et al., 2004], et est couramment utilisé aux USA pour l'évaluation des compétences techniques des internes au sein du programme *Fundamentals of Laparoscopic Surgery* (FLS). La fiabilité de la méthode de notation associée et la validité du dispositif pour prédire les compétences techniques requises au bloc opératoire ont donc été largement évaluées et démontrées.

Le dispositif consiste en un banc d'entraînement présentant deux points d'insertion pour placer des trocarts, situés de part et d'autre d'une caméra CCD dont la position et le grossissement sont fixes. Cette disposition standardisée correspondant à celle rencontrée en général en chirurgie laparoscopique.

Plusieurs exercices doivent être réalisés, chacun mimant une tâche faisant appel à différentes compétences techniques du chirurgien. Le niveau de granularité auquel je définis les métriques prédictives de la qualité est donc situé à un niveau inférieur à celui classiquement analysé dans la littérature du domaine (qui se situe plutôt au niveau des sous-tâches ou des gestes, voir section 2.2.3.ii)).

Je présente ici les différents exercices à réaliser (illustrés figure 5.1), accompagnés des paramètres permettant le calcul de leur score [Derossis et al., 1998, Fraser et al.,

2003]. D'une façon générale les scores sont calculés selon la règle suivante :

$$S = \max(0, 100\frac{T_{seuil} - T_{volontaire} - P}{T_{normalisation}})$$

où $T_{volontaire}$ est le temps nécessaire au chirurgien pour compléter la tâche, T_{seuil} est un temps limite à ne pas dépasser, P est une pénalité et $T_{normalisation}$ est un temps permettant de comparer les scores des différents exercices entre eux. Les trois paramètres T_{seuil}, $T_{normalisation}$ et P sont spécifiques à chaque exercice. Le groupe de Fried propose de ne pas affecter de score négatif, ce choix sera discuté plus loin.

Exercice 1 → transfert d'objets : le volontaire doit saisir chacun des 6 pions depuis un côté du plateau, le transférer à sa seconde main et le placer sur l'autre côté du plateau. La procédure est ensuite inversée. T_{seuil} est fixé à 300s, $T_{normalisation}$ est fixé à 237s, P est calculée comme le pourcentage de pions non transférés suite à une chute hors du champ de vision de la caméra. Cet exercice est supposé explorer la coordination œil -main, la dextérité bi-manuelle et la perception de la profondeur.

Exercice 2 → découpage d'un patron en forme de cercle : le volontaire doit découper, le long d'un pourtour circulaire dessiné de 5cm de diamètre, une compresse non-tissée de 10×10cm. La compresse est suspendue entre une pince horizontale et deux pinces crocodile. T_{seuil} est fixé à 300s, $T_{normalisation}$ est fixé à 142s. P est calculée comme le pourcentage d'aire comprise dans les écarts par rapport aux bords du patron rapportée à l'aire totale du cercle. Cet exercice est supposé explorer la dextérité bi-manuelle.

Exercice 3 → placement d'un Endoloop : un nœud coulant pré-noué doit être positionné au niveau d'un marquage spécifique sur un tube en mousse modélisant un appendice. T_{seuil} est fixé à 180s, $T_{normalisation}$ est fixé à 280s, P est calculée comme la distance entre le marquage sur le tube en mousse et la position finale du nœud (en mm). Une pénalité de 50 unités est ajoutée si le nœud est mal serré. Cet exercice est supposé explorer la dextérité bi-manuelle.

Exercice 4 → suture extra-corporelle : une suture doit être réalisée pour fermer une ouverture pratiquée longitudinalement dans un drain de penrose, en passant par deux points marqués sur ses bords. Le fil monobrin utilisé mesure 90cm avec une aiguille courbe. La suture doit être fermée en faisant les nœuds en extra-corporel, puis en les poussant en utilisant un instrument adapté (au moins 3 nœuds plats requis). T_{seuil} est fixé à 420s, $T_{normalisation}$ est fixé à 297s, P est calculée comme la somme des distances entre les points d'entrée/sortie du fil et les points marqués sur le drain, à laquelle s'additionne l'écart entre les deux bords si la suture a échoué à les rapprocher (distances en mm). Une pénalité de 20 unités est ajoutée si les nœuds sont mal serrés. Cet exercice est supposé explorer la préhension des tissus, la dextérité bi-manuelle et la perception de la profondeur.

Exercice 5 → suture intra-corporelle : l'objectif est le même que pour l'exercice 4, à la différence près que la suture doit être fermée en utilisant des nœuds plats intra-corporels (double nœud / simple nœud / simple nœud). Le fil tressé utilisé mesure 15cm avec une aiguille courbe. T_{seuil} est fixé à 600s, $T_{normalisation}$ est fixé à 520s, P est calculée de la même manière que pour l'exercice 4. Cet exercice est supposé explorer la préhension des tissus, la dextérité bi-manuelle et la perception de la profondeur.

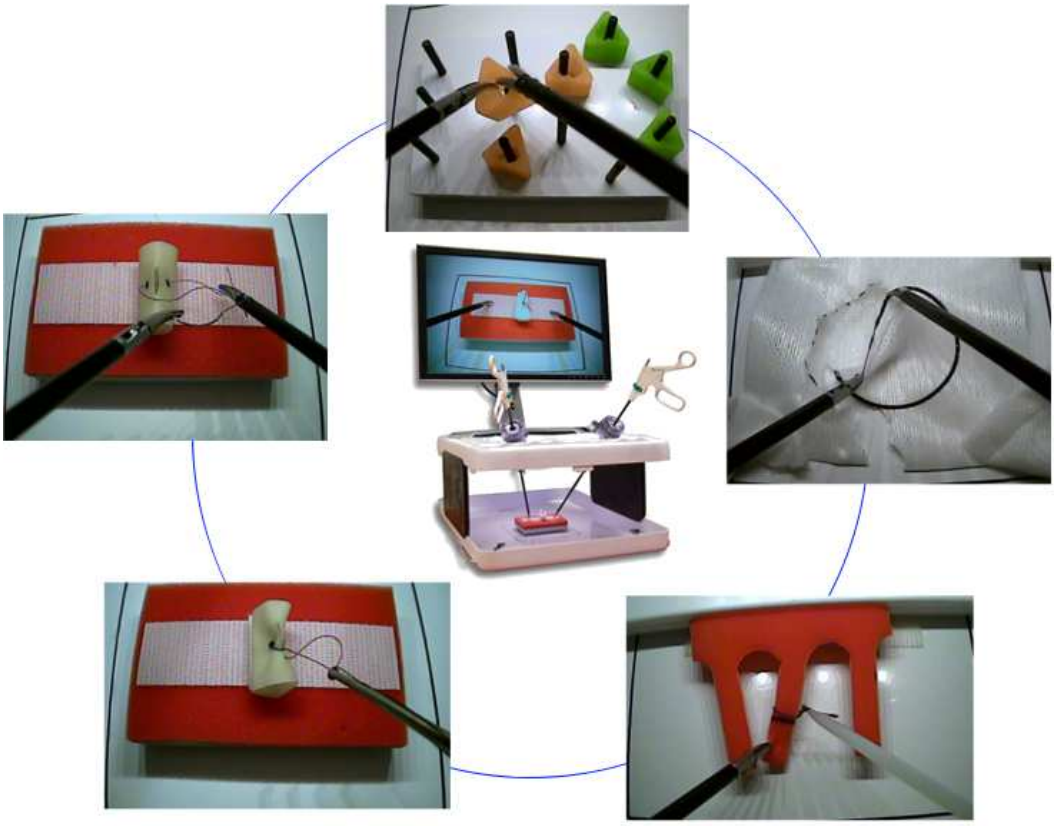


FIGURE 5.1. – Illustration des exercices du banc d'entraînement MISTELS. Exercices 1 à 5 présentés dans le sens des aiguilles d'une montre, en partant de midi. Vue globale du dispositif au centre.

Avant la réalisation de chaque exercice, une vidéo présentant l'exécution de celui-ci par un chirurgien expert en laparoscopie est présentée au volontaire. Les règles de calcul des pénalités sont explicitées oralement, et il est demandé aux volontaires de ne pas se focaliser sur le temps de complétion de l'exercice, mais plutôt sur l'exécution correcte de celui-ci (pas de pénalités). Cette ordre de priorité est le plus réaliste par rapport à l'exécution d'une procédure chirurgicale réelle. Les volontaires ont la possibilité de demander des conseils au cours de la réalisation de l'exercice s'ils ne

parviennent pas à progresser vers l'objectif. Chaque exercice est réalisé deux fois pour augmenter le volume de données.

Je réalise le chronométrage des exercices, la mesure des écarts permettant le calcul des pénalités et l'apport de conseils pour l'ensemble des volontaires (observateur externe unique).

Le score proposé pour évaluer la qualité du geste du chirurgien dans le cadre du programme FLS présente cependant plusieurs limites :

- son calcul n'est pas automatisé, c'est-à-dire qu'il nécessite un observateur externe pour mesurer les pénalités (le chronométrage pourrait facilement être effectué de façon automatique)
- il retourne une mesure globale pour chaque tâche réalisée, mais sans apport explicatif : il ne précise pas au chirurgien quelles sont les composantes de son habileté technique qui ont été mises en défaut (ou qui à l'inverse sont optimales) lors de l'exécution de la tâche. Or cet aspect explicatif est primordial pour guider et accélérer l'apprentissage et l'amélioration du geste chirurgical.

Je réalise donc une seconde évaluation objective de la qualité du geste à l'aide d'une échelle de notation plus explicative, présentée ci-dessous.

5.2 Évaluation parallèle de la qualité du geste à l'aide du score GOALS

Je m'oriente vers l'utilisation de l'échelle GOALS (*Global Operative Assessment of Laparoscopic Skills*, présentée section 2.2.2ii)) pour l'évaluation de la qualité du geste au cours de la réalisation des exercices décrits plus haut, en raison de son caractère explicatif (évaluation de différentes composantes de la qualité du geste chirurgical) et de sa portabilité vers le bloc opératoire, qui est le contexte vers lequel on cherche à s'orienter à moyen terme. On pourra donc disposer pour différents contextes d'une échelle de référence unique pour établir les modèles permettant de prédire la qualité du geste à partir des métriques enregistrées.

L'évaluation de la qualité du geste par le score GOALS est réalisée en temps différé, à partir des enregistrements vidéo de l'exécution des exercices, encodés à l'aide d'un algorithme de compression sans perte (HuffYUV). La notation des quatre premières composantes du score (dextérité bi-manuelle, perception de la profondeur, efficience et préhension) est effectuée par deux chirurgiens, un expert et un intermédiaire en laparoscopie, de façon aveugle par rapport à la qualité du volontaire évalué, et indépendamment l'un de l'autre. Je réalise la notation de la cinquième composante (autonomie) en fonction de l'aide apportée oralement.

La grille de notation GOALS n'a cependant jamais été évaluée de manière rigoureuse pour sa validité et sa fiabilité dans le contexte d'utilisation du banc d'entraînement MISTELS, en particulier avec une évaluation différée à l'aide d'enregistrements vidéo. La validité de concurrence a été démontrée entre le score FLS et une échelle de notation OSATS, associant une échelle globale relativement proche de celle de GOALS et des *check-lists* spécifiques aux exercices, mais je n'ai pas trouvé dans la littérature de preuve de la validité de concurrence avec une échelle globale utilisable au bloc opératoire.
Je montrerai donc, dans la section 6.2, en pré-requis de la suite de l'analyse, que les scores FLS et GOALS sont bien corrélés pour l'évaluation de la qualité de réalisation de chaque exercice sur le banc d'entraînement MISTELS.

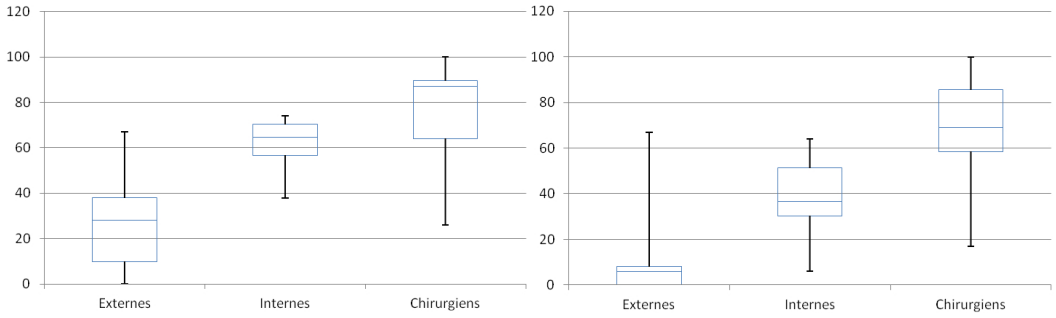
5.3 Données acquises

5.3.1 Volontaires

J'ai pu recruter pour cette première étude utilisant ma méthode de détection de l'instrument 32 volontaires au sein du CHU de Grenoble et de cliniques du bassin grenoblois.

Les volontaires se répartissaient schématiquement en trois groupes :

- 13 externes (étudiants en $4^{\text{ème}}$, $5^{\text{ème}}$ ou $6^{\text{ème}}$ année des études médicales), considérés comme novices. Ils avaient des expériences variées en chirurgie, certains ayant déjà assisté à des interventions par laparoscopie en tant que spectateurs (médiane : 10 interventions, min : 0, max : 20), ou en tant qu'aide-opérateur (médiane : 0 intervention, min : 0, max : 10).
- 7 internes de chirurgie (digestive, urologique, gynécologique ou pédiatrique), dans des années de formation différentes ($3^{\text{ème}}$ semestre à $10^{\text{ème}}$ semestre), auxquels s'ajoute un chef de clinique en chirurgie orthopédique qui n'a pas d'autre expérience en laparoscopie que sa pratique récente au cours de son internat. Tous les internes avaient déjà participé en tant qu'aide opératoire à des chirurgies laparoscopiques (médiane : 120 interventions, min : 40, max : 400), certains ayant déjà conduit seuls des interventions simples de type appendicectomie ou laparoscopie exploratrice (médiane : 3, min : 0, max : 300).
- 11 chirurgiens provenant de différents services (urologie, chirurgie digestive, chirurgie pédiatrique, chirurgie thoracique et endocrinienne) dans plusieurs établissements de santé, et ayant des expériences variées en laparoscopie, allant de la pratique occasionnelle à la pratique quasi-exclusive de la chirurgie endoscopique. Tous les chirurgiens avaient assisté à plus de 100 interventions en tant qu'aides opératoires, et avaient déjà conduit des interventions complexes en laparoscopie en tant que responsable de l'équipe chirurgicale (médiane : 175, min : 2, max : $>$ 5000).



(a) Compréhension de la chirurgie laparoscopique

(b) Expertise en chirurgie laparoscopique

FIGURE 5.2. – Diagramme à moustaches des auto-évaluations de la compréhension et de l'expertise en chirurgie laparoscopique par les volontaires, à l'aide d'une échelle EVA, en fonction de leur statut. Le rectangle s'étend du premier au troisième quartile, coupé par la médiane. Les segments rejoignent les valeurs extrêmes.

Aucun des participants n'avait d'entraînement préalable sur le banc d'entraînement MISTELS. Ils devaient réaliser à deux reprises chacun des cinq exercices, après avoir visionné la vidéo de démonstration, au cours d'une seule séance pour limiter le biais d'apprentissage. La durée d'une séance s'étalait entre 30 minutes pour les meilleurs chirurgiens et 2 heures pour les novices les plus en difficulté.

Afin d'avoir une première estimation, sommaire et subjective, de la séparabilité des groupes, j'ai demandé aux participants d'évaluer leur degré de compréhension et leur degré d'expertise en chirurgie laparoscopique, à l'aide d'une échelle visuelle analogique, avant de démarrer les exercices. Je définis la compréhension par la connaissance des gestes à exécuter et du rôle des instruments au cours de la chirurgie, et l'expertise par l'aisance manuelle et la maîtrise au cours de la réalisation de gestes techniques.

Je présente les résultats de cette auto-évaluation dans la figure 5.2. On constate que les groupes apparaissent assez bien séparés, en particulier pour l'évaluation du degré d'expertise. Je discute plus loin des limites de ce type d'évaluation et des biais qu'elle comporte en fonction des résultats obtenus lors de la réalisation des exercices.

5.3.2 Chaîne d'acquisition des données

Les enregistrements vidéo ont été réalisés en récupérant le signal RCA fourni par la caméra (25 images par seconde) installée de façon fixe dans le banc d'entraînement MISTELS (VTI Medical, Inc), en le numérisant grâce un boîtier d'acquisition (DFG/USB2pro, The Imaging Source), puis en stockant chaque image dans un fi-

chier bitmap séparé, et en reconstruisant la séquence vidéo au format AVI avec une compression sans perte HuffYUV a posteriori.

Je stocke chaque image séparément car j'enregistre, en parallèle du signal vidéo, la position des instruments acquise par le dispositif accuTrack présenté plus haut (marqueurs optiques). L'objectif de cet enregistrement parallèle est d'augmenter ma base de données en vue de comparer les deux approches, et éventuellement de mesurer les métriques sur les trajectoires acquises par le dispositif accuTrack. La matrice de calibrage extrinsèque entre la caméra et le repère lié à l'accuTrack était déterminée pour chaque volontaire.

Afin de ne pas modifier la réalisation du geste, j'ai demandé aux volontaires de ne pas prendre en compte la visibilité des marqueurs optiques dans le dispositif accuTrack au cours des exercices. En conséquence, il existe trop de données manquantes sur la position des instruments (situations où la ligne de vue entre les marqueurs et la caméra accuTrack est interrompue en raison de la rotation de l'instrument sur son propre axe) pour envisager de calculer des métriques pertinentes.

Le volontaire disposait d'un affichage spécifique (écran LCD 20 pouces avec entrée RCA) pour visualiser le signal vidéo, indépendant de la partie enregistrement des données, grâce à un dédoublement du signal RCA en amont du boîtier d'acquisition. Cet affichage était placé de façon similaire à un écran utilisé dans un bloc chirurgical (légèrement sur-élevé et décalé latéralement par rapport au banc d'entraînement).

Les paramètres nécessaires au fonctionnement de ma méthode de suivi d'instruments, c'est-à-dire la matrice de calibrage intrinsèque de la caméra et les positions des points d'insertion, étaient déterminés à chaque redémarrage du dispositif. Il n'existe pas de modification significative de ces paramètres d'une détermination à l'autre. Chaque vidéo était cependant analysée avec les paramètres déterminés lors de son enregistrement.

5.3.3 Exemple de trajectoires obtenues

Les trajectoires des instruments ont été déterminées avec la méthode que j'ai décrite dans le chapitre 3, en utilisant 10000 particules pour représenter l'orientation de l'instrument à l'aide d'une grille géodésique d'ordre 8 (655362 cellules), et 500 segments répartis le long de l'axe de l'instrument pour déterminer la position de la pointe le long de celui-ci.
Les trajectoires ainsi obtenues ont été filtrées pour éliminer les mesures aberrantes puis lissées par une spline cubique. La figure 5.3 illustre quelques trajectoires obtenues pour différents exercices.

Les métriques définies section 5.4 ont alors été calculées, pour chaque volontaire, pour chaque répétition de chaque exercice, avant l'analyse statistique décrite dans le chapitre suivant.

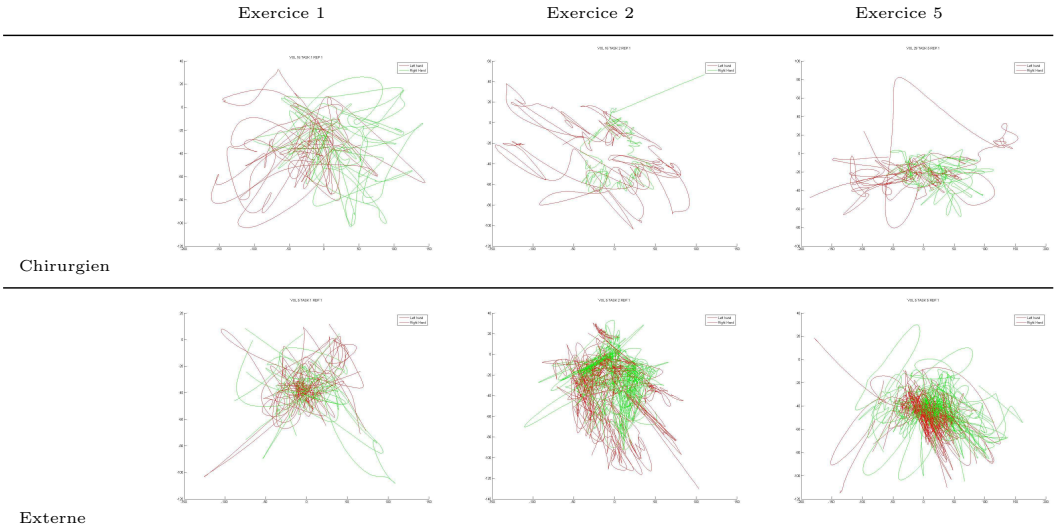


FIGURE 5.3. – Exemples de trajectoires pour différents exercices. On observe chez le chirurgien une organisation ordonnée pour chacun des gestes, que l'on ne retrouve pas chez le novice. Ainsi, pour l'exercice 1, chaque instrument explore son domaine avec des mouvements directs pour aller chercher les plots (boucles latérales) avec une zone de transfert d'une main à l'autre au centre de l'espace de travail. De même, pour l'exercice 2, la main gauche réalise une traction sur la compresse dans différentes directions (boucles de la trajectoire rouge) pendant que les ciseaux en main droites découpent le cercle (trajectoire verte au centre).

5.4 Métriques explorées

Les métriques dérivées de la trajectoire dont je souhaite analyser le caractère prédictif pour quantifier la qualité du geste chirurgical, en étudiant leur corrélation au score GOALS et à ses composantes, sont multiples. Certaines sont issues de la littérature, d'autres sont proposées ici pour la première fois à ma connaissance. Je présente ci-dessous les métriques étudiées, la façon dont je les obtiens à partir de la trajectoire des instruments dans l'espace et la raison pour laquelle elles me paraissent intéressantes à intégrer au modèle prédictif.
Je note $X(t)$ la position de la pointe de l'instrument dans l'espace à l'instant t dans la trajectoire filtrée et lissée par la spline, $X_o(t)$ la position de la pointe de l'instrument dans l'espace à l'instant t dans la trajectoire filtrée mais pas encore lissée.
Certaines métriques sont des fonctions de la durée du mouvement et de la longueur de chemin parcouru, raison pour laquelle j'applique des normalisations par ces valeurs.

temps de complétion de l'exercice T : il s'agit du temps durant lequel un instru-

ment est présent dans l'image au cours de la réalisation de l'exercice (c'est à dire le nombre de positions retenues après filtrage de la trajectoire). Cette métrique paraît évidente, et est retrouvée dans la quasi-totalité de la littérature sur le sujet.

longueur de chemin parcouru PL **:** cette mesure est susceptible de refléter l'efficience du geste ; pour une même action, un chirurgien expert minimisera le déplacement de l'instrument afin d'économiser ses gestes. On a $PL = \int_0^T \left|\left|\frac{\mathrm{d}X(t)}{\mathrm{d}t}\right|\right| \mathrm{d}t$. Cette métrique est également fréquemment retrouvée dans la littérature.

niveau de bruit B **:** il s'agit de la différence entre la trajectoire avant et après lissage par la spline. Le paramètre de lissage étant identique pour l'ensemble des volontaires, on peut supposer que cette mesure risque d'éliminer les mouvements parasites autour de la trajectoire imputables à l'opérateur en même temps qu'elle lisse le bruit de la mesure. Le bruit de la mesure pouvant être considéré comme constant, la différence entre la trajectoire originale et la trajectoire lissée peut servir d'indicateur de la quantité de mouvements parasites liés à l'opérateur. Je pose $B = \int_0^T ||X(t) - X_o(t)|| \, \mathrm{d}t$. Je normalise cette valeur d'une part par le temps (B_T), d'autre part par la longueur de chemin parcouru (B_{PL}).

vitesse moyenne de déplacement de l'instrument dans l'image V **:** cette métrique est davantage discutée dans la littérature, vitesse ne rimant pas nécessairement avec efficacité. On a $V = PL/T$.

le chemin angulaire parcouru AP **:** cette métrique mesure la concentration de l'orientation de l'instrument autour d'une direction donnée, c'est à dire que le chirurgien minimise l'aire angulaire explorée et est capable de travailler dans des espaces étroits. Elle est assez similaire à l'aire angulaire explorée proposée par Oropesa et al. [2012]. En notant $\mathbf{e}(t)$ l'orientation de l'instrument dans l'espace à un instant t, on obtient $AP = \int_0^T \left|\left|\frac{\mathrm{d}\theta}{\mathrm{d}t}\right|\right| \mathrm{d}t$ avec $\frac{\mathrm{d}\theta}{\mathrm{d}t} = \mathrm{atan2}\left(||\mathbf{e}(t) \otimes \mathbf{e}(t - \mathrm{d}t)||, \mathbf{e}(t) \cdot \mathbf{e}(t - \mathrm{d}t)\right)$, où $\mathrm{atan2}(x, y)$ est la fonction arctangente à deux arguments. Je normalise ce chemin angulaire par le temps (AP_T) et par la longueur de chemin parcouru (AP_{PL}).

l'accélération A **:** des grandes accélérations au cours du mouvement sont susceptibles de traduire des mouvements brutaux de la pointe de l'instrument. L'accélération s'écrit $A = \int_0^T \left|\left|\frac{\mathrm{d}V(t)}{\mathrm{d}t}\right|\right| \mathrm{d}t$, je la normalise par le temps (A_T) et par la longueur de chemin parcourue (A_{PL}).

la rugosité du mouvement S **:** la dérivée troisième de la position est appelée *jerk* dans la littérature anglo-saxonne : $J(t) = \frac{d^3|X(t)|}{dt^3}$. Les sciences du mouvement décrivent comme mesure de fluidité d'un geste l'intégrale du *jerk* : d'après Flash et Hogan [1985], le contrôle moteur des mouvements chez l'homme cherche à minimiser l'intégrale du *jerk* de l'effecteur final lors de la réalisation d'un geste. J'appelle dans ce travail rugosité l'intégrale du *jerk*, plutôt

que fluidité qui est son inverse. Chmarra et al. [2007] et Oropesa et al. [2012] ont étudié ce paramètre en le calculant sous la forme $S = \sqrt{\frac{T^5}{2.PL^2} \int_{t=0}^{T} J(t)^2 \mathrm{d}t}$, sans parvenir à montrer de corrélation avec la qualité du geste, tandis que Cotin et al. [2002] mettent en évidence une corrélation négative entre une forme plus simple de cette métrique ($S_T = \frac{1}{T} \int_0^T ||J(t)|| \, \mathrm{d}t$) et le niveau de pratique. J'étudie donc ces deux métriques, ainsi que l'intégrale du jerk normalisée par la longueur de chemin parcouru $S_{PL} = \frac{1}{PL} \int_0^T ||J(t)|| \, \mathrm{d}t$, pour évaluer leur caractère prédictif.

nombre de passage par zéro du jerk ZJ **:** une autre approche de la fluidité du geste est proposée par Brooks et al. [1974], qui dénombrent les pics d'accélération ou de vitesse pour mesurer la fluidité du geste. Cette mesure est assez proche du dénombrement des mouvements effectués au cours de la chirurgie, comme proposé par Aggarwal et al. [2007] entre autres. J'étudie donc le caractère prédictif du nombre de passages par zéro du *jerk*, avec une normalisation par le temps ZJ_T et par la longueur de chemin parcouru ZJ_{PL}.

variations de la profondeur d'insertion VP **:** celles-ci peuvent être calculées de deux façons, d'une part en mesurant la variation de profondeur le long de l'axe de l'instrument à chaque instant t [Cotin et al., 2002], d'autre part en projetant cette différence de profondeur sur l'orientation moyenne de l'instrument au cours de la tâche, qui est a priori dans la direction du centre de la zone de travail. Les variations de profondeur avec une orientation de l'instrument ne se dirigeant pas vers l'espace de travail sont probablement davantage liées à des mouvements non contributifs pour la réalisation du geste que celles se dirigeant vers la zone de travail. J'étudie donc les variables VP^a_T, VP^a_{PL} pour les variations le long de l'axe de l'instrument, et VP^d_T, VP^d_{PL} pour les variations le long de l'orientation moyenne de l'instrument.

volume exploré par la pointe de l'instrument dans l'espace VOL **:** le volume occupé par la pointe de l'instrument dans l'espace est calculé en utilisant des α-formes (ou α-*shapes* dans la littérature), qui sont une généralisation des enveloppes convexes pour des formes complexes [Akkiraju et al., 1995]. Une α-forme est un polyèdre, pas forcément convexe ni connecté, dérivé de la triangulation de Delaunay d'un ensemble de points, dont on peut contrôler le degré de détail en ajustant le paramètre α. J'utilise une fonction Matlab [1] pour calculer l'α-forme de chacune de mes trajectoires, avec un paramètre $\alpha = 1/10$ unique déterminé après ajustement sur les premières trajectoires enregistrées pour obtenir un volume visuellement pertinent (cela revient à considérer qu'un point n'est connecté à ses voisins que si la distance les séparant est inférieure à 10mm), puis pour calculer le volume de cette forme. On s'attend à ce que la qualité du geste soit inversement corrélée au volume exploré (figure 5.4).

1. http://www.mathworks.com/matlabcentral/fileexchange/28851-alpha-shapes/content/alphavol.m

Cette métrique n'a à ma connaissance jamais été étudiée sous cette forme.

taux de recouvrement des volumes RV : les chirurgiens experts croisent très peu leurs instruments au cours d'une procédure chirurgicale, c'est-à-dire que chaque instrument (main gauche et main droite) possède son volume d'évolution, et que le recouvrement de ces volumes est faible. Je calcule donc le volume de l'α-forme issue des points des deux trajectoires VOL_{global} et calcule $RV = \frac{VOL_{gauche}+VOL_{droite}}{2VOL_{global}}$. De même que pour le volume, cette métrique n'a jamais été proposée.

économie de volume EV : il s'agit du rapport volume exploré sur chemin parcouru. Cette métrique avait déjà été étudiée par Oropesa et al. [2012] avec une définition rudimentaire du volume exploré. Pour une longueur de chemin égale, on s'attend à ce que le volume exploré par le chirurgien soit plus faible que pour le novice. On a $EV = VOL/PL$.

mesures de coordination entre les mains du chirurgien : cette métrique paraît pertinente pour mesurer la dextérité bi-manuelle du chirurgien, mais n'a à ma connaissance pas été explorée jusqu'à présent. Il s'agit d'étudier les distributions de probabilité de métriques définies pour chacune des mains, et de rechercher des points communs entre ces distributions. Je définis plusieurs métriques :

- une première approche simple consiste à étudier le coefficient de corrélation de la régression linéaire entre les vitesses de la main gauche et de la main droite, et d'observer dans quelle mesure ces vitesses sont linéairement corrélées entre elles avec la métrique R^2_V.
- l'observation de la répartition des couples de vitesses (V_{gauche}, V_{droite}), d'accélération (A_{gauche}, A_{droite}) et de *jerk* (J_{gauche}, J_{droite}) dans le plan $(main_{gauche}, main_{droite})$ permet d'identifier une répartition spatiale de ces couples sous forme de croix définie par les axes du plan (en particulier pour l'accélération et le *jerk*, figure 5.4), traduisant le fait que les opérateurs ne font pas varier simultanément les vitesses des instruments de la main gauche et de la main droite. Je calcule donc la distance totale à la croix de l'ensemble de ces points, en normalisant par le nombre de points, ainsi que la dispersion des points par rapport à l'origine.
- enfin je recherche l'information mutuelle contenue dans les distributions de ces trois paramètres (V, A et J), qui mesure la dépendance statistique entre les variables (l'information contenue dans l'observation d'une variable apporte-t-elle des informations sur la réalisation de la variable jointe ?). Cette mesure de la coordination s'écrit

$$I(X,Y) = \sum_{y\in Y}\sum_{x\in X} p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right)$$

où $p(x,y)$ est la fonction de distribution de probabilité jointe des variables x et y (qui représentent les paramètres de la main gauche et de la main

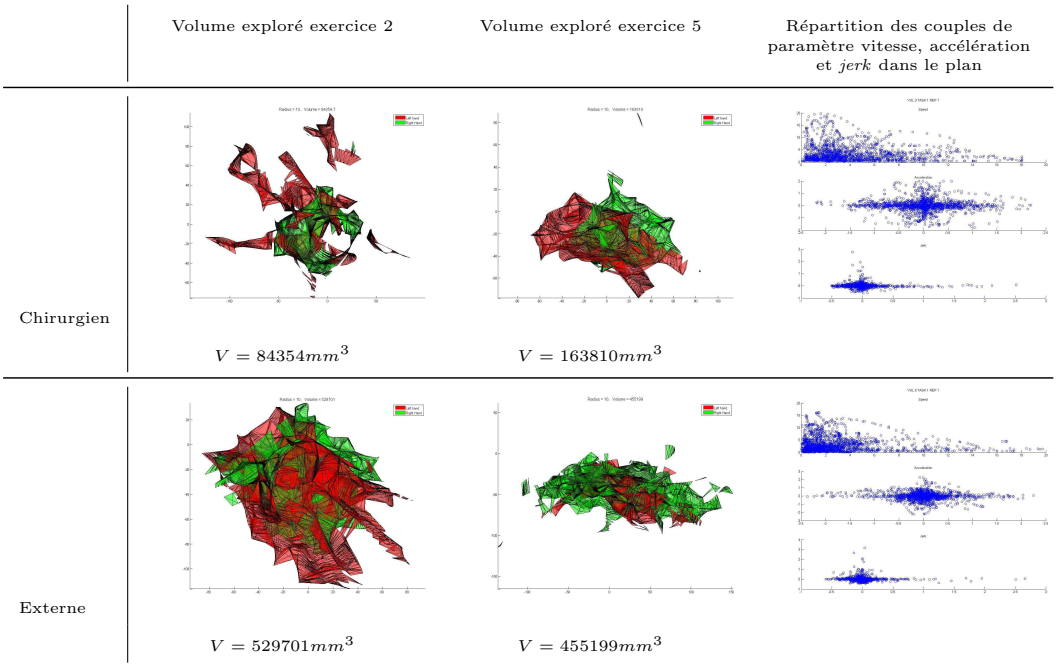


FIGURE 5.4. – Exemples de volumes et de couplage entre les paramètres vitesse, accélération et jerk pour différents exercices : volumes explorés pour les exercices 2 (découpe d'un patron) et 5 (suture intra-corporelle), pour lesquels l'espace de travail est bien défini dans l'espace donc où l'économie de volume paraît pertinente ; couples de paramètres pour l'exercice 1 (transfert d'objet) où la coordination entre les deux mains paraît pertinente.

droite) et $p(x)$ et $p(y)$ sont leurs distributions marginales. J'utilise pour calculer l'information mutuelle la méthode proposée par Peng et al. [2005], et implémentée dans Matlab [2].

Pour l'analyse de l'exercice 3 (placement d'un Endoloop), je n'utilise que les paramètres de la main dominante, le dispositif de positionnement de l'Endoloop ne permettant pas un suivi fiable (diamètre non constant, non standardisé). De même dans l'exercice 4 (suture extra-corporelle), j'arrête l'analyse de la trajectoire de l'instrument à partir de l'instant où les deux extrémités du fil sont sorties de la cavité de travail, les instruments n'étant plus visibles à partir de ce moment.
Si les deux instruments sont utilisés, chaque métrique utilisée ultérieurement est la moyenne des métriques correspondantes de chaque main.

2. http://www.mathworks.com/matlabcentral/fileexchange/14888

Les métriques ainsi définies ont pu être mises en œuvre dans des scénarios d'entraînement réels, ce qui fait l'objet du chapitre suivant.

Références

R. Aggarwal, T. Grantcharov, K. Moorthy, T. Milland, P. Papasavas, A. Dosis, F. Bello et A. Darzi : An evaluation of the feasibility, validity, and reliability of laparoscopic skills assessment in the operating room. Annals of surgery, 245(6):992, 2007. 175

N Akkiraju, H Edelsbrunner, M Facello, P Fu, E Mücke et C Varela : Alpha shapes : definition and software. pages 63–66, 1995. 175

V. B. Brooks, J. D. Cooke et J. S. Thomas : The continuity of movements. In R. B. Stein, K. G. Pearson, R. S. Smith et J. B. Redford, éditeurs : Control of Posture and Locomotion, numéro 7 in Advances in Behavioral Biology, pages 257–272. Springer US, janvier 1974. 175

M. K. Chmarra, W. Kolkman, F. W. Jansen, C. A. Grimbergen et J. Dankelman : The influence of experience and camera holding on laparoscopic instrument movements measured with the TrEndo tracking system. Surgical Endoscopy, 21(11):2069–2075, novembre 2007. 175

Stephane Cotin, Nicholas Stylopoulos, Mark Ottensmeyer, Paul Neumann, David Rattner, Steven Dawson, Takeyoshi Dohi et Ron Kikinis : Metrics for laparoscopic skills trainers : The weakest link ! Medical Image Computing and Computer-Assisted Intervention — MICCAI 2002, pages 35–43, janvier 2002. 175

Anna M. Derossis, Gerald M. Fried, Michal Abrahamowicz, Harvey H. Sigman, Jeffrey S. Barkun et Jonathan L. Meakins : Development of a model for training and evaluation of laparoscopic skills. The American Journal of Surgery, 175(6):482–487, 1998. 166

T Flash et N Hogan : The coordination of arm movements : an experimentally confirmed mathematical model. The Journal of neuroscience : the official journal of the Society for Neuroscience, 5(7):1688–1703, juillet 1985. PMID : 4020415. 174

S.A. Fraser, D.R. Klassen, L.S. Feldman, G.A. Ghitulescu, D. Stanbridge et G.M. Fried : Evaluating laparoscopic skills. Surgical Endoscopy, 17(6):964–967, 2003. 10.1007/s00464-002-8828-4. 166

Gerald M. Fried, Liane S. Feldman, Melina C. Vassiliou, Shannon A. Fraser, Donna Stanbridge, Gabriela Ghitulescu et Christopher G. Andrew : Proving the value of simulation in laparoscopic surgery. Annals of Surgery, 240(3):518–528, septembre 2004. PMID : 15319723 PMCID : PMC1356442. 166

Ignacio Oropesa, Magdalena K Chmarra, Patricia Sánchez-González, Pablo Lamata, Sharon P Rodrigues, Silvia Enciso, Francisco M Sánchez-Margallo, Frank-Willem Jansen, Jenny Dankelman et Enrique J Gómez : Relevance of Motion-Related assessment metrics in laparoscopic surgery. Surgical innovation, septembre 2012. PMID : 22983805. 174, 175, 176

Hanchuan Peng, Fuhui Long et Chris Ding : Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 27(8):1226–1238, 2005. 177

CHAPITRE 6

ANALYSE DE MÉTRIQUES DÉRIVÉES DES TRAJECTOIRES POUR ÉVALUER LA QUALITÉ DES GESTES CHIRURGAUX

Sommaire

6.1 Reproductibilité intra- et inter-observateurs 180
6.2 Validité du score GOALS pour le banc d'essai MISTELS . 182
6.2.1 Validité concurrente 182
6.2.2 Étude des composantes du score GOALS expliquant le score MISTELS 184
6.2.3 Étude de la validité prédictive 186
6.3 Prédiction d'un score à partir des métriques acquises 189
6.3.1 Outils statistiques utilisés 189
6.3.2 Plan d'analyse 192
6.3.3 Métriques significatives dans le modèle prédictif pour chaque exercice 193
6.3.4 Prédiction d'un score à partir de l'ensemble des métriques . . 195
6.3.5 Prédiction d'un score à partir des métriques significatives de chaque exercice 200
6.4 Classification en fonction des métriques uniquement 203
6.5 Prédiction de chaque aptitude du score GOALS 204
6.6 Discussion et conclusion pour l'analyse de gestes chirurgicaux 206
6.6.1 Banc d'entraînement MISTELS 207
6.6.2 Définition du score de référence 208
6.6.3 Choix des métriques explorées 208
6.6.4 Méthode de prédiction 209
6.6.5 Limites de la prédiction 209
6.6.6 Apports principaux de la méthode 210

6.6.7 Conclusion 211

Ce chapitre est consacré à l'analyse du caractère significatif des métriques définies au chapitre précédent pour quantifier la qualité du geste chirurgical dans le cadre du banc d'entraînement MISTELS, ainsi qu'à la création d'un modèle permettant de prédire cette qualité à partir de la connaissance a priori de la distribution de ces métriques. Cette analyse a été réalisée avec l'aide des statisticiens du Centre d'Investigation Clinique - Innovation Technologique rattaché au Centre Hospitalier Universitaire de Grenoble, à l'aide du logiciel R v. 2.15.0 [R Core Team, 2012].

▷ **Notations** Afin de simplifier les notations, je confonds dans la suite de ce chapitre le nom du programme d'évaluation FLS et le nom du support associé, MISTELS. Le score MISTELS de chaque exercice est donc noté M_j, le score moyen pour l'ensemble des exercices M. Je note également $G_j^{(i)}$ le score GOALS, avec en exposant la composante étudiée ($1 \rightarrow$ *perception de la profondeur*, $2 \rightarrow$ *dextérité bimanuelle*, $3 \rightarrow$ *efficience*, $4 \rightarrow$ *préhension*, $5 \rightarrow$ *autonomie*) et en indice l'exercice sur lequel le score est analysé. Je note également $G^{(i)} = \frac{1}{5}\sum_{j=1}^{5} G_j^{(i)}$, $G_j = \frac{1}{5}\sum_{i=1}^{5} G_j^{(i)}$ et $G = \frac{1}{5}\sum_{j=1}^{5} G_j$, respectivement les moyennes selon une composante, un exercice et la moyenne globale d'un volontaire.

6.1 Reproductibilité intra- et inter-observateurs pour le score GOALS

L'évaluation du score GOALS a été portée par deux chirurgiens de façon indépendante, après anonymisation des vidéos. La répartition des scores en fonction du chirurgien qui porte l'évaluation est présentée figure 6.1 ; afin de quantifier la reproductibilité inter-observateur, nous avons calculé les coefficients de corrélation intra-classe, tandis que nous avons calculé le coefficient α de Cronbach [Cronbach, 1951] pour évaluer la reproductibilité intra-observateur pour deux répétitions du même exercice. Les résultats sont présentés tableau 6.1.

		Coefficient	IC95%
Reproductibilité intra-observateur	Observateur 1	$\alpha = 0.858$	[0.837 ;0.879]
	Observateur 2	$\alpha = 0.858$	[0.832 ;0.877]
Reproductibilité inter-observateur		ICC = 0.75	[0.69 ;0.8]

Tab. 6.1. – Reproductibilité intra-observateur (coefficient α de Cronbach) et inter-observateur (coefficient de corrélation intra-classe) du score GOALS

La reproductibilité intra-observateurs est très bonne, la reproductibilité inter-

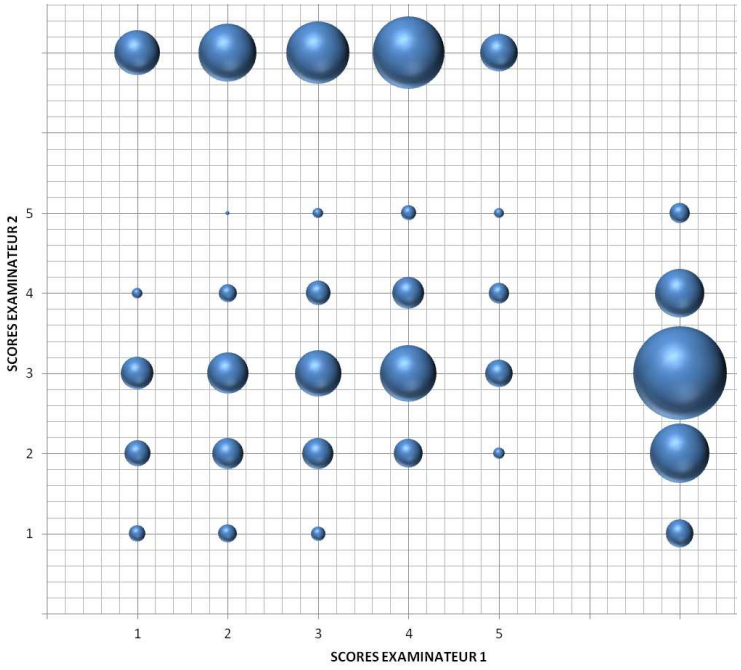


FIGURE 6.1. – Comparaison des scores GOALS moyens G entre les deux évaluateurs. La taille des disques est proportionnelle à la fréquence de survenue de ce couple de score. Il existe un couple par composante du score GOALS évaluée par les chirurgiens (l'autonomie n'apparaît pas), pour chaque répétition de chaque exercice, pour chaque volontaire. La ligne du haut et la colonne de droite de disques supplémentaires représentent la distribution des scores affectés par chaque évaluateur.

observateurs est bonne. J'utilise dans la suite de l'analyse pour chaque composante du score GOALS la moyenne des scores affectés par les évaluateurs.

6.2 Étude de la validité du score GOALS pour le banc d'essai MISTELS

Comme expliqué dans la section 5.2 du chapitre précédent, il n'existe pas d'étude ayant permis de valider l'utilisation du score GOALS pour évaluer la performance des volontaires lors de la réalisation des exercices du banc d'entraînement MISTELS.

6.2.1 Étude de la validité concurrente : corrélation entre les scores MISTELS et GOALS

i) Outils statistiques utilisés pour la corrélation

Afin de mettre en évidence une corrélation entre le score MISTELS et le score GOALS, qui sont deux variables unidimensionnelles quantitatives, je calcule les coefficients de corrélation de Pearson et de Spearman.

- le coefficient de Pearson permet de mesurer la relation linéaire existant entre deux variables, en étudiant le rapport de la covariance de ces variables sur le produit de leurs écarts-type. Il est compris entre -1 (variables parfaitement négativement corrélées) et 1 (variables parfaitement positivement corrélées). Un score de Pearson nul correspond à deux variables statistiquement indépendantes.
- le coefficient de Spearman recherche une corrélation linéaire entre les rangs des variables dans leurs distributions respectives. C'est donc une mesure non paramétrique de la dépendance statistique entre ces variables. Cette méthode de mesure de la corrélation de rang est utile lorsqu'on ne connaît pas a priori la loi permettant de relier les variables (non nécessairement linéaire) ; elle permet de mettre en évidence des corrélations monotones. Elle est donc adaptée pour comparer des scores visant à établir un classement.

ii) Moyenne des répétitions et adaptation du score MISTELS

Les scores obtenus pour la répétition d'un exercice par un volontaire ne sont pas indépendants. Afin de me restreindre à des observations indépendantes, en vue de l'interprétation de la corrélation entre les scores MISTELS et GOALS, je moyenne les scores obtenus pour les deux répétitions de chaque exercice.
On a donc dans la suite de l'analyse un score MISTELS et un score GOALS par exercice et par volontaire. Cette réduction du nombre d'observations sera également utile au moment de la classification des volontaires en différents groupes.

D'autre part, le score MISTELS étant développé avec l'objectif d'être un score de validation (avec une note discriminante), il ne s'étend pas vers des valeurs négatives (sans intérêt si la valeur discriminante est supérieure à zéro). Le recrutement d'un panel de volontaires aux niveaux d'expériences variés conduit, pour certains exercices (en particulier l'exercice 4), à l'obtention de nombreux scores ayant une valeur nulle. Ce seuillage risque d'être gênant pour la détermination de groupes, ce qui me conduit à étudier pour débuter mon analyse un score MISTELS avec ce seuillage, et un score MISTELS sans seuillage (donc autorisant des valeurs négatives). Les coefficients de corrélation décrits ci-dessus, entre le score MISTELS d'une part et les niveaux auxquels s'auto-évaluent les chirurgiens d'autre part (compréhension et expertise de la procédure chirurgicale laparoscopique), ne sont finalement pas significativement différents en fonction de l'utilisation ou non du seuillage (tableau 6.2).

	Score MISTELS seuillé à 0		Score MISTELS sans seuil	
	Compréhension	Expertise	Compréhension	Expertise
Pearson	0.54	0.62	0.44	0.48
Spearman	0.54	0.62	0.51	0.54
	IC95% :[0.27 ;0.71]	IC95% :[0.34 ;0.77]	IC95% :[0.22 ;0.72]	IC95% :[0.27 ;0.72]

Tab. 6.2. – Coefficients de corrélation entre le score MISTELS, avec et sans seuillage par la valeur 0, et l'auto-évaluation des chirurgiens.

iii) Corrélation entre le score MISTELS et le score GOALS

Afin de montrer une validité de concurrence pour l'évaluation de la qualité de la réalisation des exercices du banc d'entraînement MISTELS à l'aide du score GOALS, je compare les résultats obtenus avec ce score et ceux obtenus avec le score MISTELS. J'effectue cette analyse sur le score moyen des exercices, ainsi que pour chaque exercice. Les coefficients de corrélation sont regroupés dans le tableau 6.3.

La figure 6.2 illustre la relation linéaire qui existe entre ces deux scores, pour la moyenne des exercices. Les graphiques pour les exercices pris séparément sont semblables à celui-ci.

On note donc à partir des coefficients de corrélation que l'on peut accepter l'hypothèse d'une relation linéaire entre le score MISTELS et le score GOALS, à la fois pour le score moyen et pour les scores spécifiques des exercices.

La force de l'association est variable d'un exercice à l'autre, elle est particulièrement faible pour l'exercice 4 qui consistait en une suture extra-corporelle, technique jamais utilisée par les chirurgiens de mon échantillon et que certains découvraient. Les scores MISTELS étaient donc particulièrement bas pour cet exercice, tandis

	Moyenne	Ex. 1	Ex. 2	Ex. 3	Ex. 4	Ex. 5
Coefficient de Pearson avec MISTELS non seuillé	0.79	0.79	0.8	0.67	0.69	0.87
Coefficient de Pearson avec MISTELS seuillé	0.85	0.79	0.73	0.77	0.66	0.88
Coefficient de Spearman avec MISTELS non seuillé	0.78	0.85	0.78	0.79	0.77	0.88
Coefficient de Spearman avec MISTELS seuillé	0.84	0.77	0.77	0.79	0.79	0.86

Tab. 6.3. – Coefficients de corrélation entre les scores MISTELS, avec et sans seuillage, et le score GOALS obtenu par chaque volontaire, en moyenne et pour chaque exercice.

que le score GOALS, ne s'appuyant pas uniquement sur le résultat mais évaluant des aptitudes, n'était pas forcément abaissé (hormis la composante autonomie).

À l'inverse, la force de l'association est élevée pour l'exercice de suture intracorporelle, qui est une technique couramment utilisée mais techniquement difficile, et qui permet de mettre en évidence les aptitudes techniques du chirurgien expérimenté.

J'ai donc montré ici la validité concurrente du score GOALS pour évaluer les exercices du banc d'entraînement MISTELS. L'intérêt du premier par rapport au second est de proposer une évaluation qualitative du geste en analysant les aptitudes du chirurgien selon plusieurs composantes.

6.2.2 Étude des composantes du score GOALS expliquant le score MISTELS

Je cherche dans cette section à étudier si certaines composantes (ou aptitudes) du score GOALS sont plus significativement liées que les autres au score MISTELS, ou si toutes ces aptitudes sont liées de façon équivalente au score. En formulant le problème à l'envers, il s'agit d'étudier si certaines aptitudes du score GOALS ne sont pas explorées par le score MISTELS.

Je considère que les aptitudes du score GOALS ne sont pas colinéaires entre elles (elles ne mesurent pas la même qualité), et recherche les aptitudes les plus corrélées au score MISTELS à l'aide d'une régression linéaire multiple sans, puis avec, sélection de variables. Le critère pour intégrer une variable au modèle avec sélection de variable est le critère d'information d'Akaike (AIC), qui maximise la vraisemblance du modèle testé avec les données [Burnham et Anderson, 2004]. Je n'étudie dans cette partie que les quatre aptitudes du score GOALS évaluées en aveugle par les chirurgiens experts, l'autonomie étant évaluée par l'observateur délivrant également

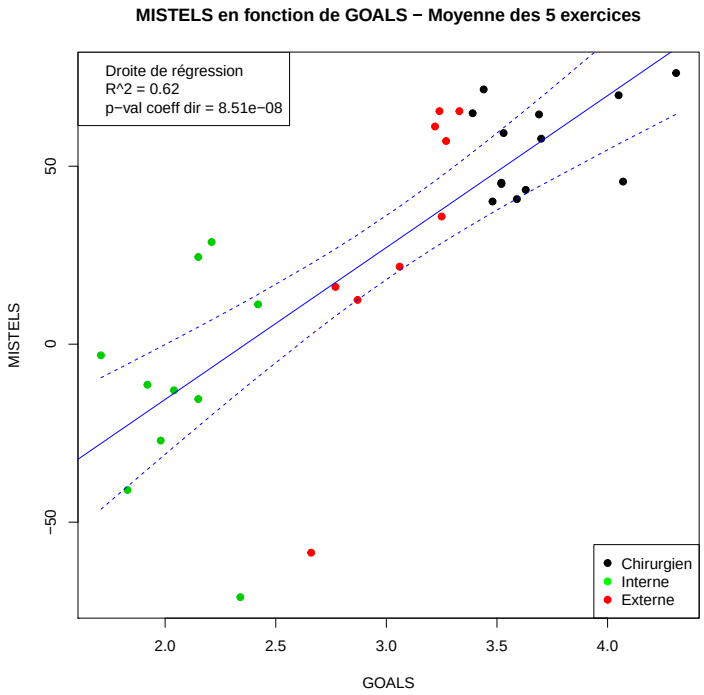


FIGURE 6.2. – Droite de régression linéaire entre le score GOALS et le score MISTELS. Les droites en pointillé représentent l'intervalle de confiance à 95%.

le score MISTELS, ce qui peut être une source de confusion.

La régression linéaire multiple entre le score MISTELS M et les scores $G^{(i)}$ ($M \sim G^{(1)} + G^{(2)} + G^{(3)} + G^{(4)}$) montre que les aptitudes *efficience* et *préhension* sont les plus significativement ($p - value < 0.05$) liées au score MISTELS pour la moyenne des exercices, avec une $p - value$ de 0.0001 pour l'*efficience* et 0.03 pour la *préhension*. Seule l'efficience est significative en absence de sélection de variable ($p - value = 0.005$).

Les résultats des régressions linéaires multiples pour chaque exercice ($M_j \sim G_j^{(1)} + G_j^{(2)} + G_j^{(3)} + G_j^{(4)}$) sont compilés dans le tableau 6.4.

Le score MISTELS est donc principalement expliqué par deux des aptitudes du score GOALS, la *perception de la profondeur* et l'*efficience*. Il paraît naturel de retrouver l'efficience comme déterminant principal du score MISTELS, vu la construction de ce score, basée principalement sur le temps nécessaire à la réalisation de la tâche. La perception de la profondeur semble également être une aptitude significative, ce qui est cohérent avec le reste de la littérature [Maan et al., 2012] ; on la

Exercice	Aptitude	Sans sélection de variable	Avec sélection de variable
Ex. 1	$G_1^{(1)}$	0.0997	$5.25 \cdot 10^{-7}$
Ex. 2	$G_2^{(1)}$	0.039948	$8.01 \cdot 10^{-5}$
Ex. 3	$G_3^{(3)}$	0.311	0.00037
Ex. 4	$G_4^{(3)}$	0.013178	$3.09.10^{-5}$
Ex. 5	$G_5^{(3)}$	0.00435	$9.3.10^{-9}$

Tab. 6.4. – Aptitudes du score GOALS les plus significatives pour expliquer le score MISTELS en fonction des exercices réalisés. Les valeurs données dans les colonnes de droite sont les $p - values$ des aptitudes dans les régressions linéaires multiples sans, puis avec, sélection de variable. G^1 →*perception de la profondeur* ; G^2 →*dextérité bi-manuelle* ; G^3 →*efficience* ; G^4 →*préhension* ; G^5 →*autonomie*.

retrouve dans deux exercices faisant appel au placement des instruments par rapport à un plan (compresse à découper) ou à un objet de référence (plots). L'absence de significativité de cette aptitude pour expliquer le score MISTELS des autres exercices, du 4 et du 5 en particulier (qui font théoriquement également appel à la perception de la profondeur) peut sans doute être expliquée par le fait que le fond utilisé pour ces exercices, de couleur rouge vif, perturbait fortement les chirurgiens. Ces derniers m'ont fait remarquer à plusieurs reprises que leur perception de la profondeur était modifiée dans ce contexte, les renvoyant au niveau des novices pour cette aptitude.

On note enfin que la dextérité bi-manuelle et la préhension n'apparaissent pas dans les résultats significatifs de cette analyse, contrairement à ce qu'on attend à partir de la description des exercices MISTELS (section 5.1).

Les aptitudes supplémentaires explorées par le score GOALS, qui ne sont pas significativement liées au score MISTELS, apportent cependant probablement une information supplémentaire quant à la qualité du geste réalisé par rapport à ce dernier. Pour mettre en évidence cette information supplémentaire, je compare la performance de ces deux scores pour classer les volontaires dans des groupes de niveau, c'est-à-dire que j'évalue les validités prédictives de chacune de ces méthodes.

6.2.3 Étude de la validité prédictive : classification des volontaires en fonction de leur score GOALS

J'étudie et je compare ici la validité prédictive de chacun des scores, c'est-à-dire leur capacité à classer correctement les volontaires en fonction de leur niveau attendu.

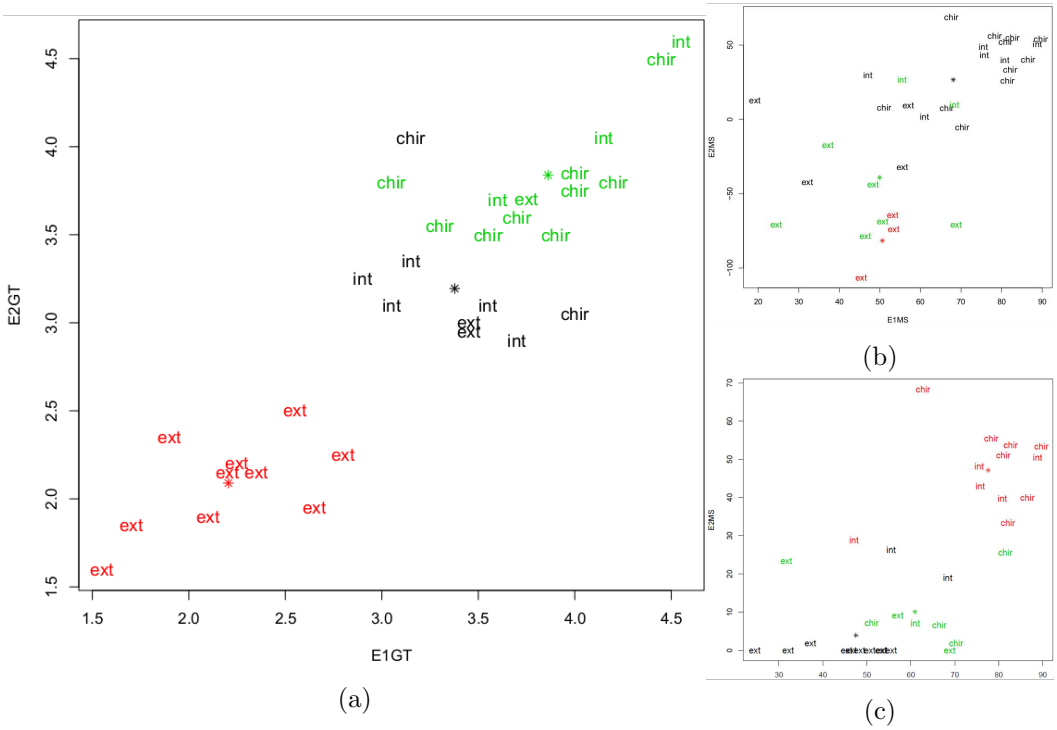


FIGURE 6.3. – Classification des volontaires en trois groupes de niveau en fonction de leur score GOALS moyen pour chaque exercice G_j (a), de leur score MISTELS sans seuillage (b) et de leur score MISTELS avec seuillage (c). Les couleurs correspondent aux groupes déterminés à partir du classifieur, les labels à côté de chaque point (*chir, int, ext*) aux groupes de référence.

▷ **Outils statistiques utilisés** J'utilise pour atteindre cet objectif un classifieur des k-moyennes (*k-means*), qui réalise de manière non-supervisée une partition de mon échantillon de volontaires en trois groupes (ou *clusters*) de niveau (appelés *experts*, *intermédiaires*, *novices*), de façon à minimiser la distance entre les points à l'intérieur de chaque groupe. Je réalise cette partition dans un espace à cinq dimensions, chaque dimension représentant un exercice. Chaque volontaire est donc affecté à l'un des groupes en fonction de son vecteur de scores $[G_1\ G_2\ G_3\ G_4\ G_5]$ pour la classification basée sur le score GOALS, $[M_1\ M_2\ M_3\ M_4\ M_5]$ pour la classification basée sur le score MISTELS. Le nombre d'itérations pour choisir les graines de la partition est choisi suffisamment élevé pour avoir une probabilité forte de convergence vers un minimum global. Le choix de trois groupes de niveau est effectué pour permettre la comparaison avec les groupes de niveau théoriques liés au statut du volontaire. Les groupes ainsi formés par la méthode des *k-means*, pour chacun des scores utilisés, sont illustrés figure 6.3.

J'utilise, pour étudier la pertinence des classements, des matrices de confusion (illustrées tableau 6.5) me permettant de calculer un taux de reconnaissance global avec un classement de référence ($Acc = \frac{nombre\ de\ volontaires\ bien\ classés}{nombre\ de\ volontaires\ total} = \frac{\sum ßC_{ii}}{\sum_i \sum_j C_{ij}}$ avec les notations décrites dans le tableau 6.5).

Les groupes de référence pour cette première partie de l'étude sont les groupes formés en fonction des statuts des volontaires (externe, interne ou chirurgien).

		Référence			
		Chirurgiens	Internes	Externes	Erreur d'excédent
Classifieur	Expert	9	3	1	0.31
	Intermédiaire	2	5	2	0.45
	Novice	0	0	10	0
	Erreur de déficit	0.18	0.27	0.23	$Acc = 0.75$

Tab. 6.5. – Exemple de matrice de confusion : matrice de confusion utilisée pour comparer le classement en groupe obtenu avec le classifieur *k-means* sur le score GOALS avec les groupes basés sur le statut des volontaires. Chaque élément C_{ij} de la matrice correspond au nombre de volontaires classés dans le groupe i appartenant au groupe j. Les erreurs de déficit correspondent au pourcentage de volontaires d'un groupe de référence affectées à d'autres groupes par la classification : $Ea_j = \frac{\sum_{i \neq j} C_{ij}}{\sum_i C_{ij}}$. Les erreurs d'excédent correspondent au pourcentage de volontaires d'un groupe issu de la classification qui appartiennent à un autre groupe de référence : $Ee_i = \frac{\sum_{j \neq i} C_{ij}}{\sum_j C_{ij}}$. Le taux de reconnaissance global s'écrit $Acc = \frac{\sum_i C_{ii}}{\sum_i \sum_j C_{ij}}$

Cette approche classique permet d'obtenir les taux de reconnaissance présentés dans le tableau 6.6. On remarque que la classification basée sur le score GOALS a un taux de reconnaissance nettement meilleur que le score MISTELS. Cette comparaison est cependant à manipuler avec précaution, car le classement basé sur les statuts des volontaires est très réducteur dans les exercices de simulation, d'autres paramètres pouvant rentrer en jeu (entraînement préalable, habitude de jeux vidéos, perception spatiale visuelle,...) [Maan et al., 2012]. On le voit ici avec des internes, entraînés sur d'autres types de banc d'entraînement, qui réalisent des meilleurs scores (à la fois MISTELS et GOALS) que certains chirurgiens.

J'ai donc démontré dans cette section que le score GOALS était adapté pour l'évaluation de la qualité de réalisation d'exercices sur le banc d'entraînement MISTELS (en particulier la validité concurrente et la validité prédictive ont été vérifiées) ; les résultats de la classification par *k-means* suggèrent également que le score GOALS classe mieux les volontaires que le score MISTELS, car il apporte des informations supplémentaires qui n'apparaissent pas le score MISTELS (en particulier les aptitudes $G^{(2)}$ et $G^{(4)}$, la dextérité bi-manuelle et la préhension). De plus, ce score a

	Acc	Variance de *Acc*
GOALS	0.750	0.006
MISTELS seuillé	0.562	0.008
MISTELS non seuillé	0.500	0.008

Tab. 6.6. – Taux de reconnaissance entre les groupes basés sur les différents scores et les groupes basés sur le statut des chirurgiens.

l'intérêt d'apporter une interprétation de la qualité du geste en fonction des compétences requises pour le réaliser correctement de façon explicite, interprétation que je cherche à retrouver à partir de l'analyse des métriques définies au chapitre précédent.

Dans la suite de ce travail, je considère donc que le score de référence que je cherche à prédire, en construisant un modèle détaillé dans la section suivante, est le score GOALS ; la nouvelle classification de référence des volontaires ayant participé à l'étude est celle basée sur la partition de ces volontaires en groupes grâce à la méthode des *k-means* basée sur le score GOALS (groupes de la figure 6.3a).

6.3 Prédiction d'un score de qualité du geste chirurgical à partir des métriques acquises

Je cherche dans cette section à identifier, parmi l'ensemble des métriques définies au chapitre précédent, celles qui sont prédictives du score GOALS global G d'un volontaire, et à définir un score prédisant le score GOALS évalué par des examinateurs externes. La prédiction obtenue à partir de ces métriques peut être considérée comme une notation similaire à celle attribuée par le score MISTELS, c'est-à-dire une évaluation du niveau global de l'opérateur, sans apport explicatif. Cette étape est le premier pas vers une notation plus explicative du geste, basée sur la prédiction de chaque aptitude du score GOALS en fonction des métriques enregistrées, que j'étudierai plus loin.

6.3.1 Outils statistiques utilisés

J'utilise pour construire ces modèles prédictifs une régression PLS (*partial least squares*) [Haenlein et Kaplan, 2004, Mevik et Wehrens, 2007] méthode fréquemment utilisée en chimie moléculaire pour prédire les propriétés d'une molécule à partir de sa structure (domaine des relations quantifiées structure-propriété).

L'analogie avec la situation étudiée dans le cadre de ce travail est directe : je connais la structure du geste chirurgical, décrite par l'ensemble des métriques enregistrées, et je cherche à dériver les propriétés de ce geste, c'est-à-dire sa qualité

selon le score GOALS. L'approche basée sur la régression PLS apparaît donc comme particulièrement adaptée à mon problème.

La régression PLS peut être vue, de façon sommaire, comme une généralisation de l'analyse en composantes principales (ACP) conjuguée avec une régression linéaire multiple (approche dite régression en composantes principales, RCP). Elle bénéficie des avantages de la RCP, c'est à dire qu'elle permet de traiter des observations de grande dimension dont certaines peuvent être colinéaires (grâce à l'orthogonalisation des prédicteurs dans un sous-espace de l'espace des observations). Une limite de la RCP est la stratégie à adopter pour choisir les vecteurs du sous-espace des observations qui seront les meilleurs prédicteurs de la variable à prédire. Dans l'approche des régressions linéaires multiples, la principale limite est la nécessité d'obtenir un nombre d'observations largement supérieur au nombre de dimensions de ces dernières (afin de pouvoir inverser la matrice de corrélation des observations), tandis que la régression PLS supporte un nombre d'observations proche voire inférieur au nombre de dimensions de celles-ci. Ainsi, je peux étudier dans ce travail 30 métriques, dont certaines sont probablement colinéaires entre elles, avec seulement 32 volontaires, ce qui ne serait pas permis avec une approche de type RCP standard.

Pratiquement, une régression PLS consiste en la recherche de variables dites latentes (ou composantes), dérivées des nombreuses variables explicatives des observations, dans un espace de dimensionalité plus faible que celui des variables explicatives originales. Ce point rapproche la méthode PLS d'une analyse en composantes principales, avec la création de composantes orthogonales entre elles (donc non corrélées), qui sont des combinaisons linéaires des variables explicatives intiales.

La différence avec une approche RCP standard se situe dans le choix des variables latentes, qui sont ici choisies de façon itérative, en optimisant à chaque itération l'information supplémentaire apportée au modèle pour prédire la variable à expliquer, tout en préservant la contrainte d'orthogonalité avec les composantes déjà construites. L'optimisation est réalisée en maximisant la corrélation de la nouvelle composante avec la variable à expliquer.

Le modèle est donc construit de manière itérative, en utilisant un nombre de composantes variable. Le nombre de composantes optimal pour le modèle est défini par une approche de validation croisée, avec un ratio cible à atteindre entre l'erreur résiduelle du modèle à n composantes $RESS_n$ et l'erreur prédite du modèle à $n+1$ composantes $PRESS_{n+1}$. Plus précisément, on n'ajoute pas la $(n+1)^{\text{ième}}$ composante si $\sqrt{PRESS_{n+1}} > 0.95 \cdot \sqrt{RESS_n}$. Les erreurs résiduelles correspondent à des erreurs quadratiques moyennes entre la prédiction et la variable prédite. Ce choix du nombre de composantes est crucial, un trop grand nombre de composantes conduisant à un sur-ajustement du modèle. La méthode d'optimisation du nombre de composantes est déjà implémentée dans la fonction `plsr` de R.

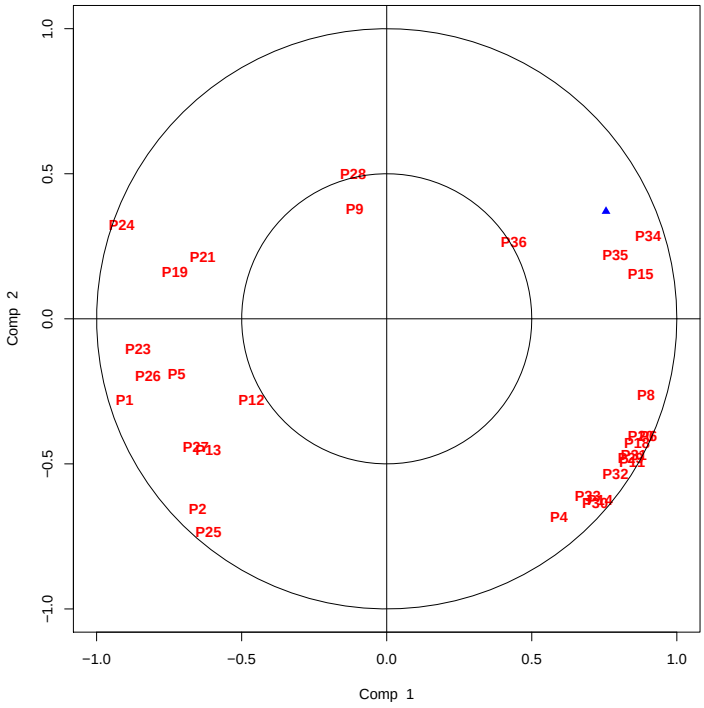
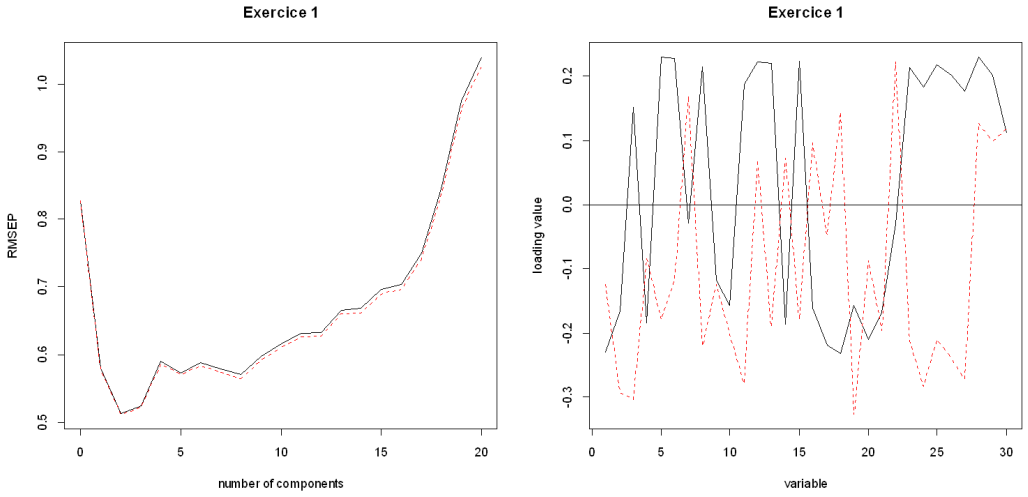


FIGURE 6.4. – Projection dans le plan des composantes des observations et de la variable à expliquer (triangle bleu), pour une situation avec deux composantes. Le coefficient de régression entre la projection de l'observation et la projection de la variable à prédire est égal au produit scalaire des vecteurs reliant l'origine à ces projections. Les métriques positivement corrélées au score GOALS se situent dans le même cadran, les observations négativement corrélées dans le cadran opposé.

Géométriquement, la régression PLS consiste donc à projeter les observations X dans un sous-espace de l'espace des X (réduction de la dimensionalité des observations), afin que ces projections soient les meilleurs prédicteurs de la projection de la variable à prédire Y sur ce sous-espace (principe illustré par la figure 6.4, pour un nombre de composantes optimal égal à 2).

Pour construire les modèles prédictifs à partir de N observations, j'adopte une approche de validation croisée de type *leave one out* qui permet de construire un modèle à partir de N-1 observations et de prédire le score de la N$^{\text{ième}}$ observation, ceci de façon itérative sur les N observations.

Avant d'être utilisée dans le modèle PLS, chaque métrique X est centrée et nor-



(a) (b)

FIGURE 6.5. – Construction de la régression PLS pour les données de l'exercice 1. (a) : choix du nombre de composantes intégrées à la régression, qui correspond au premier minima local de l'erreur de prédiction ; (b) : poids des différentes variables dans les deux premières composantes.

malisée ($X_i = \frac{x_i - < x >}{\sigma_x}$), où $< x >$ est la moyenne des x_i des volontaires et σ_i leur écart-type.

La méthode PLS permet également de hiérarchiser les métriques selon leur pouvoir explicatif dans la prédiction, grâce à la variable VIP (*Variable Importance in Prediction*), qui indique la pertinence de chaque métrique pour la prédiction finale du score GOALS avec le nombre de composantes retenues. On considère qu'une métrique ayant une valeur de VIP supérieure à 1 sera fortement significative. Cependant, une variable dont la VIP est inférieure à 1 ne sera pas forcément inopérante dans la classification.

J'évalue l'erreur de prédiction du modèle en étudiant le coefficient de corrélation entre le score GOALS réel et le score GOALS prédit.

6.3.2 Plan d'analyse

Pour chaque exercice j, je détermine un modèle prédictif du score GOALS G_j à l'aide de la régression PLS en optimisant le nombre de composantes. Les figures 6.4 et 6.5 résument les différentes étapes de la méthode.

Cette analyse est conduite sur chaque exercice séparément.

6.3.3 Métriques significatives dans le modèle prédictif pour chaque exercice

J'étudie tout d'abord quelles sont les métriques les plus significatives (VIP≥1) dans la régression PLS de chaque exercice. Le tableau 6.7 présente quelles sont les métriques significatives ainsi que le signe de leur corrélation avec le score GOALS de référence. Je présente dans l'annexe A l'ensemble des VIP des métriques pour chaque exercice.

L'analyse de ce tableau permet de noter :
- que le temps, la longueur de chemin parcouru, le volume exploré et l'économie de volume (rapport du volume exploré sur la longueur de chemin parcouru) sont négativement corrélés au score, quel que soit l'exercice réalisé : les trois premières corrélations apparaissent évidentes, la dernière témoigne que pour réaliser un trajet d'une longueur donnée, le chirurgien expert minimisera le volume exploré ;
- que l'intégrale du *jerk* normalisé par la distance parcourue est systématiquement positivement corrélée au score obtenu : cette corrélation est contre-intuitive, et est probablement liée à un défaut de normalisation de la métrique (les variations du *jerk* sont trop faibles par rapport au dénominateur, qui devient le déterminant principal de la métrique) ;
- que la métrique rugosité en revanche, qui normalise le *jerk* différemment, est bien négativement corrélée au score lorsqu'elle est significative, de même que le nombre de passages par zéro du *jerk* par unité de temps et de longueur (mesure du nombre de pics d'accélération) ;
- que les métriques liées à l'information mutuelle entre les mains sont positivement corrélées au score obtenu, ce qui traduit une forme de coordination bi-manuelle chez les experts ;
- que le taux de recouvrement des volumes explorés par les instruments est négativement corrélé au score, ce qui traduit une séparabilité nette entre les volumes affectés à chaque instrument chez les experts.

Certaines autres métriques ne sont ponctuellement significatives que pour les modèles spécifiques à chaque exercice.
De plus, les métriques les plus significatives dans les modèles pour l'ensemble des exercices sont le temps, la longueur de chemin parcouru et l'économie de volume (métriques cohérentes avec le reste de la bibliographie) ainsi que le volume exploré (qui n'avait pas été précisément exploré dans les travaux étudiés). Dans les exercices faisant intervenir les deux mains du chirurgien, j'identifie deux nouvelles métriques qui sont systématiquement significatives dans les modèles de prédiction construits, le taux de recouvrement des volumes explorés par chaque main (un faible taux de recouvrement est associé à un score GOALS élevé) et l'information mutuelle entre les vitesses de chaque main (les distributions de vitesse des mains ne sont pas indépendantes l'une de l'autre).

Métriques		Ex. 1	Ex. 2	Ex. 3	Ex. 4	Ex. 5
temps	P1	-	-	-	-	-
longueur de chemin parcouru	P2	-	-	-	-	-
niveau de bruit par unité de temps	P4					
niveau de bruit par unité de longueur	P5	-				
vitesse moyenne	P6					
vitesse angulaire moyenne	P8			+		
déplacement angulaire par unité de longueur	P9					
accélération temporelle	P11					
accélération spatiale	P12		-	-		
rugosité	P13		-	-		
jerk temporel	P14					
jerk spatial	P15	+	+	+	+	+
variation de profondeur le long de l'axe moyen par unité de temps	P18					
variation de profondeur le long de l'axe moyen par unité de longueur	P19		-	-		-
variation de profondeur le long de l'axe par unité de temps	P20					
variation de profondeur le long de l'axe par unité de longueur	P21		-	-		-
nombre de passage par zéro du *jerk* par unité de temps	P23	-		-		
nombre de passage par zéro du *jerk* par unité de longueur	P24	-		-		
volume exploré	P25	-	-	-	-	-
taux de recouvrement des volumes	P26	-	-	n/a	-	-
économie de volume	P27	-	-	-	-	
R^2 des vitesses de chaque main	P28			n/a		+
distance aux axes des couples de vitesse	P29			n/a		
distance aux axes des couples d'accélération	P30			n/a		
distance aux axes des couples de *jerk*	P31			n/a		
distance à l'origine des couples d'accélération	P32			n/a		
distance à l'origine des couples de *jerk*	P33			n/a		
information mutuelle des vitesses	P34	+	+	n/a	+	+
information mutuelle des accélérations	P35	+		n/a	+	
information mutuelle des *jerks*	P36			n/a		
Nombre de composantes dans le modèle		2	2	4	2	4

Tab. 6.7. – Résumé des paramètres les plus significatifs (VIP$\geq$ 1) pour les différents exercices dans les modèles prédictifs PLS. '+' traduit une corrélation positive avec le score GOALS, '-' une corrélation négative. n/a : non applicable $\rightarrow$ l'exercice 3 ne retourne la trajectoire que d'une main.

6.3.4 Prédiction d'un score à partir de l'ensemble des métriques

J'étudie ici le résultat de la prédiction dans laquelle l'ensemble des métriques est utilisé dans la régression PLS, quelle que soit la valeur de leur VIP (30 métriques pour les exercices 1, 2, 4 et 5, 20 métriques pour l'exercice 3 après exclusion des métriques décrivant les mouvements relatifs des instruments).

i) Prédiction du score GOALS pour chaque exercice et du score GOALS moyen

Les modèles construits à l'aide de la régression PLS permettent de prédire les scores de chaque volontaire pour chaque exercice, $\tilde{G}_j$. J'obtiens donc une nouvelle distribution de scores GOALS prédits, ainsi qu'un score GOALS moyen prédit $\tilde{G} = \frac{1}{5}\sum_{j=1}^{5}\tilde{G}_j$, que je peux comparer aux scores GOALS initiaux à l'aide d'une régression linéaire (figure 6.6) ainsi qu'à l'aide des coefficients de Spearman et de Pearson, déjà décrits plus haut (tableau 6.8).

	Moyenne	Ex. 1	Ex. 2	Ex. 3	Ex. 4	Ex. 5
Pearson	0.88	0.77	0.58	0.83	0.72	0.87
Spearman	0.88	0.82	0.68	0.79	0.8	0.87
IC95%	[0.76 ;0.93]	[0.62 ;0.91]	[0.43 ;0.82]	[0.57 ;0.89]	[0.63 ;0.89]	[0.73 ;0.92]

Tab. 6.8. – Coefficients de corrélation entre les socres GOALS prédits et réels.

On note que le score GOALS moyen prédit, $\tilde{G} = \frac{1}{5}\sum_{j=1}^{5}\tilde{G}_j$, est fortement corrélé au score moyen GOALS réel, G. L'analyse séparée de chaque exercice (figure 6.7 et tableau 6.8) montre que la prédiction du score GOALS pour un exercice est en général fortement corrélée au score GOALS réel de cet exercice, sauf pour l'exercice 2 (découpe de patron) pour lequel les scores GOALS prédits sont très dispersés, et ne semblent pas corrélés au score GOALS réel. Cette dispersion des scores est visible en particulier pour les volontaires des groupes *novice* et *intermédiaire*.
Le coefficient de corrélation de l'exercice 4, légèrement inférieur à ceux des autres exercices, peut s'expliquer par les temps de réalisation particulièrement longs de cet exercice pour certains externes (classés ici en *novice* et *intermédiaire*), qui ne sont que peu pris en compte lors de la construction du modèle grâce à l'approche de type *leave one out*.

Il est intéressant d'étudier la répartition des erreurs de la prédiction en fonction des groupes auxquels appartiennent les volontaires, afin de définir plus précisément la population à laquelle cette méthode peut être appliquée avec confiance. Je compare donc les erreurs quadratiques moyennes entre les scores GOALS moyen prédits $\tilde{G}$ et réels G, ainsi qu'entre les scores GOALS prédits et réels par exercice ($\tilde{G}_j$ et G_j), en fonction des groupes basés sur le statut des volontaires et des groupes basés

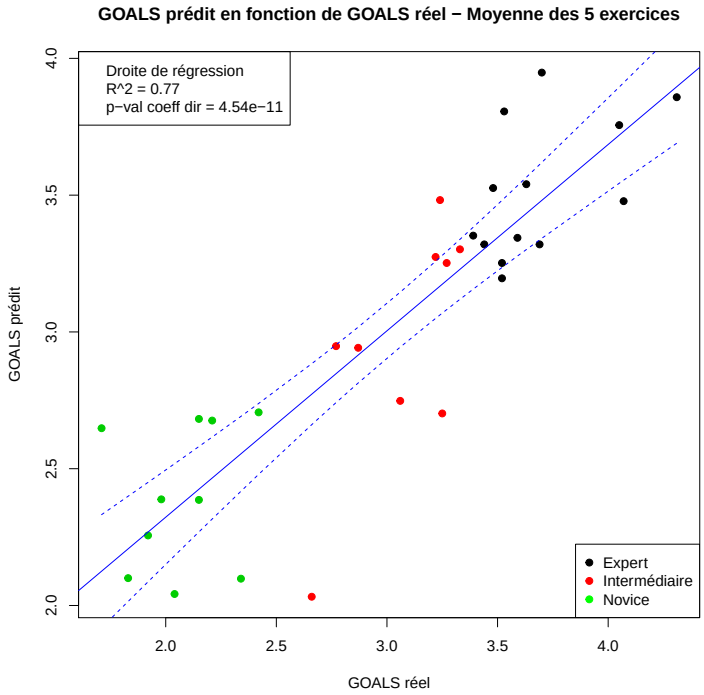
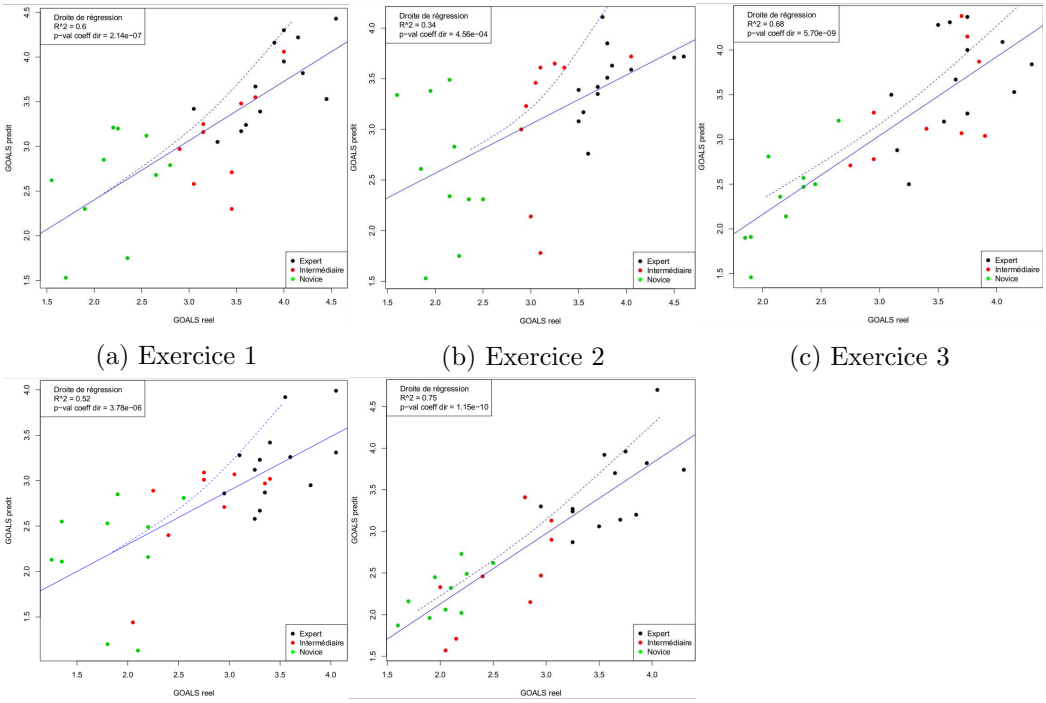


FIGURE 6.6. – Régression linéaire entre le score GOALS moyen réel G et le score GOALS moyen prédit $\tilde{G}$ par le modèle obtenu par régression PLS. Le score situé très loin de la courbe de régression correspond au point d'une externe, classée intermédiaire avec la classification basée sur le score GOALS réel, dont le score GOALS prédit est fortement abaissé à cause d'un exercice (suture extra-corporelle) pour lequel le temps de réalisation a été particulièrement long.



(a) Exercice 1 (b) Exercice 2 (c) Exercice 3

(d) Exercice 4 (e) Exercice 5

FIGURE 6.7. – Score GOALS prédit $\tilde{G}_j$ en fonction du score GOALS réel G_j pour chaque exercice. Les volontaires du groupe *novice* apparaissent en vert, ceux du groupe *intermédiaire* en rouge et ceux du groupe *expert* en noir.

sur la classification obtenue avec le score GOALS réel à a section précédente. Ces erreurs sont présentées dans le tableau 6.9.

		Ex. 1	Ex. 2	Ex. 3	Ex. 4	Ex. 5	Moyenne
Erreur globale		0.090	0.112	0.080	0.106	0.060	0.090
Groupe Statut	Externe	0.189	0.217	0.115	0.205	0.089	0.163
	Interne	0.072	0.202	0.208	0.135	0.122	0.148
	Chirurgien	0.126	0.139	0.111	0.163	0.109	0.130
Groupe GOALS	Novice	0.198	0.248	0.129	0.236	0.099	0.182
	Intermédiaire	0.166	0.184	0.198	0.191	0.120	0.172
	Expert	0.113	0.155	0.0994	0.133	0.096	0.119

Tab. 6.9. – Erreurs quadratiques moyennes des prédictions en fonction des groupes de volontaires.

L'erreur quadratique moyenne entre la prédiction et le score GOALS réel est globalement plus faible pour les groupes de niveau élevé (*chirurgien* ou *expert* selon la classification utilisée). Cet effet est moins marqué pour les exercices qui ont déjà une corrélation élevée entre le score prédit et le score réel d'après le tableau 6.8 (exercices 1,3 et 5). On note que l'erreur est systématiquement plus faible pour le groupe des *experts*, ce qui s'explique probablement par une moins grande variabilité entre les valeurs des métriques d'un volontaire de ce groupe à l'autre : il existe moins de valeurs aberrantes des métriques pour les experts (valeur de la métrique supérieure à 2 après la normalisation présentée dans la section 6.3.1).

ii) Validité prédictive du score obtenu avec la régression PLS

De la même manière qu'avec le score GOALS réel, je cherche à évaluer la validité prédictive du score GOALS prédit par ma méthode d'acquisition des trajectoires. J'utilise pour cela à nouveau un classifieur *k-means*, sur les scores de chaque exercice $\tilde{G}_j$.
Afin de comparer la classification obtenue à la classification obtenue avec le score GOALS réel, qui sert de référence, je calcule le taux de reconnaissance, en introduisant une pondération en fonction de la "gravité" de l'erreur de classification [Cohen, 1968]. Cette pondération permet de calculer un taux de reconnaissance pondéré Acc_w.

En reprenant les notations précédentes de la matrice de confusion C, dont les éléments C_{ij} sont le nombre de volontaires du groupe de référence j affectés au groupe i par le classifieur, on calcule la matrice de proportion P dont les éléments sont $P_{ij} = \frac{C_{ij}}{\sum_i \sum_j C_{ij}}$.

J'introduis la matrice de pondération W :

$$W = \begin{bmatrix} 1 & 0.1 & 0 \\ 0.1 & 1 & 0.1 \\ 0 & 0.1 & 1 \end{bmatrix}$$

Les coefficients de cette matrice traduisent le fait que l'on accorde davantage d'importance aux erreurs de classifications qui correspondent à un décalage de deux groupes de niveau qu'à celles qui ne correspondent qu'à un décalage d'un groupe. On a alors $Acc_w = \frac{\sum_i \sum_j W_{ij} \cdot C_{ij}}{\sum_i \sum_j C_{ij}}$.

▷ **Classification en fonction des scores prédits** Les résultats des classements à l'aide de la méthode *k-means* des volontaires basés sur le score GOALS prédit est illustrée sur la figure 6.8.

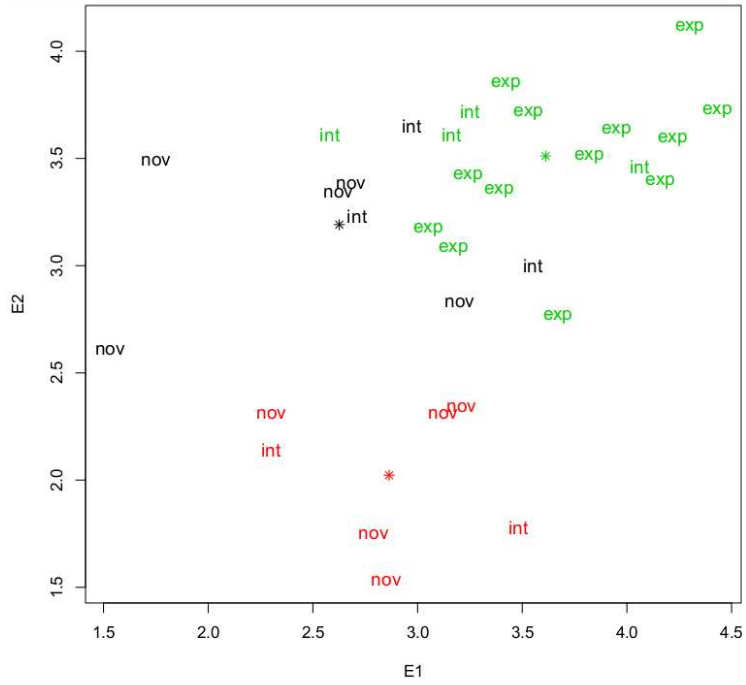


FIGURE 6.8. – Classification des volontaires par un classifieur *k-means* basé sur le score GOALS prédit dans les cinq exercices. Les volontaires sont projetés ici dans le plan formé par les scores obtenus aux premier et second exercice, à partir de l'espace à cinq dimensions représentant les scores de l'ensemble des exercices. Les couleurs correspondent aux groupes déterminés à partir du classifieur, les labels à côté de chaque point (*exp, int, nov*) aux groupes de référence.

La matrice de confusion est explicitée dans le tableau 6.10.

		Référence		
		Expert	Intermédiaire	Novice
k-means	Expert	13	4	1
	Intermédiaire	0	3	5
	Novice	0	2	5

Tab. 6.10. – Matrice de confusion obtenue avec la classification *k-means* effectuée sur le score GOALS prédit.

Le taux de reconnaissance pondéré est de 69,1% (65,6% sans pondération, le taux de reconnaissance pondéré avec le statut original du volontaire est de 63.1%).

D'une façon générale on note que le groupe expert est trop peuplé par rapport au groupe original. En analysant la classification *k-means* obtenue séparément sur chaque exercice, en prenant toujours les groupes obtenus avec le GOALS réel moyen comme références (figure 6.9, avec le taux de reconnaissance pondérée pour chaque exercice), on note que la classification échoue principalement pour les exercices *découpe de patron* et *suture extra-corporelle*, qui sont les exercices où les métriques aberrantes sont les plus marquées (en particulier le temps). Ces métriques aberrantes isolent une sous-population du groupe novice dans un groupe spécifique très éloigné des autres, ce qui conduit le classifieur *k-means* à créer seulement deux groupes à partir de la population restante.

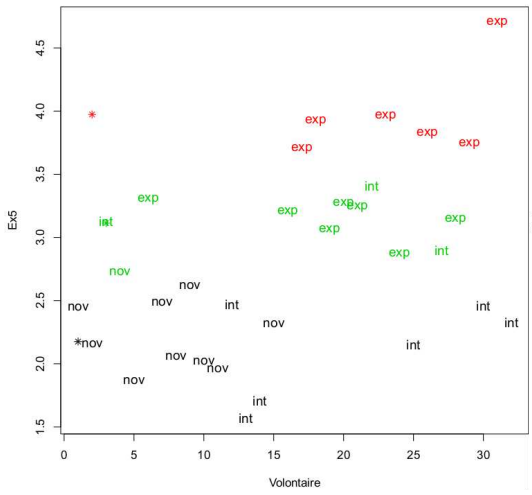
6.3.5 Prédiction d'un score à partir des métriques significatives de chaque exercice

Cette étude vise à étudier s'il est possible de réduire le nombre de métriques mesurées en conservant une prédiction fiable du score GOALS. Je n'effectue donc les régressions PLS qu'avec le sous-ensemble de métriques qui ont une VIP$\geq$1 lors de la régression PLS avec l'ensemble des métriques (les cellules non vides du tableau 6.7), en optimisant le nombre de composantes pour ce cas particulier.
Le principe de l'analyse est identique à la partie précédente. J'étudie la corrélation entre le score GOALS prédit à partir du sous-ensemble de métriques significatives pour chaque exercice, $\tilde{g}_j$, et le score GOALS réel pour cet exercice, G_j, ainsi qu'entre les moyennes de ces scores (tableau 6.11 et figure 6.10).

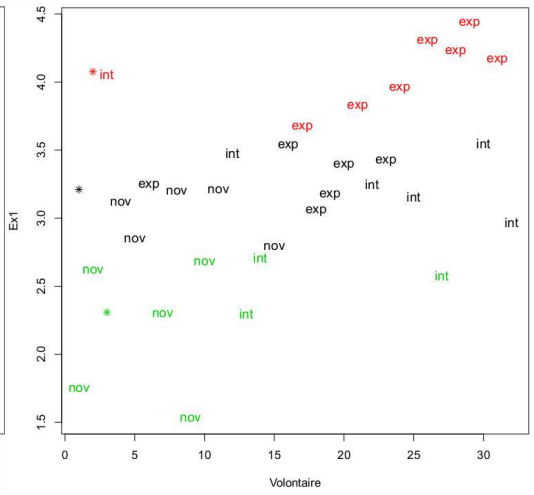
	Moyenne	Ex. 1	Ex. 2	Ex. 3	Ex. 4	Ex. 5
Pearson	0.85	0.78	0.69	0.80	0.64	0.84
Spearman	0.85	0.82	0.71	0.81	0.70	0.86
IC95%	[0.71 ;0.92]	[0.62 ;0.91]	[0.49 ;0.82]	[0.63 ;0.9]	[0.46 ;0.84]	[0.72 ;0.92]

Tab. 6.11. – Coefficients de corrélation entre les scores GOALS prédits et réels.

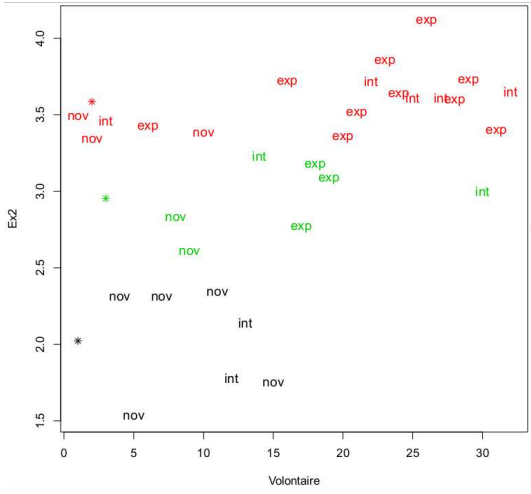
Il n'existe pas de différence notable entre le score GOALS prédit avec un sous-



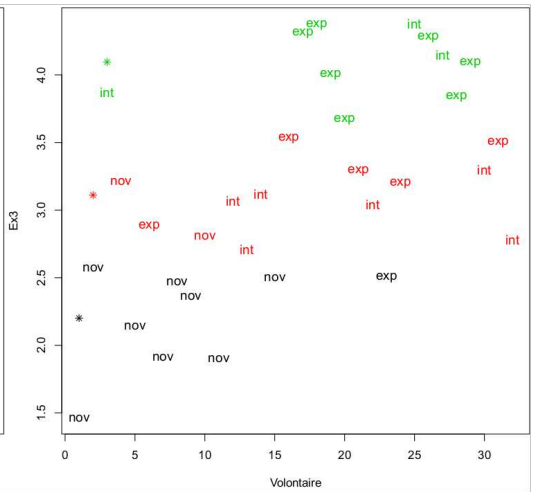
(a) Exercice 1



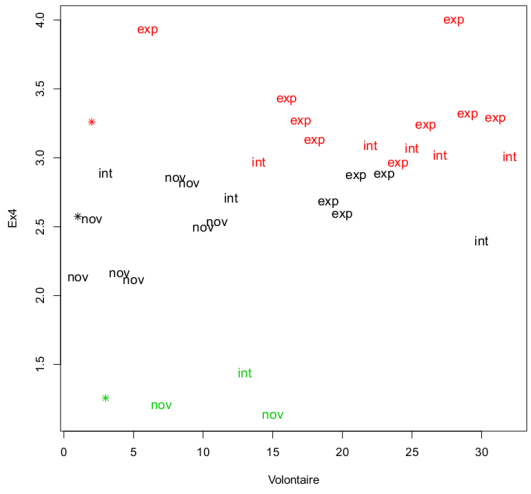
(b) Exercice 2



(c) Exercice 3



(d) Exercice 4



(e) Exercice 5

	Acc_w
Ex. 1	57.8%
Ex. 2	56.8%
Ex. 3	68.8%
Ex. 4	49.4%
Ex. 5	60.6%

(f) Taux de reconnaissance pondérée de chaque exercice

FIGURE 6.9. – Classification *k-means* pour chaque exercice basée sur le score GOALS prédit spécifique à cet exercice. Les couleurs correspondent aux groupes déterminés à partir du classifieur, les labels à côté de chaque point (*exp, int, nov*) aux groupes de référence.

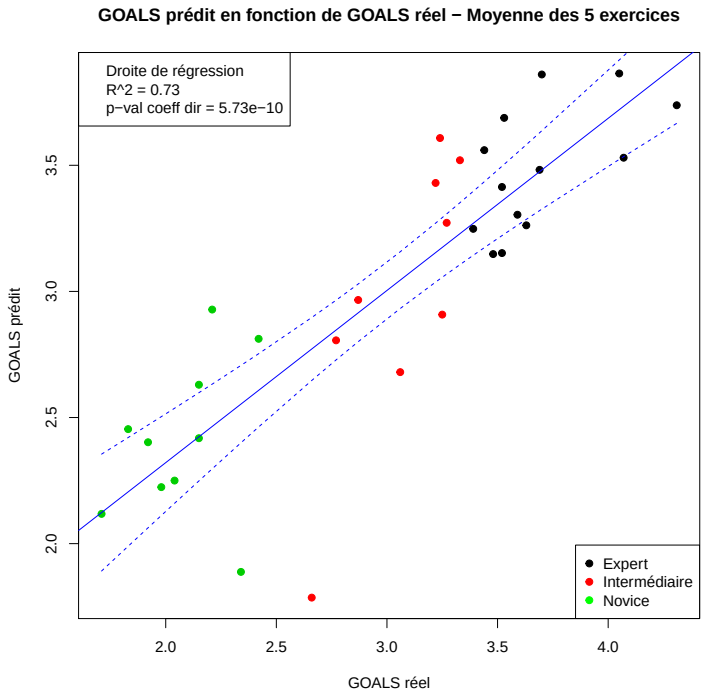


FIGURE 6.10. – Régression linéaire entre le score GOALS moyen réel G et le score GOALS moyen prédit à partir d'un sous-ensemble de métriques $\tilde{g}$.

ensemble de métriques particulièrement significatives et le score GOALS prédit à partir de l'ensemble des métriques. La classification obtenue avec ce score est également très proche de celle obtenue avec le score $\tilde{G}$, le taux de reconnaissance avec la classification basée sur le score GOALS réel est identique ($Acc_w = 69.1\%$) à celui déterminé dans la section précédente.

On peut donc supposer que l'information contenue dans les métriques dont la VIP est supérieure à 1 est suffisante pour classer les volontaires en groupes de niveau, avec un taux de reconnaissance satisfaisant.

6.4 Classification en fonction des métriques uniquement

J'étudie en parallèle la capacité des métriques choisies à classer les volontaires en groupes, sans passer par l'étape de prédiction effectuée à l'aide de la régression PLS. Je fais varier le nombre de métriques pour déterminer le sous-ensemble de métriques qui permet d'obtenir le meilleur taux de reconnaissance, en sélectionnant soit :

- toutes les métriques définies pour chaque exercice (soit 140 métriques) → sous-ensemble M
- toutes les métriques significatives pour chaque exercice (soit 51 métriques) → sous-ensemble M_S
- les métriques communes à l'ensemble des exercices (en excluant les métriques de coordination bi-manuelle pour l'exercice 3, soit 28 métriques) → sous-ensemble M_C

Pour chaque sous-ensemble, une classification par le classifieur *k-means* est réalisée, et les groupes obtenus sont comparés aux groupes obtenus avec le score GOALS réel. Les taux de reconnaissance sont regroupés dans le tableau 6.12.

	M	M_S	M_C
Acc_w	57.5%	71.9%	74.7%

Tab. 6.12. – Taux de reconnaissance pour les classifications obtenues à partir d'un sous-ensemble de métriques.

Je constate que les classifications basées sur la méthode des *k-means* après réduction du nombre de métriques présentent un taux de reconnaissance supérieur au taux de reconnaissance obtenu avec le score GOALS prédit par la régression PLS. L'information contenue dans ces métriques est donc suffisante pour classer correctement un volontaire ayant effectué les cinq exercices. En revanche les taux de reconnaissance pour chaque exercice pris individuellement sont identiques (exercices *transfert d'objet* et *découpe de patron*) ou inférieurs au taux de reconnaissance obtenus avec la prédiction du score GOALS pour chaque exercice (tableau 6.13). Il

	Acc_w
Ex. 1	57.8%
Ex. 2	56.8%
Ex. 3	52.2%
Ex. 4	48.4%
Ex. 5	57.8%

Tab. 6.13. – Taux de reconnaissance entre le classement obtenu par *k-means* à partir des métriques significatives pour chaque exercice et le classement de référence basé sur le score GOALS réel.

existe en particulier davantage d'erreurs de classification correspondant à un décalage de deux groupes de niveau.

De plus, cette méthode ne permet pas de comparer deux volontaires appartenant au même groupe entre eux, ce qui est possible avec le score GOALS prédit au vu du coefficient de corrélation de Spearman (qui corrèle des rangs).

6.5 Régressions partial least squares pour prédire chaque aptitude du score GOALS

Afin d'enrichir l'évaluation non supervisée obtenue par la régression PLS sur les scores GOALS de chaque exercice, je cherche à identifier les métriques significatives plus spécifiquement de chaque aptitude $G^{(i)}$, afin de retourner davantage qu'un score moyen par exercice au volontaire, en lui précisant quelles aptitudes sont mises en défaut lors de la réalisation d'un geste.

Je moyenne donc les métriques sur l'ensemble des cinq exercices, ainsi que le score GOALS de chaque composante $G^{(i)} = \sum_j G_j^{(i)}$. J'effectue ensuite une régression PLS sur l'ensemble des métriques, en recherchant les plus significatives pour prédire chaque aptitude. Les résultats sont regroupés dans le tableau 6.14, les VIP de chaque métrique sont présentées dans l'annexe B.

L'analyse des métriques significatives pour la régression PLS pour chaque aptitude fait apparaître quelques métriques plus fortement associées que les autres à certaines aptitudes. Ainsi le temps et la longueur de chemin parcouru sont davantage significatifs pour les aptitudes *autonomie* et *efficience* que pour les autres, la variation de profondeur le long de l'axe de l'instrument est également fortement significatives pour l'*autonomie*. On note également que l'aptitude *préhension* est la seule associée significativement à l'accélération par unité de longueur, ce qui pourrait traduire l'existence de pics d'accélération

Il est toutefois difficile d'identifier clairement quelles sont les métriques les plus pertinentes pour évaluer séparément chaque aptitude, il semble que les paramètres significatifs sont plus dépendants des exercices que des aptitudes du score GOALS

Métriques		$G^{(1)}$	$G^{(2)}$	$G^{(3)}$	$G^{(4)}$	$G^{(5)}$
temps	P1	-	-	-	-	-
longueur de chemin parcouru	P2	-	-	-	-	-
niveau de bruit par unité de temps	P4					
niveau de bruit par unité de longueur	P5	-	-			
vitesse moyenne	P6					
vitesse angulaire moyenne	P8					
déplacement angulaire par unité de longueur	P9					
accélération temporelle	P11					
accélération spatiale	P12				-	
rugosité	P13	-	-	-		
jerk temporel	P14					
jerk spatial	P15	+	+	+	+	+
variation de profondeur le long de l'axe moyen par unité de temps	P18					
variation de profondeur le long de l'axe moyen par unité de longueur	P19	-	-	-	-	-
variation de profondeur le long de l'axe par unité de temps	P20					
variation de profondeur le long de l'axe par unité de longueur	P21	-	-	-	-	-
nombre de passage par zéro du *jerk* par unité de temps	P23	-		-	-	
nombre de passage par zéro du *jerk* par unité de longueur	P24					
volume exploré	P25	-	-	-	-	-
taux de recouvrement des volumes	P26	-	-	-	-	-
économie de volume	P27		-	-	-	-
R^2 des vitesses de chaque main	P28	+	+	+		
distance aux axes des couples de vitesse	P29					
distance aux axes des couples d'accélération	P30					
distance aux axes des couples de *jerk*	P31					
distance à l'origine des couples d'accélération	P32					
distance à l'origine des couples de *jerk*	P33					
information mutuelle des vitesses	P34	+	+		+	+
information mutuelle des accélérations	P35	+	+	+	+	+
information mutuelle des *jerks*	P36					
Nombre de composantes dans le modèle		4	4	4	2	4

Tab. 6.14. – Résumé des paramètres les plus significatifs (VIP≥ 1) pour les différentes aptitudes élavluées par le score GOALS dans les modèles prédictifs PLS. '+' traduit une corrélation positive avec le score GOALS, '-' une corrélation négative. G^1 →*perception de la profondeur* ; G^2 →*dextérité bi-manuelle* ; G^3 →*efficience* ; G^4 →*préhension* ; G^5 →*autonomie*.

explorées.

Les mêmes évaluations que dans les sections précédentes ont cependant été réalisées. Les coefficients de corrélation entre les scores prédits et les scores réels pour les aptitudes étaient inférieurs aux coefficients de corrélation des mêmes scores pour les aptitudes. Le taux de reconnaissance était également inférieur à celui obtenu avec les scores déterminés par les exercices ($Acc_w = 63.4\%$).

Il semble donc que les métriques ne permettent pas de prédire finement les aptitudes du volontaire. La raison probable à cette prédiction peu discriminante est la trop faible variation des valeurs des aptitudes du score GOALS pour la réalisation d'un exercice donné, c'est à dire que ces valeurs ne varient que d'une unité (rarement deux) d'un item à l'autre pour chaque volontaire : un chirurgien semblera avoir l'ensemble de ses aptitudes dans la même zone de notation. Ce biais de classement était évité pour la notation des exercices grâce à l'anonymisation des vidéos, qui permettait de juger indépendamment deux exercices réalisés par le même volontaire.
Il pourrait être évité pour juger les aptitudes de deux façons :

- en modifiant le type d'exercice réalisé pour se concentrer au cours de chaque exercice spécifiquement sur une aptitude (par exemple en permettant une vue en trois dimensions lorsqu'on ne cherche pas à mesurer la perception de la profondeur, ou en faisant réaliser des tâches du type test de Purdue [Tiffin, 1948] pour mesurer la dextérité bi-manuelle), avec en contre-partie la nécessité de faire un compromis en diminuant la validité apparente du dispositif ;
- en réduisant le recrutement des volontaires à des niveaux plus homogènes (uniquement des chirurgiens par exemple) et en insistant auprès des évaluateurs pour qu'ils utilisent toute le spectre des notations proposées par l'échelle GOALS, en essayant d'évaluer chaque aptitude de façon la plus indépendante possible.

6.6 Discussion et conclusion pour l'analyse de gestes chirurgicaux

J'ai présenté, dans cette partie de ma thèse, une méthode d'analyse de la trajectoire d'un geste chirurgical permettant de prédire le score d'un opérateur lors de la réalisation d'exercices standardisés sur un banc d'entraînement de chirurgie laparoscopique. Cette méthode exploite les informations contenues dans la trajectoire de la pointe de l'instrument ainsi que dans son orientation, déterminées dans la partie précédente, pour définir des métriques significatives en vue de prédire la qualité du geste et de construire un modèle permettant d'obtenir un score pour quantifier cette qualité.

Ces modèles sont construits de façon robuste par une méthode de validation

croisée *leave one out*, ils constituent la connaissance a priori nécessaire à la quantification de la qualité du geste.

Je discute dans cette section les différents choix effectués dans cette analyse, les limites et les apports principaux de cette dernière avant de conclure.

6.6.1 Banc d'entraînement MISTELS

Le choix d'utiliser un banc d'entraînement pour l'acquisition des données plutôt que d'enregistrer des procédures chirurgicales réelles permet de standardiser l'environnement dans lequel va évoluer le volontaire, en excluant les variations anatomiques susceptibles de perturber le déroulement d'un geste. Il autorise le recrutement d'un panel de volontaires avec des niveaux davantage variés, ce qui peut faciliter l'identification des métriques les plus significatives pour quantifier la qualité du geste.

Ce choix répond également à des considérations pratiques :

- tout d'abord la méthode de suivi des instruments n'était pas suffisamment évaluée dans des conditions réelles pour pouvoir exploiter les trajectoires obtenues de façon fiable ;
- de plus, l'utilisation du banc d'entraînement permet de se placer dans une situation où la caméra est fixe, ce qui aurait été difficilement réalisable au bloc opératoire : l'asservissement visuel du robot porte-endoscope n'était pas au point au début de la phase d'acquisition des trajectoires, et l'utilisation d'une caméra fixe pour une intervention en laparoscopie rend l'exposition délicate, perturbant le geste du chirurgien. A contrario, l'absence de structures adjacentes à la structure d'intérêt dans le cadre du banc d'entraînement exclut ce problème d'exposition ;
- le banc d'entraînement permet une disponibilité accrue pour enregistrer les gestes des volontaires. L'acquisition de trajectoires au bloc opératoire aurait nécessité une période d'acquisition des gestes particulièrement longue pour obtenir suffisamment de matériel pour réaliser les analyses décrites dans cette partie. L'acquisition de trajectoires sur spécimens anatomiques ou sur animal aurait nécessité un nombre important de spécimens anatomiques frais, difficilement disponibles, et une disponibilité des chirurgiens synchrone avec les horaires d'ouverture du laboratoire d'anatomie, ce qui était rarement le cas.

Le choix d'utiliser le banc d'entraînement MISTELS [Derossis et al., 1998] plutôt qu'un autre (ADEPT [Hanna et al., 1998] par exemple) s'est fait en raison de la littérature abondante en faveur des validités apparente, de construction, concurrente et prédictive de l'évaluation portée par le score MISTELS, basée sur la réalisation des exercices de ce banc d'essai, ainsi que sur la reproductibilité des scores obtenus.

6.6.2 Définition du score de référence

J'ai choisi d'introduire un score de référence différent du score MISTELS, qui est basé sur le temps nécessaire à la réalisation de chaque exercice associé à des pénalités, afin de limiter l'influence de la métrique *temps* sur la prédiction de la qualité du geste. En effet, si le facteur *temps* est probablement corrélé à celle-ci, une régression basée sur le score MISTELS (qui dépend presque exclusivement de celui-ci) risque de minorer l'importance d'autres métriques.

Le score GOALS [Vassiliou et al., 2005] a été choisi pour plusieurs raisons :

- il a été largement étudié, et ses validités prédictive, concurrente et de construction ont été démontrées dans le contexte du bloc opératoire, qui est le contexte de travail auquel dans lequel on souhaite évaluer la qualité des gestes *in fine* ;
- il est indépendant de l'exercice effectué, contrairement au score OSATS qui requiert l'utilisation de *check-lists* spécifiques aux exercices ;
- la notation des items de ce score est réalisable par des observateurs externes à partir d'enregistrements vidéo des exercices, ce qui permet d'anonymiser les volontaires, et de limiter le biais de classification subjectif ;
- enfin, ce score évalue plusieurs items, qui reflètent différentes aptitudes du chirurgien. Cette notation séparée des facteurs entrant en jeu dans la qualité globale de réalisation du geste devait me permettre de créer un score pour chacune des aptitudes, et de retourner à l'opérateur une information plus qualitative qu'un score unique sur son geste, en mettant en évidence les aptitudes maîtrisées et celles qui faisaient défaut.

6.6.3 Choix des métriques explorées

Un grand nombre de métriques peuvent être définies dans l'objectif d'analyser la qualité d'un geste chirurgical. Je me suis efforcé dans le cadre de ce travail de ne définir que des métriques qui paraissent avoir une signification "physique" pour modéliser les différentes aptitudes décrites par le score GOALS. Certaines parmi celles-ci ont déjà été décrites dans la littérature (en particulier le temps, la longueur de chemin parcouru, le nombre de mouvements, la rugosité (inverse de la fluidité) et l'économie de volume), d'autres sont décrites ici pour la première fois à ma connaissance.

Il s'agit en particulier de la mesure précise du volume exploré grâce à la définition d'une α-forme à partir de la trajectoire et des mesures de coordination bi-manuelle : taux de recouvrement des volumes explorés par chaque main et information mutuelle au sein des couples (*main gauche, main droite*) pour les paramètres vitesse et accélération. En effet, le score GOALS évalue spécifiquement cette aptitude, ce qui incite à définir de nouvelles métriques qui puissent la prédire.

La liste des métriques définies dans ce travail n'est pas exhaustive, et de nombreuses autres métriques pourraient être imaginées. En particulier, la méthode de détermination de la trajectoire de l'instrument que j'ai développée ne me permet

pas de retrouver la rotation de l'instrument sur son propre axe, qui est pourtant probablement pertinente pour mesurer l'optimisation de l'orientation de l'instrument pour saisir un objet.

Enfin je ne mesure pas les contraintes appliquées sur le système, comme peuvent le faire par exemple Rosen et al. [2006] avec le dispositif BlueDragon. Les métriques dérivées de ces dernières sont pourtant sans doute pertinentes pour mesurer l'aptitude *préhension*.

Le choix des métriques explorées est donc dépendant du dispositif d'acquisition des données utilisé, qui lui même est dépendant du contexte dans lequel on souhaite analyser la qualité du geste du chirurgien. Les forces appliquées au système ne peuvent pas être enregistrées dans le contexte du bloc opératoire, tandis que la trajectoire de l'instrument peut l'être plus aisément. En revanche, l'orientation de celui-ci autour de son propre axe peut probablement être mesurée pour certains types d'instrument à l'aide d'un procédé de reconnaissance de forme dans l'image.

6.6.4 Méthode de prédiction

Le choix de la régression PLS pour identifier les métriques significatives pour la prédiction, puis pour prédire un score GOALS en fonction du score GOALS réel de chaque exercice, a été fait en concertation avec les statisticiens du CIC-IT afin de prendre en compte la probable colinéarité en certaines variables définies, ainsi que leur nombre important par rapport au nombre de volontaires recrutés.

L'intérêt de cette méthode est d'ordonner les variables en fonction de leur importance dans la prédiction, son inconvénient est d'être plus complexe à interpréter lorsque le nombre de composantes utilisées est important (il faut interpréter le signe du produit scalaire entre la projection de la métrique et la projection de la variable à prédire dans l'espace des composantes).

La méthode de validation croisée de type *leave one out* s'est imposée naturellement, vu la taille réduite de mon échantillon de volontaires.

La méthode de classification par *k-means* est une méthode de partition non-supervisée simple, qui me permet de définir trois groupes pour le score prédit de chaque exercice et le score moyen. D'autres algorithmes de partitionnement de données, tels des regroupements hiérarchiques ou des algorithmes d'espérance-maximisation, pourraient également convenir. La principale modification à apporter à cette classification serait une méthode d'exclusion des valeurs aberrantes du GOALS (qui expliquent les mauvais résultats de la classification sur l'exercice *suture extra-corporelle* en particulier) et une classification "floue" qui retourne une probabilité d'appartenir à chacun des groupes.

6.6.5 Limites de la prédiction

La méthode de prédiction du score GOALS présentée dans ce chapitre présente plusieurs limitations.

Tout d'abord, je me base pour calculer les métriques étudiées sur les trajectoires obtenues avec la méthode de suivi d'instruments que j'ai présentée dans le chapitre 3, filtrées avec l'approche présentée dans le chapitre 4. Or, comme je le montre dans la discussion de ces chapitres, la précision de cette méthode est limitée, et les trajectoires obtenues ne sont pas les trajectoires exactes. L'addition d'erreurs d'estimation de la trajectoire rend la détermination des métriques significatives moins robuste. Le fait d'avoir recruté des volontaires de niveaux fortement distincts me permet cependant d'identifier des paramètres prédictifs de la qualité du geste, mais une augmentation de la précision de la méthode de suivi est nécessaire pour quantifier plus finement la qualité des gestes réalisés au cours des exercices.

À l'inverse, une fois le modèle établi, on note que les principales erreurs de prédiction (erreur quadratique moyenne par groupe) et de classification se retrouvent chez les volontaires novices, en raison d'une plus grande variabilité dans la distribution des métriques de ce groupe et de valeurs de métrique aberrantes pour certains volontaires.

Il est probablement nécessaire d'augmenter la puissance de l'analyse en recrutant davantage de volontaires, en particulier parmi les internes et les chirurgiens, pour obtenir une prédiction plus fine de la qualité du geste.

De plus, comme le montre la section 6.5, je ne parviens pas à identifier de façon fiable des modèles permettant de prédire séparément les différentes aptitudes des chirurgiens, ce qui était un des objectifs de cette étude. Cet échec peut être expliqué par la faible variabilité des scores décrivant les aptitudes d'un volontaire pour un exercice donné, qui ne paraissent pas indépendants les uns des autres. Comme je le propose plus haut, il faut probablement développer des exercices plus spécifiques pour parvenir à isoler ces métriques, ce qui est délicat si l'on souhaite conserver des exercices ayant une validité apparente satisfaisante.

6.6.6 Apports principaux de la méthode

L'objectif principal de l'étude est cependant atteint, à savoir obtenir un score prédisant la qualité de réalisation globale d'un exercice de façon non supervisée, uniquement à partir de l'analyse des images endoscopiques. Le score prédit est fortement corrélé au score GOALS réel pour l'ensemble des exercices, à l'exception de l'exercice *découpe de patron*, ainsi qu'avec le score GOALS moyen.

Pour parvenir à cette prédiction, un modèle de régression PLS a été utilisé, ce qui est à ma connaissance une nouveauté dans le domaine. Cette méthode présente de nombreux avantages dans le cadre de la recherche de métriques pertinentes pour évaluer la qualité d'un geste chirurgical, où l'espace des métriques peut être important par rapport au nombre de données différentes, en sélectionnant les métriques explicatives de manière optimisée par rapport à une approche par régression en

composantes principales standard.

L'apport principal de ce travail reste l'identification de nouvelles métriques en plus de la validation de certaines déjà décrites dans la littérature, pour prédire la qualité d'un geste chirurgical. En particulier, je m'intéresse à des métriques étudiant les mouvements des deux mains et l'organisation de l'espace de travail. Par rapport à la littérature existante, je valide dans le contexte du banc d'entraînement l'importance des métriques *temps, longueur de chemin parcouru, rugosité, variation de la profondeur de l'instrument* et *économie de volume*. J'ajoute à ces métriques les métriques décrivant le *volume* exploré, le *taux de recouvrement des volumes* explorés par chaque main et l'*information mutuelle* dans les changements de position et de vitesse des mains (information mutuelle variables des couples (V_{gauche}, V_{droite}) et (A_{gauche}, A_{droite})).

6.6.7 Conclusion

J'ai identifié dans cette partie un sous-ensemble de métriques significatives pour la prédiction de la qualité du geste chirurgical, à partir de l'ensemble des métriques dérivées des trajectoires des instruments. Ces métriques permettent de prédire un score pour chaque exercice, qui est corrélé au score GOALS affecté par un observateur expert évaluant la qualité du geste de façon aveugle à partir d'un enregistrement vidéo.

Elles décrivent la rapidité d'exécution de la tâche, l'économie des mouvements lors de la réalisation de ces tâches, l'organisation de l'espace de travail et la coordination entre les déplacements des instruments, qui sont des descriptions pertinentes de la compétence des chirurgiens d'après les retours des chirurgiens les plus expérimentés que j'ai pu interroger.

Les modèles de prédiction obtenus sont spécifiques à la réalisation d'un exercice, il sera nécessaire de les généraliser pour qu'ils puissent s'adapter à un panel de tâches plus larges, et être utilisés dans des contextes différents. Cette généralisation pourrait par exemple s'effectuer en évaluant les métriques sur des fenêtres temporelles glissantes, et en adaptant le modèle de prédiction à la tâche active identifiée par une approche de type *Surgical Process Model*.

Enfin, les modèles décrits dans cette partie ne prennent pas en compte les erreurs procédurales, qui sont tout aussi importantes que les erreurs d'exécution pour évaluer la qualité des gestes en chirurgie. Dans une approche similaire à celle développée dans ce travail, ces erreurs pourraient être décrites par un score de type OCHRA [Joice et al., 1998] porté par un observateur externe sur un ensemble de procédures définies, avec comme métriques des mesures de concordance entre la procédure suivie par le chirurgien et une modélisation générique de la procédure.

Références

Kenneth P. Burnham et David R. Anderson : Multimodel inference understanding AIC and BIC in model selection. Sociological Methods & Research, 33(2):261–304, novembre 2004. 184

Jacob Cohen : Weighted kappa : Nominal scale agreement provision for scaled disagreement or partial credit. Psychological Bulletin, 70(4):213–220, 1968. 198

Lee J Cronbach : Coefficient alpha and the internal structure of tests. Psychometrika, 16(3):297–334, 1951. 180

Anna M. Derossis, Gerald M. Fried, Michal Abrahamowicz, Harvey H. Sigman, Jeffrey S. Barkun et Jonathan L. Meakins : Development of a model for training and evaluation of laparoscopic skills. The American Journal of Surgery, 175(6):482–487, 1998. 207

Michael Haenlein et Andreas M. Kaplan : A beginner's guide to partial least squares analysis. Understanding Statistics, 3(4):283–297, 2004. 189

G. B. Hanna, T. Drew, P. Clinch, B. Hunter et A. Cuschieri : Computer-controlled endoscopic performance assessment system. Surgical Endoscopy, 12(7):997–1000, juillet 1998. 207

P Joice, G.B Hanna et A Cuschieri : Errors enacted during endoscopic surgery—a human reliability analysis. Applied Ergonomics, 29(6):409–414, décembre 1998. 211

Z. N. Maan, I. N. Maan, A. W. Darzi et R. Aggarwal : Systematic review of predictors of surgical performance. British Journal of Surgery, 99(12):1610–1621, 2012. 185, 188

Björn-Helge Mevik et Ron Wehrens : The pls package : principal component and partial least squares regression in r. Journal of Statistical Software, 18(2):1–24, 2007. 189

R Core Team : R : A Language and Environment for Statistical Computing. Vienna, Austria, 2012. ISBN 3-900051-07-0. 180

Jacob Rosen, Jeffrey D Brown, Lily Chang, Mika N Sinanan et Blake Hannaford : Generalized approach for modeling minimally invasive surgery as a stochastic process using a discrete markov model. IEEE Trans Biomed Eng, 53(3):399–413, mars 2006. 209

Joseph Tiffin : Purdue pegboard test. 1948. 206

M.C. Vassiliou, L.S. Feldman, C.G. Andrew, S. Bergman, K. Leffondré, D. Stanbridge et G.M. Fried : A global assessment tool for evaluation of intraoperative laparoscopic skills. The American Journal of Surgery, 190(1):107–113, 2005. 208

Discussion, conclusion et perspectives

CHAPITRE 7

DISCUSSION GÉNÉRALE

Les limitations et les apports des différentes méthodes développées dans le cadre de ce travail de thèse ont déjà été largement discutés à la fin de chaque partie les présentant. Je m'applique dans ce chapitre à discuter de l'intérêt de l'approche proposée à un niveau d'analyse supérieur, c'est-à-dire en pensant aux problématiques liées à son utilisation en pratique courante. Je me focalise en particulier sur les notions d'acceptabilité par les utilisateurs finaux, sur l'intégration de l'approche à des modèles de procédure chirurgicale et sur son intérêt pour se diriger vers une optimisation de la qualité de la chirurgie pour les chirurgiens experts.

7.1 Acceptabilité par les chirurgiens de dispositifs d'assistance par ordinateur

Le développement de dispositifs d'évaluation de la qualité du geste d'une part, et d'aide opératoire d'autre part, doit s'efforcer de prendre en compte les contraintes des utilisateurs finaux, ici les chirurgiens, à tous les stades de la mise au point.

7.1.1 Acceptabilité des dispositifs d'évaluation

Actuellement le principe de l'évaluation au cours de la phase d'apprentissage de la chirurgie est largement accepté, et même encouragé, par les chirurgiens. Plusieurs modalités ont été discutées, et l'importance du développement d'"écoles de chirurgie" permettant l'entraînement sur des simulateurs virtuels, des bancs d'entraînement ou des modèles animaux a été souligné [Conseil National de la Chirurgie, 2005, Vallancien, 2006].

En revanche, le principe de la ré-évaluation de la qualité des pratiques chirurgicales au cours de l'exercice en tant que spécialiste est plus récent. Si son intérêt n'est pas discuté [Cuschieri et al., 2001], les modalités d'une telle ré-évaluation font encore débat. Actuellement en France, la ré-évaluation est assurée par les procédures d'accréditation de la Haute Autorité de Santé (HAS), qui sont basées sur le volontariat. La procédure d'accréditation en urologie est disponible sur le site de la HAS [Haute Autorité de Santé et Association Française d'Urologie, 2012]. On note qu'il s'agit principalement de formation continue et de recherche d'évènements porteurs de risque autour de la prise en charge du patient, et que ces procédures n'évaluent pas la qualité de la chirurgie *stricto sensu*.

Une étude récente [Ahmed et al., 2013], basée sur des interviews de chirurgiens au Royaume-Uni, met en évidence une demande pour le développement de dispositifs d'évaluation étudiant davantage les déterminants de la compétence et de la performance du chirurgien tels que définis par des experts, à savoir la prise de décision, les habiletés techniques et la relation patient-médecin. Les participants recommandaient d'étudier d'une part les résultats de la chirurgie en termes de morbi-mortalité, d'autre part les compétences du chirurgien avec des méthodes observationnelles, génériques et spécifiques à des procédures, similaires à celles décrites dans ce travail. D'une façon générale, les chirurgiens considéraient que l'évaluation devait avoir lieu au bloc opératoire, et non sur simulateur ou modèle animal, pour prendre en compte l'ensemble de l'environnement péri-opératoire, et que des critères d'objectivité devaient être définis (ils n'étaient pas en faveur d'une évaluation locale par les pairs).

Il semble donc acquis que les chirurgiens acceptent aujourd'hui d'être évalués dans leur domaine de compétence, dans une démarche d'amélioration de la qualité des soins apportés au patient. Il reste nécessaire de développer des méthodes suffisamment non-intrusives et abordables, tout en demeurant fiables et robustes, pour permettre le déploiement de ces procédures d'évaluation au sein du bloc opératoire. Ce travail s'inscrit dans cet effort, en permettant notamment le suivi de la trajectoire des instruments au cours du geste, et en identifiant certaines métriques significatives pour la prédiction de la qualité d'un geste sur banc d'entraînement, lesquelles pourront être explorées plus en détail pour la prédiction de cette qualité dans le contexte d'une chirurgie réelle.

7.1.2 Acceptabilité de dispositifs d'aide per-opératoire

Si la notion de "valeur pronostique" du chirurgien était impopulaire jusqu'à récemment, limitant le développement de dispositifs d'évaluation de la qualité du geste, les chirurgiens sont demandeurs depuis longtemps de dispositifs leur permettant d'améliorer la qualité de celui-ci.

Cependant, et comme expliqué dans le chapitre 1, la prise de décision du chirurgien peut être altérée si trop d'informations lui parviennent simultanément, la hiérarchisation de ces informations en fonction de l'objectif de la chirurgie n'étant pas

toujours triviale.
De nombreux travaux se sont penchés sur l'optimisation de l'organisation de la salle d'opération dans le futur [Cleary et al., 2005, Bharathan et al., 2012], afin de n'apporter au chirurgien que les informations les plus pertinentes pour l'étape chirurgicale en cours. Dans le cadre de cette thèse, cette problématique s'applique à la forme que l'on souhaite donner à l'évaluation de la qualité du geste, lorsque celle-ci sera fonctionnelle dans le contexte du bloc opératoire : doit-on retourner au chirurgien en permanence la position de ses instruments par rapport à différentes structures (cibles ou à éviter), ou faut-il uniquement l'alerter en cas de distance inférieure à un seuil ? De même, doit-on l'informer sur des signes de fatigue (augmentation des tremblements des instruments par exemple) en temps réel, ou faut-il lui retourner cette information à la fin de la chirurgie ?
Ces décisions doivent être prises en concertation avec les utilisateurs finaux, en optimisant le compromis entre l'apport d'informations supplémentaires au chirurgien et les perturbations qu'un trop grand nombre d'informations peuvent engendrer.

7.2 Fusion de l'approche avec des modèles de procédure chirurgicale

Dans cette optique d'apporter au chirurgien les informations dont il a besoin au moment où elles sont nécessaires, le développement de modèles des procédures chirurgicales apparaît comme une étape indispensable. Je ne reviens pas sur les travaux de ce domaine, déjà abordés dans le chapitre 2, mais présente les dépendances relatives entre les deux approches et les possibilités d'intégration de l'une à l'autre.
Mon approche de suivi de l'instrument peut servir de base pour des méthodes de reconnaissance du type d'instrument utilisé, en identifiant le type de pointe ou la couleur du manche, afin d'obtenir automatiquement l'information sur le type d'instrument utilisé à un instant donné. Elle permet également de localiser la position des instruments dans le repère du patient (après recalage du repère de la caméra avec un repère monde), ce qui peut apporter des informations supplémentaires sur la tâche en cours pour identifier la tâche active. De même, des trajectoires typiques peuvent être observées pour différencier certains gestes. Ces éléments de bas-niveau remplissent les caractéristiques énoncées par Bouarfa et al. [2011] pour l'information à intégrer dans la construction de modélisations de procédures chirurgicales de plus haut niveau (facilité d'acquisition, caractère discriminant entre les tâches, invariant aux variations de réalisation d'une tâche).
Dans une approche inverse, certaines métriques présentées ici, ainsi que d'autres à développer dans le futur, peuvent s'appliquer plus volontiers à certaines étapes de la chirurgie qu'à d'autres. La reconnaissance automatique de la tâche active en utilisant des modélisations génériques de la procédure chirurgicale sera donc utile pour sélectionner les métriques appropriées à un instant donné, ainsi que pour choisir la

forme de l'information que l'on souhaite retourner au chirurgien.

7.3 Vers l'optimisation du geste chirurgical

7.3.1 Compétences non-techniques du chirurgien

Je me suis beaucoup concentré dans ce travail de thèse sur la quantification de l'habileté technique du chirurgien, en dérivant des métriques de la trajectoire de ses instruments pour évaluer ses aptitudes sur un banc d'entraînement. Or il est acquis que la qualité d'un geste chirurgical ne dépend pas uniquement des compétences techniques du chirurgien, mais également des connaissances théoriques et de la capacité de jugement de ce dernier. Le poids relatif de ces composantes dans la qualité de réalisation d'un geste chirurgical a été évalué à 75% de prise de décision (regroupant connaissances théoriques et capacité de jugement) contre 25% d'habileté technique, le poids relatif de l'habileté technique augmentant dans le cas de la chirurgie minimalement invasive [Spencer, 1978].
Les connaissances et la capacité à prendre une décision adaptée à une situation donnée peuvent être évaluées par des scores de type OCHRA, mais leur évaluation automatisée est un des enjeux principaux du domaine des gestes chirurgicaux assistés par ordinateur à l'heure actuelle.
Les modélisations de la procédure chirurgicale sont une des approches les plus prometteuses dans l'objectif d'éviter les erreurs procédurales, en mesurant les écarts entre le déroulé habituel de la procédure et la procédure en cours. Si une erreur de procédure est détectée précocement, il est possible d'alerter le chirurgien avant que l'erreur n'ait des conséquences irréversibles.

7.3.2 Apport d'informations contextualisées au chirurgien

Enfin, afin d'aider le chirurgien dans le processus de prise de décision, il est nécessaire de lui apporter des informations adaptées au contexte chirurgical à un instant donné. Speidel et al. [2008] ont mesuré des distances entre les instruments, suivis visuellement, et des structures à éviter, pour déclencher des alertes en cas de proximité trop importante. En utilisant d'autres modalités d'acquisition (fluorescence, échographie per-opératoire [Voros et al., 2013]), il sera possible d'étudier le geste du chirurgien dans un environnement riche, et d'améliorer le résultat de la chirurgie en minimisant le risque de survenue d'aléas chirurgicaux, tels des marges carcinologiques positives (fluorescence) ou des troubles fonctionnels (marquage des bandelettes vasculo-nerveuses avec de la fluorescéine pour la chirurgie prostatique par exemple).
Le suivi d'instruments n'est donc qu'un élément dans cet environnement, dont les différentes composantes devront s'articuler de façon intelligente pour apporter au chirurgien la bonne information au bon moment pour rendre le geste meilleur.

Références

Kamran Ahmed, Reenam S. Khan, Ara Darzi, Thanos Athanasiou et George B. Hanna : Recertification : What do specialists think about skill assessment ? The Surgeon, 2013. 216

Rasiah Bharathan, Rajesh Aggarwal et Ara Darzi : Operating room of the future. Best Practice & Research Clinical Obstetrics & Gynaecology, 2012. 217

L. Bouarfa, P.P. Jonker et J. Dankelman : Discovery of high-level tasks in the operating room. Journal of Biomedical Informatics, 44(3):455–462, 2011. 217

Kevin Cleary, Audrey Kinsella et Seong K. Mun : OR 2020 workshop report : Operating room of the future. International Congress Series, 1281:832–838, 2005. 217

Conseil National de la Chirurgie : Recommandations du Conseil National de la Chirurgie. Conseil National de la Chirurgie, 2005. 215

Alfred Cuschieri, Nader Francis, Joy Crosby et George B Hanna : What do master surgeons think of surgical competence and revalidation ? The American Journal of Surgery, 182(2):110–116, août 2001. 216

Haute Autorité de Santé et Association Française d'Urologie : Programme d'accréditation en chirurgie urologique. http ://www.urofrance.org/fileadmin/medias/accreditation/oa/PAS_2008_000312.pdf, 2012. 216

Stefanie Speidel, Gunther Sudra, Julien Senemaud, Maximilian Drentschew, Beat Peter Müller-Stich, Carsten Gutt et Rüdiger Dillmann : Recognition of risk situations based on endoscopic instrument tracking and knowledge based situation modeling. pages 69180X–69180X–8, mars 2008. 218

F. Spencer : Teaching and measuring surgical techniques : the technical evaluation of competence. Bulletin of the American College of Surgeons, 63:9–12, 1978. 218

Guy Vallancien : L'évaluation de la sécurité, de la qualité et de la continuité des soins chirurgicaux dans les petits hôpitaux publics en france. Rapport technique, Ministère de la Santé et des Solidarités, 2006. 215

S. Voros, A. Moreau-Gaudry, B. Tamadazte, G. Custillon, R. Heus, M.-P. Montmasson, F. Giroud, O. Gaiffe, C. Pieralli, G. Fiard, J.-A. Long, J.-L. Descotes, C. Vidal, A. Nguyen-Dinh et P. Cinquin : Devices and systems targeted towards augmented robotic radical prostatectomy. IRBM, 2013. 218

CHAPITRE 8

CONCLUSION ET PERSPECTIVES

Dans un contexte de complexification croissante des tâches chirurgicales, liée au développement de nouvelles techniques de chirurgie minimalement invasive, il apparaît nécessaire de développer les outils permettant aux chirurgiens de maîtriser cette complexité. L'enjeu est d'apporter des soins de la meilleure qualité possible au patient, en le faisant bénéficier de tous les avantages qu'offrent ces nouvelles techniques et en compensant toutes les perturbations que leur utilisation peuvent engendrer. Dans cette optique, de nombreux travaux se penchent sur le développement des composantes de la salle d'opération du futur, qui permettra au chirurgien et à son équipe de disposer de toutes les informations nécessaires à la réalisation optimale de la procédure, de façon hiérarchisée et contextualisée. Ces développements concernent des dispositifs d'acquisition de données per-opératoire, la modélisation de la procédure chirurgicale pour présenter au chirurgien ces données de façon structurée et l'aider dans ses choix. Ils portent aussi sur l'évaluation de la qualité de réalisation du geste en temps réel pour valider les choix du chirurgien ou l'alerter en cas de choix potentiellement non optimal pour une situation identifiée. Dans ce contexte, cette thèse se focalise sur l'évaluation de la qualité du geste chirurgical, à partir d'images endoscopiques et d'informations a priori, obtenues grâce à la modélisation de la procédure chirurgicale.

Le choix d'analyser des images endoscopiques s'appuie sur deux principales motivations. Tout d'abord, elles correspondent à des procédures chirurgicales réalisées par laparoscopie, qui est une technique chirurgicale complexe à maîtriser en raison de modifications de la perception visuelle et tactile du site opératoire. L'intérêt de l'évaluation de la qualité du geste est donc majoré dans ce contexte, en particulier au cours de l'apprentissage, lorsque le chirurgien est dans la partie la plus raide de sa courbe d'apprentissage. De plus, le signal de la vidéo endoscopique est le signal principal de la procédure laparoscopique, et son enregistrement est aisé. On a donc

accès, pour presque l'intégralité de la procédure, à un signal continu qui contient la majeure partie de l'information disponible pour le chirurgien.

Je me suis concentré dans ce travail sur l'analyse de la trajectoire des instruments chirurgicaux, afin d'identifier des structures contenues dans ces trajectoires susceptibles de permettre la prédictions de l'habileté technique du chirurgien, qui est une des composantes de la qualité de la chirurgie.

La première partie de cette thèse a donc été consacrée à l'extraction de la trajectoire des instruments laparoscopiques à partir des images endoscopiques. Cette étape s'appuyait sur la connaissance des caractéristiques géométriques de l'instrument et de la caméra, pour établir un lien entre la position des bords de l'instrument dans l'image et sa position dans l'espace. J'ai exploité la relative fixité du point d'insertion dans la paroi abdominale pour modéliser les orientations possibles de l'instrument dans un espace discret réduit constitué d'une grille géodésique, et j'ai utilisé un algorithme de CONDENSATION pour limiter l'espace d'exploration des orientations possibles, en modélisant les déplacements de l'instrument selon une composante déterministe et une composante stochastique. La position de la pointe de l'instrument était alors déterminée en étudiant la distribution des valeurs des pixels le long de l'axe de l'instrument.
Une étape cruciale dans cette partie de détection de la position de l'instrument est le traitement appliqué à l'image endoscopique originale pour mettre en évidence les bords de celui-ci. Ce traitement faisait intervenir successivement un changement d'espace colorimétrique pour obtenir une image de saturation, une étape de filtrage réalisé à l'aide d'un filtre bilatéral, et une détection des bords par un algorithme de Canny. La détermination des paramètres utilisés pour cette dernière étape était réalisée par une méthode statistique permettant d'optimiser leur choix.
Enfin les trajectoires obtenues ont été filtrées pour éliminer les mesures aberrantes d'une part, les bruits de faible amplitude d'autre part.
Les résultats obtenus jusqu'à présent montrent que le suivi de la position des instruments dans l'espace est réalisable dans les conditions du laboratoire, mais que la méthode doit encore être améliorée avant de pouvoir être utilisée au bloc opératoire.

À partir des trajectoires obtenues avec la méthode développée dans la première partie, j'ai défini dans la deuxième partie de cette thèse un ensemble de métriques, telles que le volume exploré par les instruments, l'économie de volume ou des mesures de coordination entre les mains du chirurgien. Ces métriques ont un sens visant à décrire plusieurs aptitudes attendues du chirurgien. Ces dernières sont définies par les différents items d'une échelle d'évaluation de la qualité du geste par des observateurs extérieurs.
Le jeu de données utilisé pour étudier le lien entre les métriques définies et les aptitudes techniques du chirurgien a été obtenu en faisant réaliser à des volontaires des

exercices sur un banc d'entraînement.
Dans un premier temps, j'ai vérifié la validité de l'échelle de notation utilisée comme référence pour évaluer la qualité de réalisation des exercices, en corrélant les résultats obtenus au statut des volontaires. Puis j'ai défini des groupes de niveau de référence en me basant sur les scores obtenus avec cette méthode à l'aide d'un classifieur *k-means*. Puis j'ai utilisé la méthode de régression *partial least squares* pour identifier les métriques les plus significatives pour prédire la qualité du geste, et pour prédire un score, de façon non supervisée, pour les exercices réalisés. Je n'ai cependant pas réussi à obtenir un score différent pour chaque aptitude, en raison d'une variabilité trop faible d'une aptitude à l'autre pour chaque volontaire de mon jeu de données. Ceci limite l'intérêt de la méthode développée pour le moment, car l'intérêt de l'évaluation de la qualité du geste chirurgical n'est pas seulement d'obtenir un score, mais bien de fournir au chirurgien un retour explicatif sur ses erreurs afin de lui permettre d'améliorer son geste.

Les limites actuelles du dispositif de suivi d'instrument sont la nécessité de traiter les vidéos en temps différé pour obtenir des trajectoires fiables pour la suite de l'analyse, et l'absence d'évaluation de ses performances en conditions réelles, avec en particulier des doutes sur la robustesse de la mesure. Ces limitations techniques doivent cependant pouvoir être levées en optimisant l'implémentation de la méthode d'une part, et en intégrant des méthodes d'analyse d'image plus élaborées d'autre part.
L'analyse des métriques significatives a été conduite sur un échantillon de volontaires de faible effectif ; le recrutement de davantage de chirurgiens, sur des exercices plus spécifiques aux aptitudes que l'on souhaite évaluer, doit permettre d'identifier de façon plus fine les métriques déterminantes pour chacune de ces aptitudes afin d'offrir au chirurgien une évaluation plus riche au cours de son apprentissage.

Ce travail constitue donc un élément d'un système d'évaluation de la qualité du geste plus contextualisé, les trajectoires obtenues pouvant servir de base de travail à d'autres dispositifs. Elles permettent actuellement d'asservir un robot porte-endoscope en temps réel à la position des instruments dans l'image endoscopique. Elles pourront dans le futur servir à évaluer en temps réel la position des instruments par rapport à d'autres organes, et grâce à la fusion d'informations multi-modales et à la modélisation des procédures chirurgicales, aider le chirurgien à améliorer la réalisation de son geste dans toutes ses composantes, techniques et non-techniques.

Annexes

ANNEXE A

RÉSULTATS DE LA RÉGRESSION PLS POUR LES EXERCICES

A.1 Exercice 1 : Transfert d'objets

VIP

```
      P1       P2       P4       P5       P6       P8       P9       P11
1.3560540 1.4577433 0.8113883 1.0436521 0.9897740 0.9751990 0.1998147 0.9234059
     P12      P13      P14      P15      P18      P19      P20      P21
0.5601058 0.9379945 0.8332465 1.2394003 0.9430856 0.8021543 0.9614839 0.6927570
     P23      P24      P25      P26      P27      P28      P29      P30
1.1816996 1.0010484 1.4922855 1.1089619 1.1147823 0.3454384 0.9171515 0.8289047
     P31      P32      P33      P34      P35      P36
0.9287351 0.8651141 0.8101127 1.3985786 1.0849181 0.9788851
```

Variables expliquant le score GOALS: VIP > 1

P1 P2 P5 P15 P23 P24 P25 P26 P27 P34 P35

Nombres de composantes utilisées :

2

A.2 Exercice 2 : Découpage de patron

VIP

```
      P1       P2       P4       P5       P6       P8       P9       P11
1.7628662 1.9280290 0.5812087 0.7757422 0.2340898 0.5441644 0.3866501 0.3243527
     P12      P13      P14      P15      P18      P19      P20      P21
1.1930211 1.5556424 0.5375742 1.0887518 0.1794426 1.2149110 0.1741307 1.2710770
```

```
      P23       P24       P25       P26       P27       P28       P29       P30
0.4917925 0.4406328 1.8235569 1.7501633 1.1770578 0.3011265 0.3172519 0.5432814
      P31       P32       P33       P34       P35       P36
0.6073159 0.3988858 0.5954540 1.5501818 0.9809949 0.4622114

Variables expliquant le score GOALS : VIP > 1

P1 P2 P12 P13 P15 P19 P21 P25 P26 P27 P34

Nombres de composantes utilisées :

2
```

A.3 Exercice 3 : Placement d'un Endoloop

```
VIP
       P1        P2        P4        P5        P6        P8        P9       P11
1.2540371 1.2548710 0.6773932 0.9598951 0.8125548 1.1001467 0.6385418 0.5822601
      P12       P13       P14       P15       P18       P19       P20       P21
1.0502026 1.1074794 0.6258392 1.4706528 0.6310628 1.0452853 0.6119244 1.1328522
      P23       P24       P25       P27
1.0713264 1.0610551 1.1676225 1.0888436

Variables expliquant le score GOALS : VIP > 1

P1 P2 P8 P12 P13 P15 P19 P21 P23 P24 P25 P27

Nombres de composantes utilisées :

4
```

A.4 Exercice 4 : Suture extra-corporelle

```
VIP
       P1        P2        P4        P5        P6        P8        P9       P11
1.8986877 1.8920239 0.2581793 0.4810338 0.5248275 0.5662541 0.5775147 0.5783944
      P12       P13       P14       P15       P18       P19       P20       P21
0.3810440 0.9233858 0.5407143 1.7191075 0.5463881 0.8145620 0.5284900 0.7577711
      P23       P24       P25       P26       P27       P28       P29       P30
0.3449479 0.5453427 1.7855308 1.8061028 1.1427937 0.9848427 0.5455829 0.5289500
      P31       P32       P33       P34       P35       P36
0.5548071 0.5506978 0.4937861 1.6755518 1.2050145 0.5980188

Variables expliquant le score GOALS : VIP > 1

P1 P2 P15 P25 P26 P27 P34 P35

Nombres de composantes utilisées :

2
```

A.5 Exercice 5 : Suture intra-corporelle

VIP

```
      P1       P2       P4       P5       P6       P8       P9      P11
1.6603754 1.6482670 0.6858818 0.5142849 0.7114506 0.5490370 0.9020693 0.7764766
     P12      P13      P14      P15      P18      P19      P20      P21
0.5109787 0.8187827 0.8101473 1.6173172 0.8574941 1.4249265 0.8073554 1.1086353
     P23      P24      P25      P26      P27      P28      P29      P30
0.5439245 0.7137365 1.5582069 1.5506799 0.7573028 1.0397914 0.7691101 0.8081767
     P31      P32      P33      P34      P35      P36
0.6893208 0.7185467 0.7160703 1.2678878 0.9781492 0.3335340
```

Variables expliquant le score GOALS: VIP > 1

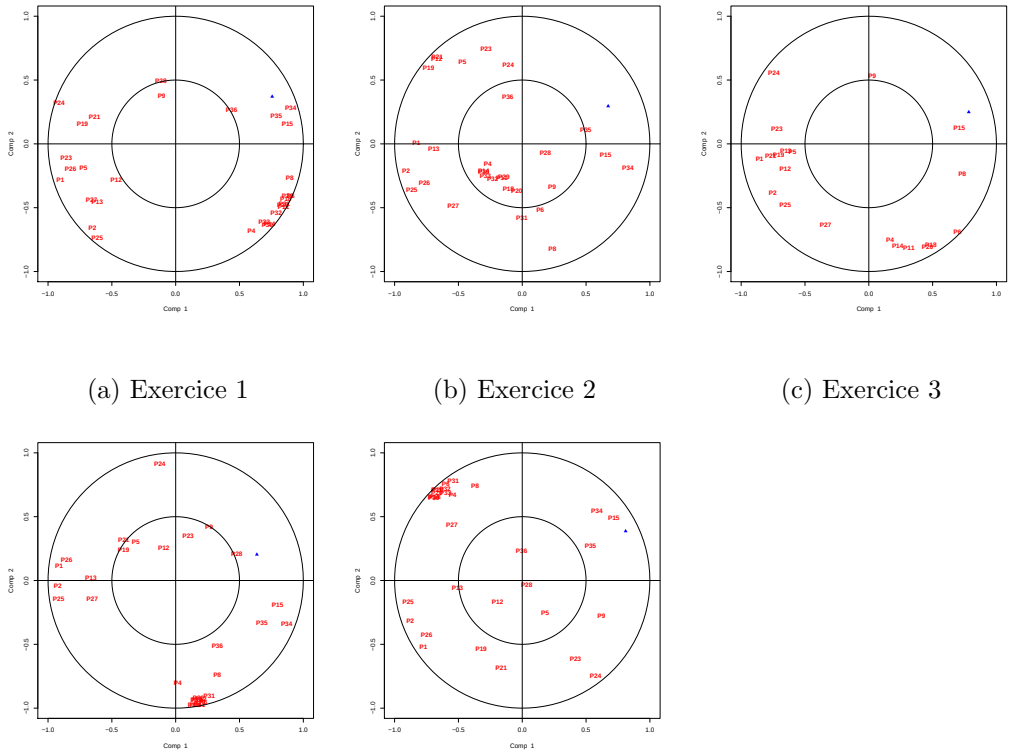
P1 P2 P15 P19 P21 P25 P26 P28 P34

Nombres de composantes utilisées :

4

A.6 Récapitulatif des paramètres explicatifs du score GOALS

P1	P12	P13	P15	P19	P2	P21	P23	P24	P25	P26	P27	P28	P34	P35	P5	P8
5	2	3	5	3	5	3	2	2	5	4	4	2	4	2	1	1



(a) Exercice 1 (b) Exercice 2 (c) Exercice 3

(d) Exercice 4 (e) Exercice 5

FIGURE A.1. – Représentation des variables sur les deux premières composantes des régressions PLS

ANNEXE B

RÉSULTATS DE LA RÉGRESSION PLS POUR LES APTITUDES

B.1 Aptitude 1 : Perception de la profondeur

VIP

```
      P1       P2       P4       P5       P6       P8       P9      P11
1.2906201 1.3320803 0.7671698 1.0111058 0.9238448 0.9152348 0.7265466 0.8564413
     P12      P13      P14      P15      P18      P19      P20      P21
0.9314878 1.2560765 0.7286897 1.3085170 0.7882876 1.0686001 0.8303138 1.0236436
     P23      P24      P25      P26      P27      P28      P29      P30
1.0667798 0.7342578 1.2795352 1.2255119 0.9414222 1.4732422 0.6868805 0.6998333
     P31      P32      P33      P34      P35      P36
0.8944091 0.7437925 0.7803421 1.2369612 1.0703794 0.5786631
```

Variables expliquant le score GOALS: VIP > 1

P1 P2 P5 P13 P15 P19 P21 P23 P25 P26 P28 P34 P35

Nombres de composantes utilisées :

4

B.2 Aptitude 2 : Dextérité bi-manuelle

VIP

```
      P1       P2       P4       P5       P6       P8       P9      P11
1.2400356 1.2963767 0.8306482 1.0912709 0.8829148 0.8668238 0.6644541 0.7997841
     P12      P13      P14      P15      P18      P19      P20      P21
0.9442102 1.3010296 0.6920250 1.4210581 0.7341981 1.1217337 0.7708862 1.0539786
```

```
      P23       P24       P25       P26       P27       P28       P29       P30
0.9997387 0.7164537 1.3117232 1.2396106 1.1664079 1.3413554 0.6815584 0.7738574
      P31       P32       P33       P34       P35       P36
0.7273917 0.6705297 0.6988510 1.2377303 1.1758743 0.5734950

Variables expliquant le score GOALS: VIP > 1

P1 P2 P5 P13 P15 P19 P21 P25 P26 P27 P28 P34 P35

Nombres de composantes utilisées :

4
```

B.3 Aptitude : Efficience

```
VIP

      P1        P2        P4        P5        P6        P8        P9       P11
1.3693162 1.4333186 0.6994118 0.9940864 0.8706784 0.8932987 0.6480973 0.7126097
      P12       P13       P14       P15       P18       P19       P20       P21
0.9457086 1.1136843 0.6572609 1.4228459 0.7335396 1.0368162 0.7519214 1.0146344
      P23       P24       P25       P26       P27       P28       P29       P30
1.0405932 0.7632622 1.4177011 1.2896427 1.1158660 1.3240864 0.6838892 0.6894642
      P31       P32       P33       P34       P35       P36
0.7590674 0.6913029 0.6867471 1.2623335 1.1518148 0.7644936

Variables expliquant le score GOALS: VIP > 1

P1 P2 P13 P15 P19 P21 P23 P25 P26 P27 P28 P34 P35

Nombres de composantes utilisées :

4
```

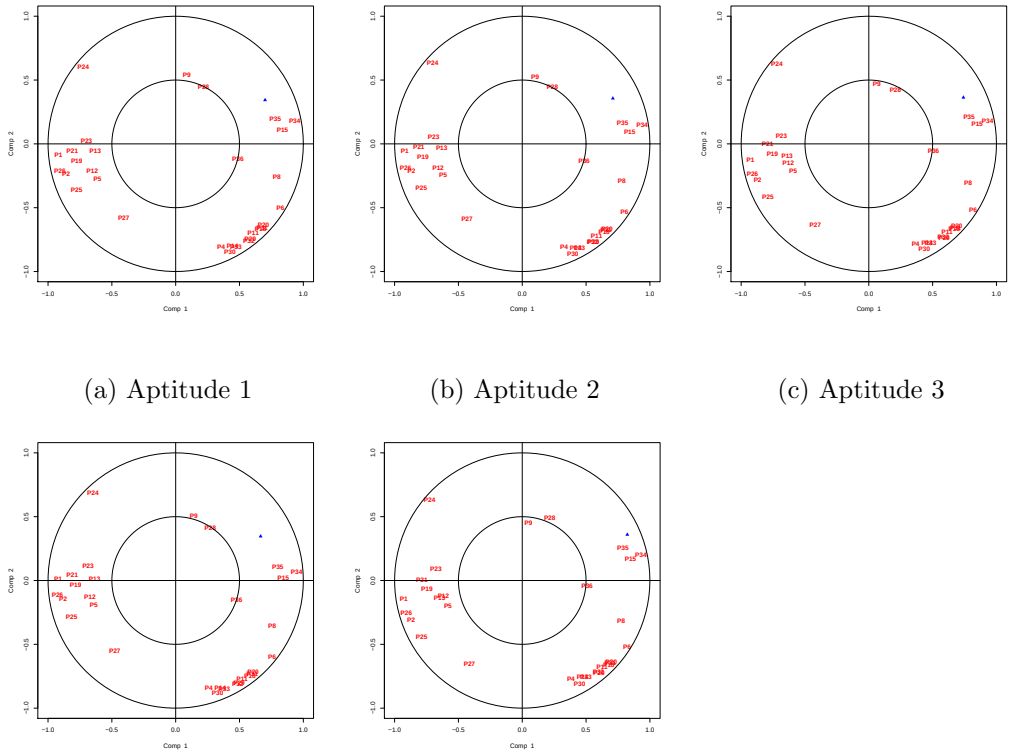
B.4 Aptitude 4 : Préhension

```
VIP

      P1        P2        P4        P5        P6        P8        P9       P11
1.4445863 1.5299526 0.5520537 0.8496652 0.8796382 0.9006255 0.5485116 0.6310110
      P12       P13       P14       P15       P18       P19       P20       P21
1.0634937 0.7281341 0.6611909 1.3767676 0.6915780 1.2140519 0.7032714 1.1489971
      P23       P24       P25       P26       P27       P28       P29       P30
1.0437395 0.7855463 1.5162727 1.4052837 1.2487802 0.3120892 0.6741941 0.7859515
      P31       P32       P33       P34       P35       P36
0.8580074 0.8145051 0.8862080 1.4141294 1.0938216 0.6405366

Variables expliquant le score GOALS: VIP > 1
```

```
P1 P2 P12 P15 P19 P21 P23 P25 P26 P27 P34 P35

Nombres de composantes utilisées :

2
```

B.5 Aptitude 5 : Autonomie

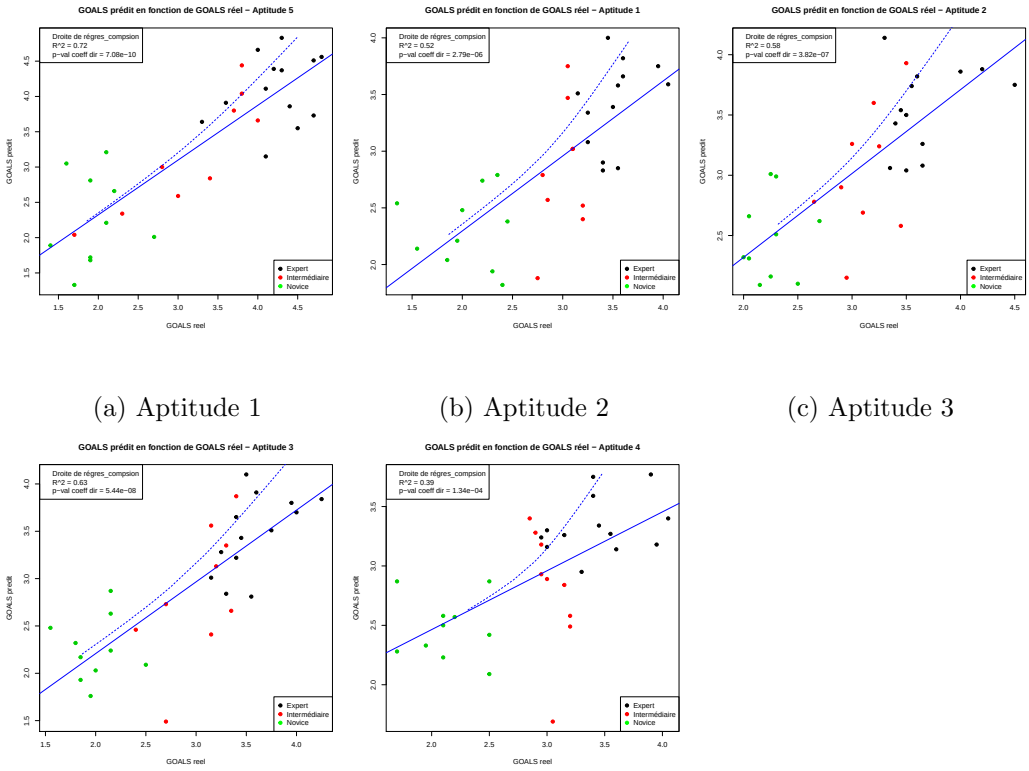
```
VIP

      P1        P2        P4        P5        P6        P8        P9       P11
1.4620862 1.5185478 0.5798068 0.9036391 0.9128671 0.9202663 0.6234210 0.7070897
     P12       P13       P14       P15       P18       P19       P20       P21
0.8871640 0.9898661 0.6118752 1.4896975 0.7576602 1.1652817 0.7723407 1.1719781
     P23       P24       P25       P26       P27       P28       P29       P30
0.9172680 0.8031364 1.5005158 1.3678119 1.1097198 0.6314310 0.6958858 0.6361405
     P31       P32       P33       P34       P35       P36
0.7804613 0.6833145 0.6260938 1.3653393 1.2201984 0.7894459

Variables expliquant le score GOALS: VIP > 1

P1 P2 P15 P19 P21 P25 P26 P27 P34 P35

Nombres de composantes utilisées :

4
```



(a) Aptitude 1 (b) Aptitude 2 (c) Aptitude 3

(d) Aptitude 4 (e) Aptitude 5

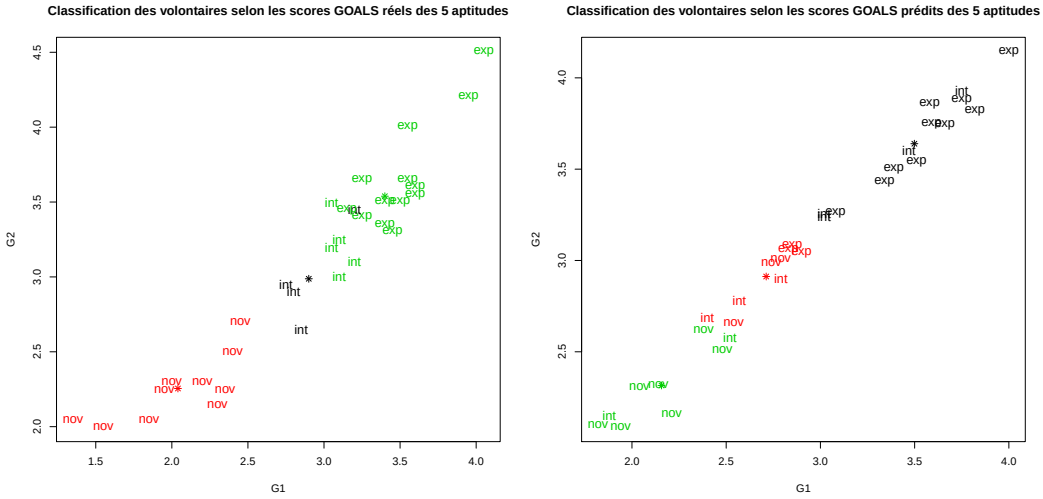
FIGURE B.1. – Représentation des variables sur les deux premières composantes des régressions PLS pour modéliser les aptitudes



(a) Aptitude 1 (b) Aptitude 2 (c) Aptitude 3

(d) Aptitude 4 (e) Aptitude 5

FIGURE B.2. – Régressions linéaires entre les scores GOALS prédits et réels pour les aptitudes.



(a) Classification basée sur les scores GOALS réels des aptitudes

(b) Classification basée sur les scores GOALS prédits des aptitudes

FIGURE B.3. – Comparaison entre les classifications *k-means* obtenues avec le score GOALS réel et le score GOALS prédit pour chaque aptitude. Les couleurs correspondent aux groupes déterminés à partir du classifieur, les labels à côté de chaque point (*exp, int, nov*) aux groupes de référence.

Résumé

Le développement de la chirurgie laparoscopique entraîne de nouveaux défis pour le chirurgien, sa perception visuelle et tactile du site opératoire étant modifiée par rapport à son expérience antérieure. De nombreux dispositifs ont été développés autour de la procédure chirurgicale afin d'aider le chirurgien à réaliser le geste avec la meilleure qualité possible. Ces dispositifs visent à permettre au chirurgien de mieux percevoir le contexte dans lequel il intervient, à planifier de façon optimale la stratégie opératoire et à l'assister lors de la réalisation de son geste. La conception d'un système d'analyse de la procédure chirurgicale, permettant d'identifier des situations à risque et d'améliorer la qualité du geste, est un enjeu majeur du domaine des Gestes Médico-Chirurgicaux Assistés par Ordinateur. L'évaluation de la qualité du geste explore plusieurs composantes de sa réalisation : les habiletés techniques du chirurgien, ainsi que ses connaissances théoriques et sa capacité de jugement. L'objectif de cette thèse était de développer une méthode d'évaluation de la qualité technique des gestes du chirurgien à partir de connaissances a priori, qui soit adaptée aux contraintes spécifiques du bloc opératoire sans modifier l'environnement du chirurgien. Cette évaluation s'appuie sur la définition de métriques prédictives de la qualité du geste chirurgical, dérivées des trajectoires des instruments au cours de la procédure. La première étape de ce travail a donc consisté en la mise au point d'une méthode de suivi de la position des instruments laparoscopiques dans la cavité abdominale au cours de la chirurgie, à partir des images endoscopiques et sans ajout de marqueurs. Ce suivi combine des modèles géométriques de la caméra, de l'instrument et de son orientation, ainsi que des modèles statistiques décrivant les évolutions de cette dernière. Cette méthode permet le suivi de plusieurs instruments de laparoscopie dans des conditions de banc d'entraînement, en temps différé pour le moment. La seconde étape a consisté à extraire des trajectoires des paramètres prédictifs de la qualité du geste chirurgical, à partir de régressions aux moindres carrés partiels et de classifieurs *k-means*. Plusieurs nouvelles métriques ont été identifiées, se rapportant à la coordination des mains du chirurgien ainsi qu'à l'optimisation de son espace de travail. Ce dispositif est destiné à s'intégrer dans un système plus large, permettant d'apporter au chirurgien, en temps réel, des informations contextualisées concernant son geste, en fusionnant par exemple les données issues de la trajectoire à des données multi-modales d'imagerie per-opératoire.

Mots clés : chirurgie, évaluation, qualité, apprentissage, suivi d'instruments, laparoscopie

Quantification of the quality of a surgical gesture from a priori knowledge

Abstract

The development of laparoscopic surgery has led to new challenges for surgeons, their visual and tactile perception of the operating field having been modified compared to their prior experience. Numerous devices have been designed around the surgical process in order to help the surgeon to perform the best possible intervention. These devices aim at enhancing the surgeon's perception of the operating context, optimally planning the surgical strategy, and assisting him/her during the intervention. The design of a system dedicated to the analysis of the surgical process, enabling the identification of risky situations and improving the quality of the surgery, is a major issue in the field of Computer Assisted Medical Interventions. The assessment of quality in surgery covers different aspects : the technical skills of the surgeon, as well as his/her theoretical knowledge and decision-making abilities. The objective of this thesis was to develop a technical skills assessment device for laparoscopic surgery, using a priori knowledge, suitable for the operating room's specific constraints without modifying the surgeon's environment. This assessment is based on the definition of predictive metrics for the quality of the surgery, derived from the instruments' trajectories during the procedure. The first step of this work consisted in the implementation of an instrument tracking method, based on endoscopic images without addition of any markers, in order to retrieve the 3D position of the instrument's tip inside the abdominal cavity during the surgical procedure. This tracking combines geometric models of the camera, the instrument and its orientation, with statistical models describing the evolutions of the latter. This method allows for simultaneous off-line tracking of multiple instruments within a training bench environment. The second step consisted in the extraction, from these trajectories, of predictive metrics for the assessment of the technical skills of the surgeon, using partial least squares regression and k-means classifiers. Several new metrics were identified, relating to the coordination of the surgeon's hands and the optimization of his/her workspace. This device is intended to be integrated in a more general system, in order to provide the surgeon with context-aware information regarding the surgical process, for example by merging data obtained from the trajectory with per-operative multi-modal image data.

Keywords : surgery, evaluation, quality, learning, instrument tracking, laparoscopy