

THÈSE

Présentée par :

SELKA Faïçal

Pour l'obtention du grade de

DOCTEUR

Spécialité Electronique Biomédicale

Suivi des points d'intérêt dans les vidéos en chirurgie coelioscopique pour la réalité augmentée : contributions à l'évaluation et à l'amélioration des performances de suivi par traitement d'images

Unité de recherche :

Laboratoire de génie biomédical - GBM Tlemcen

Institut de recherche contre le cancer de l'appareil digestif - IRCAD Strasbourg

Soutenue le 13 novembre 2014 devant le jury composé de :

Président :

M. CHIKH Mohamed Amine Professeur, Université de Tlemcen

Examineurs :

M. TALEB Nasreddine Professeur, Université de Sidi Bel Abbes

M. NICOLAU Stéphane Maître de Conférences, IRCAD Strasbourg

Invité :

M. LOUDJEDI Salim Maître de Conférences, Université de Tlemcen

Directeur de thèse :

M. BESSAID Abelhafid Professeur, Université de Tlemcen

Co-directeur de thèse :

M. SOLER Luc Professeur, IRCAD Strasbourg

TABLE DES MATIÈRES

Table des matières	i
1 Introduction	1
1.1 La chirurgie mini-invasive	2
1.2 Applications en vision par ordinateur pour la chirurgie laparoscopique	3
1.2.1 Reconstitution du champ visuel	3
1.2.1.1 Approche mosaïque d'image	4
1.2.1.2 Approche SLAM	4
1.2.2 Reconstitution 3D de la scène	5
1.2.2.1 Stéréo-vision	5
1.2.2.2 Estimation de la structure et du mouvement	6
1.2.2.3 Estimation de la structure 3D à partir des ombres	7
1.2.3 Réalité augmentée	7
1.3 Orientation des travaux de thèse	8
Les approches actives	8
Les approches passives	8
Plan de thèse	9
2 Suivi de points d'intérêt	11
2.1 Définition et généralité	12
2.2 Détection de points d'intérêt	12
2.2.1 Détecteur de Harris	14
2.2.2 Détecteur de Shi-Tomasi	16
2.2.3 Détecteur de Lowe : SIFT	17
2.2.4 Détecteur SURF	19
2.2.5 Détecteur FAST	19
2.3 Suivi de points d'intérêt	20
2.3.1 Approche récursive	20
2.3.1.1 Estimation du flux optique par les méthodes de corrélation de fenêtres	21
2.3.1.2 Algorithme de Lucas kanade	21
Limites	23
2.3.2 Approche descripteur	24
2.3.2.1 Descripteur SIFT	24
2.3.3 Descripteur SURF	26
2.3.3.1 Principe de l'appariement des points	26
2.3.3.2 Étude des appariements de points	27
Limites	27

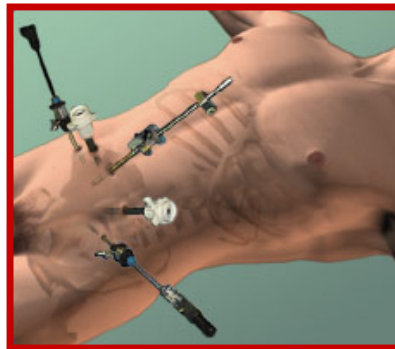
2.4	Bilan	27
3	Évaluation des performances de suivi	31
3.1	Évaluation du suivi	32
3.1.1	Approches directes	32
3.1.2	Approches indirectes	33
3.2	Détection des erreurs de suivi pour les approches récursives	34
3.2.1	Suivi avant-arrière basé sur les images paires-impaires	34
3.2.1.1	Qualité du suivi	35
3.2.1.2	Détection d'hystérésis	36
3.3	Méthodologie de comparaison des différents algorithmes de détection et de suivi	37
3.3.1	État de l'existant	37
3.3.2	Mise au point de la comparaison des techniques de suivi de points	38
3.3.2.1	Distribution	38
3.4	Évaluations	39
3.4.1	Influence des seuils des détecteurs	39
3.4.2	Influence du nombre des points à la détection	41
3.4.3	Pertinence de l'évaluation de la distribution spatiale	41
3.4.4	Évaluation du phénomène d'hystérésis	44
3.4.5	Synthèse	44
4	Amélioration des performances de suivi	47
4.1	Le suivi de points dans les images en chirurgie laparoscopique	48
4.2	Détection et suivi simultané sur les canaux RGB	49
4.2.1	Détection et sélection de points sur les canaux RGB	51
4.2.2	Suivi simultané de points sur les canaux RGB	51
4.2.3	Évaluation de la détection et du suivi simultané sur les canaux RGB	52
4.2.3.1	L'analyse RGB simultanée pour augmenter le nombre de points détectés	53
4.2.3.2	Évaluation de la robustesse du suivi de points en utilisant les canaux RGB	53
4.3	Pré-traitement pour l'amélioration des performances de suivi	55
4.3.1	Égalisation lissée	55
4.3.2	Chapeau haut de forme	56
4.3.3	Toggle	58
4.3.4	Toggle basé top-hat	58
4.3.5	Évaluation de l'influence d'un pré-traitement	58
4.3.5.1	Évaluation de l'influence du pré-traitement durant la phase du suivi de points	59
4.3.5.2	Évaluation de l'influence du pré-traitement à la détection de points	61
4.4	Dérivation d'une approche contexte spécifique pour le suivi de points	61
4.4.1	Description des indicateurs de performances	63
4.4.2	Évaluation du contexte-spécifique pour le suivi de points	64
4.4.2.1	Exemple d'utilisation	64
4.4.2.2	Évaluation de la persistance du contexte spécifique	65
4.4.2.3	Évaluation de l'influence du contexte spécifique sur les performances de suivi	68

4.4.2.4	Évaluation du temps de calcul	68
4.5	Synthèse	69
5	Application en réalité augmentée	71
5.1	Contexte clinique	72
5.1.1	Génération de la carte de perfusion	73
5.2	Recalage de la carte de perfusion	76
5.3	Estimation du mouvement des tissus	76
5.3.1	Calcul de la déformation	77
5.4	Évaluation du recalage de la carte de perfusion	79
5.4.1	Évaluation de la précision	80
5.4.2	Évaluation du temps de calcul	81
5.5	Synthèse	82
6	Conclusion	85
6.1	Contributions	86
6.1.1	Contributions méthodologiques	86
6.1.2	Contributions expérimentales	86
6.1.3	Contributions médicales	86
6.2	Perspectives	87
6.2.1	Perspectives théoriques	87
6.2.2	Perspectives pratiques	87
6.2.3	Perspectives médicales	87
A	Annexe - A	89
A.1	Espace de couleur RGB	89
A.2	Espace de couleur HSV	90
A.2.1	conversion RGB HSV	90
A.2.2	conversion HSV RGB	91
A.3	Espace de couleur HSL	91
B	Annexe - B	93
B.1	Les opérateurs morphologiques en traitement d'image	93
B.1.1	Élément structurant	93
B.1.2	Dilatation	94
B.1.3	Érosion	94
B.1.4	Ouverture	95
B.1.5	Fermeture	95
B.1.6	Chapeau haut de forme	95
C	Annexe - C	97
C.1	Séquences vidéos utilisées	97
C.1.1	Foie d'humain	97
C.1.1.1	Patient - A	97
C.1.1.2	Patient - B	98
C.1.1.3	Patient - C	98
C.1.1.4	Patient - D	98

C.1.2	Foie et rate de cochon	99
C.1.2.1	Foie	99
C.1.2.2	Rate	99
C.1.3	Intestins de cochon	100
C.1.3.1	cochon - A	100
C.1.3.2	cochon - B	100
Bibliographie		101

INTRODUCTION

La chirurgie mini-invasive procure de nombreux avantages pour le patient : morbidité post-opératoire réduite, une récupération plus rapide, un temps d'hospitalisation réduit, et il résulte de l'opération des cicatrices mineures par rapport à la chirurgie classique.



1.1 La chirurgie mini-invasive

L'un des progrès qu'a connu la chirurgie le développement de la chirurgie mini-invasive qui est effectuée en utilisant une caméra endoscopique, ainsi que plusieurs petits instruments chirurgicaux qui sont introduits à travers les orifices naturels du corps ou à partir de petites incisions réalisées sur le corps du patient. Réduire l'agression chirurgicale sans perdre la qualité du geste opératoire, tel est le but de cette nouvelle chirurgie. En fonction de la procédure chirurgicale et des équipes chirurgicales on parle de thoracoscopie si l'intervention se situe au niveau du thorax, de laparoscopie ou coéloscopie lorsque l'intervention se situe au niveau de la cavité abdominale, etc. Les travaux de recherche au sein de l'IRCAD ont toujours été orientés vers le développement de techniques chirurgicales de moins en moins invasives, ce qui lui confère aujourd'hui un statut de référence en chirurgie laparoscopique à travers le monde.

En chirurgie laparoscopique (cf. Fig. 1.1b) pour permettre une bonne visualisation de la cavité abdominale, la création d'un pneumopéritoine (cf. Fig. 1.2) par insufflation de gaz carbonique (CO_2) à l'intérieur de la cavité abdominale est nécessaire. Le chirurgien peut ainsi introduire l'endoscope par une petite incision et naviguer dans l'abdomen du patient. L'endoscope est muni d'un système d'éclairage et d'une caméra permettant de suivre le déroulement de l'intervention sur un moniteur (cf. Fig. 1.3b) grâce à une fibre optique reliée à la caméra. Avec l'utilisation d'instruments passant par d'autres incisions (cf. Fig. 1.3) le chirurgien peut effectuer des opérations chirurgicales ne nécessitant pas de pratiquer une laparotomie (cf. Fig. 1.1a). L'évolution de la caméra ainsi que des instruments chirurgicaux ont permis d'élargir la possibilité des interventions en laparoscopie. L'évolution des endoscopes ainsi que la miniaturisation des instruments ont permis un nouveau type d'intervention en chirurgie mini-invasive, qui dérive de la chirurgie laparoscopique. Connue sous le nom de chirurgie abdominale par voie transluminale (cf. Fig. 1.1c), elle consiste à l'utilisation d'un endoscope flexible couplé avec des micros-instruments introduit par un orifice naturel du patient. L'endoscope flexible est amené dans l'abdomen jusqu'à l'organe cible après l'incision d'une paroi interne (estomac, paroi utérine, colon) du patient. L'opération chirurgicale ne nécessite aucune cicatrice visible sur le corps du patient.

Il est clair que ce type de chirurgie mini-invasive limite les traumatismes opératoires et il résulte de l'opération des cicatrices mineures par rapport à la chirurgie classique. Les risques d'infection sont limitées car moins de tissus sont touchés, ce qui réduit la morbidité post-opératoire pour le patient.

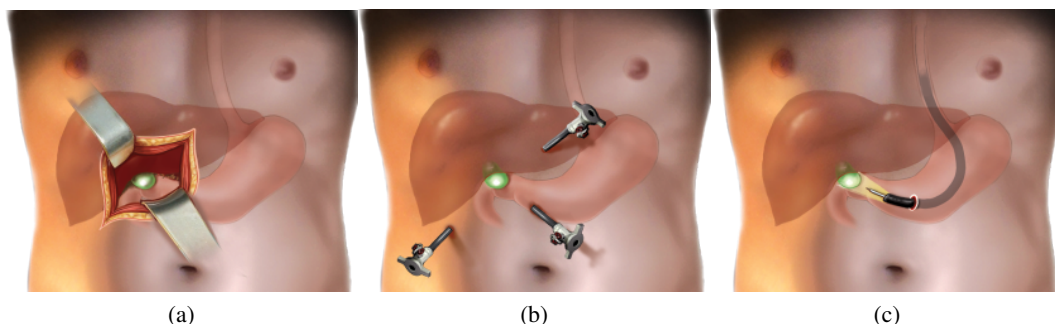


FIGURE 1.1 – Types de chirurgie : (a) chirurgie laparotomique, (b) chirurgie laparoscopique, (c) chirurgie transluminale

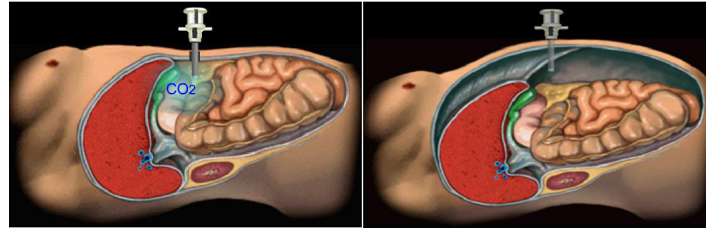


FIGURE 1.2 – Création d'un pneumopéritoine en insufflant du CO_2 pour gonfler la cavité abdominale du patient. Ceci permet la création d'un espace de travail pour le chirurgien durant l'intervention.

De plus, comme la durée d'hospitalisation et la récupération du patient est réduite, les coûts post-opératoires deviennent moins importants. Néanmoins du fait que le chirurgien n'a plus une vision directe de la zone qu'il opère, ces techniques présentent quelques inconvénients. En effet la limitation du champ visuel, la perte de la vision de profondeur, la mobilité réduite des instruments de chirurgie, le manque de rétroaction tactile du aux interactions limitées des instruments, la coordination entre vision et mouvement de l'endoscope qui est très complexe, perturbent le déroulement du geste chirurgical, qui peuvent rallonger le temps de l'intervention et aussi, être sujet à des erreurs médicales mettant la vie du patient en danger.

De ce fait, ces limitations ont encouragé plusieurs axes de recherche dans le domaine de la vision par ordinateur afin de fournir des informations pré-opératoires ou per-opératoires liées au patient (généralement des modèles 3D reconstruits à partir d'une image CT ou IRM) [BHN⁺12, HNR⁺10, HMG⁺09, HFB⁺08]. Ces informations ont pour but d'être fusionnées avec la vue du chirurgien durant l'intervention dans le but de faciliter le geste chirurgical mais aussi être considéré comme un moyen didactique ou de planification de l'intervention.

1.2 Applications en vision par ordinateur pour la chirurgie laparoscopique

Plusieurs axes de recherche dans le domaine de la vision par ordinateur ont été proposés comme un moyen pour surmonter certaines limites actuelles, donnant ainsi une véritable impulsion à la chirurgie laparoscopique. Généralement ce sont des systèmes basés sur une aide visuelle qui assiste les chirurgiens durant les interventions chirurgicales. De plus les progrès en imagerie interventionnelle et l'association au geste thérapeutique ont permis l'avènement de nouvelles perspectives dans ce domaine. Trois grandes applications dans le domaine de la vision par ordinateur en chirurgie laparoscopique qui nous intéresse, seront présentées ci-dessous.

1.2.1 Reconstitution du champ visuel

La vue endoscopique fournie au chirurgien sur le moniteur est limitée par l'angle de vue de la caméra. Le seul point d'insertion de l'endoscope à travers un trocart rends sa manipulation contraignante pour un changement de point de vue. De plus, situer le reste des organes dans la cavité abdominale pendant que le chirurgien se focalise sur un organe ou une région spécifique, nécessite un effort d'orientation et une certaine expertise.

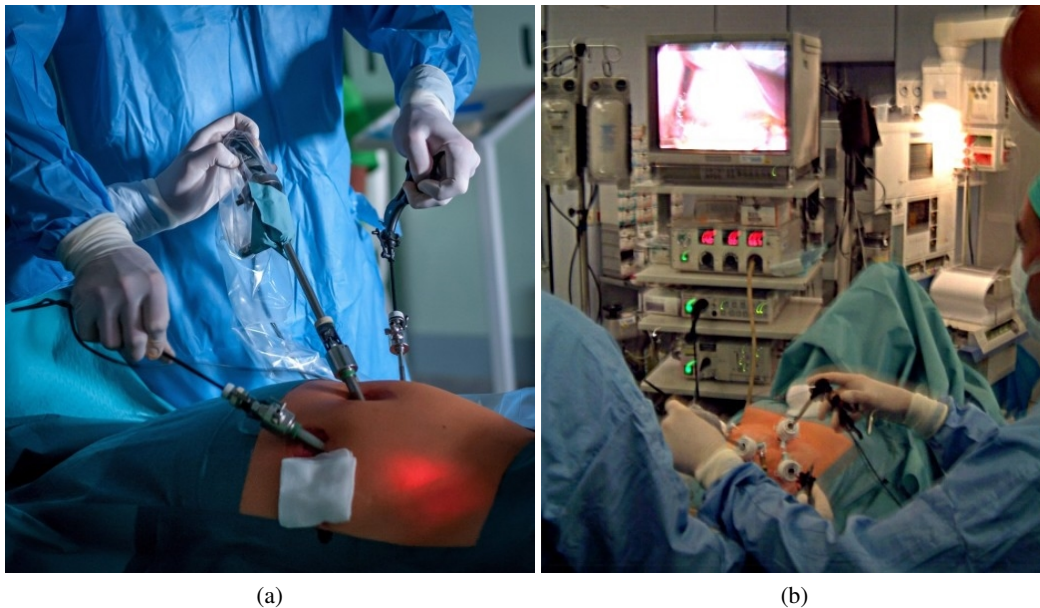


FIGURE 1.3 – Exemple d’une chirurgie laparoscopique.(a) Insertion de la caméra ainsi que des instruments chirurgicaux à travers les trocars, (b) visualisation de l’intérieur du patient avec un moniteur.

1.2.1.1 Approche mosaïque d’image

Cette technique qui est actuellement présente dans la plupart des appareils photos grand public, consiste en la création d’une image mosaïque ou appelée aussi d’une image panoramique. Le principe est de réaliser des captures de plusieurs images de la scène avec un mouvement de balayage et d’associer ces images en une seule image.

Les travaux en chirurgie laparoscopique [Vem, BBS⁺10] ont permis grâce à cette approche, d’élargir le champ visuel par association de plusieurs images grâce à un appariement de caractéristiques dans les images. En effet, à partir d’un balayage panoramique effectué par le chirurgien avec la caméra endoscopique ils reconstituent le champ visuel de la scène opératoire (cavité abdominale) (cf. Fig. 1.4) afin de faciliter la localisation d’organes non visibles.

1.2.1.2 Approche SLAM

Les approches récentes pour l’augmentation du champ de vue et facilitant ainsi la navigation *in-vivo*, reposent sur la méthode SLAM (Simultaneous Localization And Mapping) [MSDY06, MY09, Tot11]. La méthode proposée augmente le champ visuel efficace en créant incrémentalement un modèle 3D du tissu pendant que l’endoscope est dirigé autour de la scène chirurgicale. Le modèle 3D est employé pour augmenter la vue courante de l’endoscope avec l’information extérieure du champ visuel de la scène. Le modèle est projeté sur une caméra virtuelle située dans le même plan que l’endoscope, de ce fait créant un plus grand champ visuel que l’endoscope physique lui-même. La méthode établit un modèle 3D du tissu *in-situ* et récupère la position de l’endoscope relative au tissu.

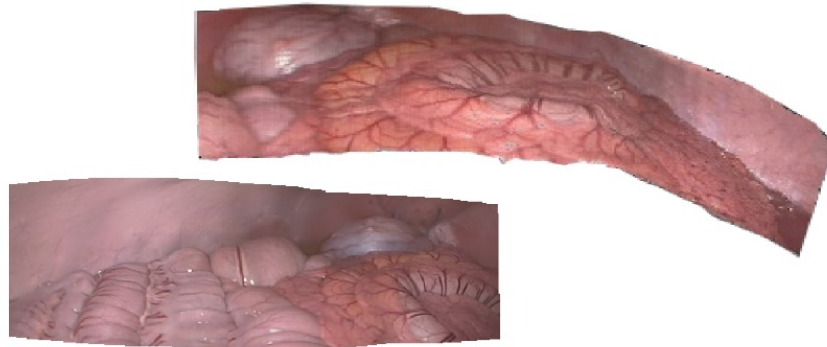


FIGURE 1.4 – Exemple de reconstitution du champ visuel.

1.2.2 Reconstitution 3D de la scène

Les endoscopes actuels capturent une image 2D de la scène, et livre cette information aux chirurgiens via un moniteur HD. Le système d'affichage ne peut donc pas transmettre la structure 3D de l'environnement. L'information de profondeur est perdue par le système d'affichage. En conséquence, les chirurgiens doivent déduire la structure 3D de la scène à partir de l'image plane pour manipuler des instruments et réaliser une tâche désirée. C'est une tâche difficile qui prend des années de formation pour être en mesure de concilier la vue fournie dans le moniteur avec la vraie scène 3D et de coordonner la position et les mouvements des instruments. La restitution de la profondeur ou la reconstruction 3D permet non seulement d'offrir une plus grande précision du geste chirurgicale mais aussi de pouvoir piloter des dispositifs robotisés dans le but d'assister le chirurgien. Plusieurs techniques sont utilisées pour atteindre cet objectif. D'une manière générale, le principe de la reconstruction 3D repose sur la détermination des coordonnées des éléments visibles de la scène 2D dans un repère de l'espace réel 3D à partir d'un ensemble d'images prises sous différents points de vue.

1.2.2.1 Stéréo-vision

L'emploi d'un système basé sur un endoscope stéréo permet en effet de restituer la vue en profondeur ceci en fournissant grâce aux deux caméras disposées à son bout, deux images de la même scène avec deux points de vue différents. Dans le robot Da Vinci le chirurgien place sa tête dans une console qui permet de renvoyer chaque vue d'un endoscope stéréo, à chaque œil (cf. Fig. 1.5). Cette double information renvoyée au cerveau permet par des mécanismes neurophysiologique de percevoir les reliefs ainsi que la distance entre les tissus et les instruments. Dans ce dispositif, aucune reconstruction 3D n'est réalisée ce qui ne permet pas d'extraire la structure réel 3D de la scène.

Cependant, en utilisant le principe de triangulation en géométrie épipolaire [HZ03] à partir de deux vues, la profondeur d'un point sur une image peut être calculée après calibration des caméras et mise en correspondance de points. Plusieurs applications en chirurgie laparoscopique [RBS⁺11, SMD⁺05, SVA⁺09, SVSPY10] utilisant des endoscopes stéréos ont lieu dans le but est de reconstruire la surface du tissu et de restituer la profondeur au chirurgien. La fiabilité de ces techniques reposent essentiellement sur la précision de la mise en correspondance de points entre les deux images. Par ailleurs, La reconstruction 3D est limitée par la distance séparant les deux caméras, du fait de la miniaturisation et du confinement de l'emplacement des caméras dans l'endoscope.



FIGURE 1.5 – Robot Da Vinci avec système de vision stéréoscopique.

1.2.2.2 Estimation de la structure et du mouvement

Dans une deuxième approche pour la reconstruction des modèles 3D en utilisant une simple caméra monoculaire, l'estimation de la structure et du mouvement (SFM) de la scène à partir d'une séquence d'images 2D [Mah12, HPE⁺07, AB12, CC11, Hu 10] permet en effet d'estimer la structure 3D d'un objet rigide à partir d'images 2D de la structure prise par une caméra en mouvement. Dans la figure 1.6 nous pouvons voir un exemple de reconstruction 3D utilisant la technique SFM, qui repose sur du suivi de points. Un pré-requis important est la détermination de la pose et du calibrage de la caméra, qui peut être exprimé par une matrice de projection. La théorie géométrique de structure du mouvement permet aux matrices de projection et des points 3D, d'être calculés utilisant simultanément des points correspondants dans chaque vue. L'appariement et/ou le suivi de points doit être précis et robuste pour assurer le bon fonctionnement de la méthode.

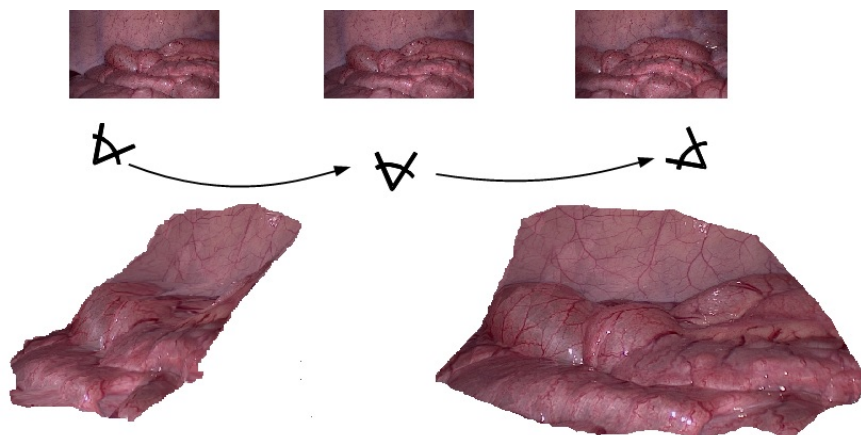


FIGURE 1.6 – Exemple de reconstitution de surface en utilisant le principe de la SFM.

1.2.2.3 Estimation de la structure 3D à partir des ombres

La technique du shape from shading (SFS) vise à reconstruire en 3D la scène à partir de l'analyse photométrique d'une seule image, en d'autres termes à analyser les niveaux de gris de l'image afin de caractériser l'orientation de la surface observée. Connaissant les caractéristiques photométriques des surfaces ainsi que la position et l'intensité de l'éclairage, il est possible d'obtenir une information différentielle sur la profondeur [Gou99]. Les travaux récents en chirurgie laparoscopique combinent les techniques SFM et/ou stéréo [SEY09, CB12, Mal12] afin de déterminer les caractéristiques photométriques des surfaces à partir de plusieurs points de vue.

1.2.3 Réalité augmentée

On appelle réalité augmentée l'insertion d'éléments virtuels dans une séquence vidéo « réelle ». À la différence de la réalité virtuelle, où la totalité de la scène est générée artificiellement, une séquence en réalité augmentée contient à la fois des éléments réels et virtuels. R. Azuma [Azu97] définit la réalité augmentée comme ayant les propriétés suivantes :

1. elle combine les éléments réels et virtuels dans un environnement réel,
2. elle est interactive, et par conséquent fonctionne en temps réel,
3. elle cherche à ajuster des objets réels et virtuels les uns par rapport aux autres.

La réalité augmentée en chirurgie laparoscopique [SVA⁺09, PSM13b, Nic11, Wan10, Iei11, HDP13] consiste le plus souvent à recaler un modèle pré-opératoire 3D ou 2D dans la vue endoscopique à partir de points anatomiques sur la surface du tissu ou de marqueurs. Généralement les modèles 3D proviennent des systèmes d'acquisitions tels qu'IRM et/ou TDM. En superposant une image pré-opératoire 3D aux images 2D qui s'affichent sur le moniteur du chirurgien, la vue virtuelle et réelle se complètent et fusionnent pour obtenir une vue plus informative. Actuellement, la plupart des systèmes de réalité augmentée pour les tissus mous sont axés sur la chirurgie du foie, en raison de sa pertinence clinique. En effet, le cancer du foie est l'une des causes les plus importantes de décès [ESR07]. Il existe donc un besoin pour améliorer le contrôle des marges de résection et des zones de risque en chirurgie laparoscopique. Par exemple, dans la figure 1.7 le chirurgien peut visuellement délimiter le tronc veineux qui alimente la tumeur et au-delà pratiquer une résection de la partie cancéreuse sans pour autant qu'il s'aide en palpant la surface du foie à la recherche de poulx indicateur d'un réseau sanguin. Disposer de cette information en temps réel offre la possibilité de rendre visible les structures internes du patient qui sont invisibles à l'œil du chirurgien et de retrouver le sens de toucher perdu en chirurgie laparoscopique. De plus, la réalité augmentée peut être utilisée dans des systèmes de navigation afin de diriger une aiguille durant les procédures d'ablation de la tumeur, qui peut également être fait sous cœlioscopie [MHTS⁺08]. Ces systèmes de navigation peuvent augmenter la précision du positionnement de l'aiguille. Le fait de la mobilité réduite de l'endoscope, des contraintes du mouvement diaphragmatique responsable de la respiration du patient et du déplacement des organes causés par les instruments introduits dans le corps du patient, le suivi de mouvements physiologiques (déformation des tissus) est nécessaire pour assurer un système fiable en réalité augmentée durant les procédures en chirurgie laparoscopique abdominale.

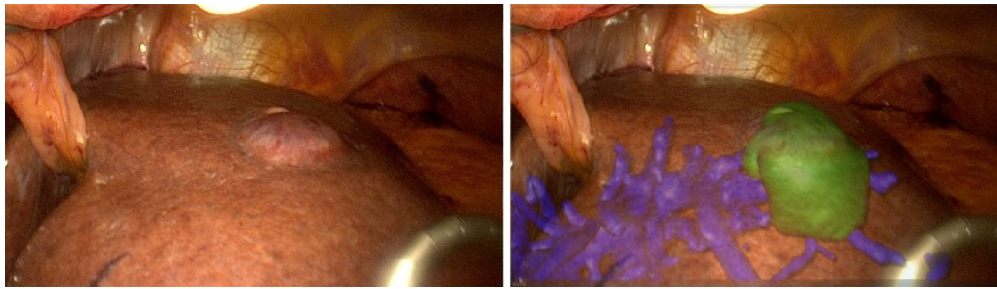


FIGURE 1.7 – Exemple de réalité augmentée. A gauche, vue endoscopique normal. A droite, recalage du modèle pré-opératoire (tumeur plus système porte veineux hépatique) sur la vue endoscopique.

1.3 Orientation des travaux de thèse

Disposer d'une information tridimensionnelle, et en temps réel, relative à l'environnement dans lequel opère un chirurgien permettrait d'envisager de nouvelles perspectives dans le domaine de la chirurgie laparoscopique. La restitution visuelle de la profondeur, l'expansion de la vue ainsi que la réalité augmentée assure un plus grand confort et une efficacité au déroulement du geste chirurgical. Le fonctionnement de ces méthodes repose sur la connaissance 3D en temps réel de la géométrie de la cavité abdominale. En d'autres termes sur la récupération et le suivi de la déformation du tissu qui doit être robuste et précise. D'autre part, le suivi de déformation des organes en temps réel pourrait aussi, par exemple, permettre de piloter des systèmes robotisés dans le but d'assister le chirurgien durant une intervention en chirurgie transluminale par compensation des mouvements physiologiques. Le problème de la localisation de la caméra ainsi que l'estimation de la déformation des tissus est donc cruciale et peut être résolu par diverses approches qui peuvent être classées comme actives et passives.

Les approches actives Elles concernent les méthodes employant un dispositif supplémentaire ou à une modification matériel de l'endoscope. Dans [HSN06] les auteurs proposent une approche basée sur la détection d'un faisceau laser projeté sur les tissus. Cette approche nécessite l'insertion d'un second endoscope pour la projection du faisceau laser ce qui rend le dispositif encombrant. Dans [AGD07] ils proposent d'utiliser la lumière structurée afin de reconstruire en 3D les scènes déformables. Une des caméras de l'endoscope stéréo est remplacée par un projecteur de motifs lumineux où chaque élément du motif est associé à un code. À partir du décodage des éléments apparaissant dans l'image ils calculent la localisation 3D des points imagés par triangulation. Cependant, aussi bien la conception du motif que la stratégie de codage associée nécessite un appariement robuste des points. De plus, la projection de motifs sur les tissus perturbe la vision du chirurgien dans le déroulement du geste chirurgical. Dans [Pen09] ils utilisent une caméra fonctionnant sur le principe du temps de vol (Time of Flight ToF), permettant de mesurer en temps réel la scène 3D. Le principe repose sur la mesure du temps que prend un pulse de lumière durant son trajet entre l'objet et la caméra. Le temps de vol de ce pulse est directement proportionnel à la distance entre la caméra et l'objet mesuré. Néanmoins à ce jour, cette technique présente une limite technologique à cause de la faible résolution des caméras ToF, qui relève son utilisation du domaine de la recherche.

Les approches passives Elles consistent en des techniques de vision basées uniquement sur les images endoscopiques. Ces points correspondent généralement à des caractéristiques visuels dans l'image. Une partie des travaux de cette thèse se portent sur le suivi de déformation des organes en chirurgie laparoscopique en se basant sur les images provenant de la caméra endoscopique. La littérature propose plusieurs méthodes dont la plupart s'appuient sur un suivi de points d'intérêt dans les images endoscopiques [MY11, GVSY12, Yip12]. Le suivi de points d'intérêt, est un problème difficile dans le cadre des images en chirurgie laparoscopique du fait de la propriété des images en chirurgie laparoscopique et de l'absence de vérité terrain pour évaluer et comparer les méthodes de suivi.

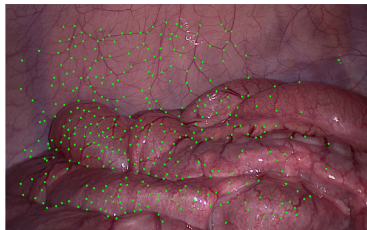
Plan de thèse Au **chapitre 2** nous donnerons une définition de la notion "**points d'intérêt**" dans les images ainsi que leur intérêt pour le suivi de déformation des organes. De plus nous décrivons les différents moyens de détection et de suivi connus dans la littérature. Dans le **chapitre 3** nous discuterons sur la méthodologie d'évaluation des performances du suivi de points dans les images endoscopiques. Nous verrons par la suite que la validation du suivi de points dans les images endoscopiques est très difficile, du fait de l'absence de vérité terrain. En général, les travaux sur l'évaluation des détecteurs et du suivi visuel [GHT11, MTS⁺05, MS05] se base sur une bibliothèque de donnée [DATA, DATb] qui permet de fournir une vérité terrain pour l'évaluation les performances de suivi. Néanmoins, du fait des différents types d'interventions, du changement de scène opératoire, de la diversité et complexité des tissus, la mise en place d'une vérité terrain standard est donc non envisageable. Nous verrons aussi dans le **chapitre 3**, comment à partir des méthodes de détection des erreurs de suivi et grâce à notre framework d'évaluation, nous pouvons fournir une vérité terrain artificielle et assurer une évaluation exhaustive des méthodes de détection et de suivi.

Les travaux sur le suivi de points dans les images endoscopiques, tendent à fournir le suivi le plus robuste et précis possible afin d'assurer le bon fonctionnement des systèmes d'aide à la navigation en chirurgie laparoscopique [BCBC12]. Plusieurs approches sur le suivi de points dans les images endoscopiques seront discutés dans le **chapitre 4**. Nous mettrons en évidence le fait que les approches de suivi en images endoscopiques, n'exploite pas la totalité de l'information de l'image provenant du capteur de la caméra et qu'aucune évaluation sur l'influence des approches basées sur le filtrage d'image afin d'améliorer les performances de suivi, n'a été réalisée. De ce fait, nous proposons une approche visant à utiliser les 3 canaux provenant du capteur de la caméra. Nous proposons également plusieurs types de pré-traitements qui seront décrit dans le **chapitre 4** afin d'évaluer leurs influences sur les performances de suivi. De plus, nous décrivons notre approche spécifique au contexte du suivi qui dérive de notre framework d'évaluation.

Finalement, au **chapitre 5**, nous présentons un système de réalité augmentée pour assister le chirurgien durant les procédures d'anastomose du colon. Notre système se base notre approche spécifique au contexte du suivi afin de fournir un suivi robuste et précis en fonction de la scène.

SUIVI DE POINTS D'INTÉRÊT

Connaitre la déformation des tissus est une partie importante pour plusieurs applications en chirurgie laparoscopique. La déformation des tissus peut être estimée à partir d'un suivi de points d'intérêt dans l'image. Un point d'intérêt est une forme ou structure facilement identifiable dans l'image.



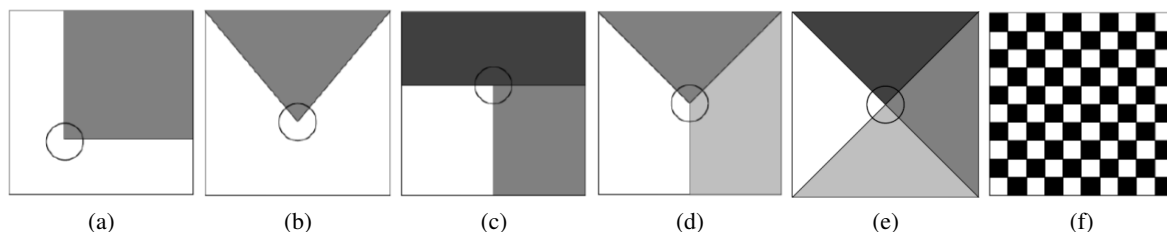


FIGURE 2.1 – Différent types de points d'intérêt (coin) : (a) jonction en «L», (b) jonction en «V», (c) jonction en «T», (d) jonction en «Y», (e) jonction en «X» et (f) jonction en «damier».

2.1 Définition et généralité

Dans de nombreuses applications de mise en correspondance d'images, reconnaissance ou suivi d'objets, la comparaison systématique de tous les pixels entre deux images n'est généralement pas envisageable car elle nécessite un temps de calcul très important. De plus, il existe des ambiguïtés de comparaison dans les régions faiblement texturées (région homogène et/ou contour cf. Fig. 2.2b). Le processus peut être simplifié en réduisant le domaine de recherche aux zones de l'image localement intéressantes. Les points référençant ces zones sont dénommés points d'intérêt. Informellement, un point d'intérêt correspond à une discontinuité des niveaux de gris (voire des couleurs), de la texture, de la géométrie, ... de l'image. Il est souvent assimilé à un coin. Les images de la figure 2.1 donnent une panoplie de types de coins qui se situent à des jonctions de type «L», «V», «T», «Y», «X» et «damier». Il est possible d'illustrer les différents types de coins dans les images réelles dans la figure 2.2. Il apparaît clairement qu'il n'y a pas une discontinuité franche entre les niveaux de gris à cause des bruits du capteur, du floutage et de la compression des images. Par ailleurs, dans le cas des images en chirurgie laparoscopique, les points spéculaires sont souvent assimilés à des coins lors de la phase de détection (cf. Fig. 2.2e). En effet, un point spéculaire correspond à une forte réflexion de la lumière sur le tissu, ce qui est fréquent dans les images en chirurgie laparoscopique, car les tissus sont fortement réfléchissants. Néanmoins ces points provoquent des erreurs de suivi car, ils ne suivent pas le mouvement des tissus, du fait qu'ils se trouvent dans l'alignement de la source de lumière localisée au bout de l'endoscope.

La capacité de suivre un point d'intérêt dépend de sa représentation qui comprend généralement : la couleur, les bords, les coins, la texture et son gradient. La sélection de ces points d'intérêts est une étape importante parce qu'elle a une influence directe sur la performance du suivi. Plusieurs stratégies dans la littérature de la vision par ordinateur ont été mises en place pour la détection de ces points d'intérêt [SMB00, TM07, Low04, Tom94] qui seront décrits ci-dessous.

2.2 Détection de points d'intérêt

De nombreuses méthodes ont été proposées pour détecter des points d'intérêts. Elles peuvent être classées suivant trois catégories :

1. Approches contours : à partir de la détection des contours dans une image, les points d'intérêts sont définis par l'intersection des contours à courbures maximales.

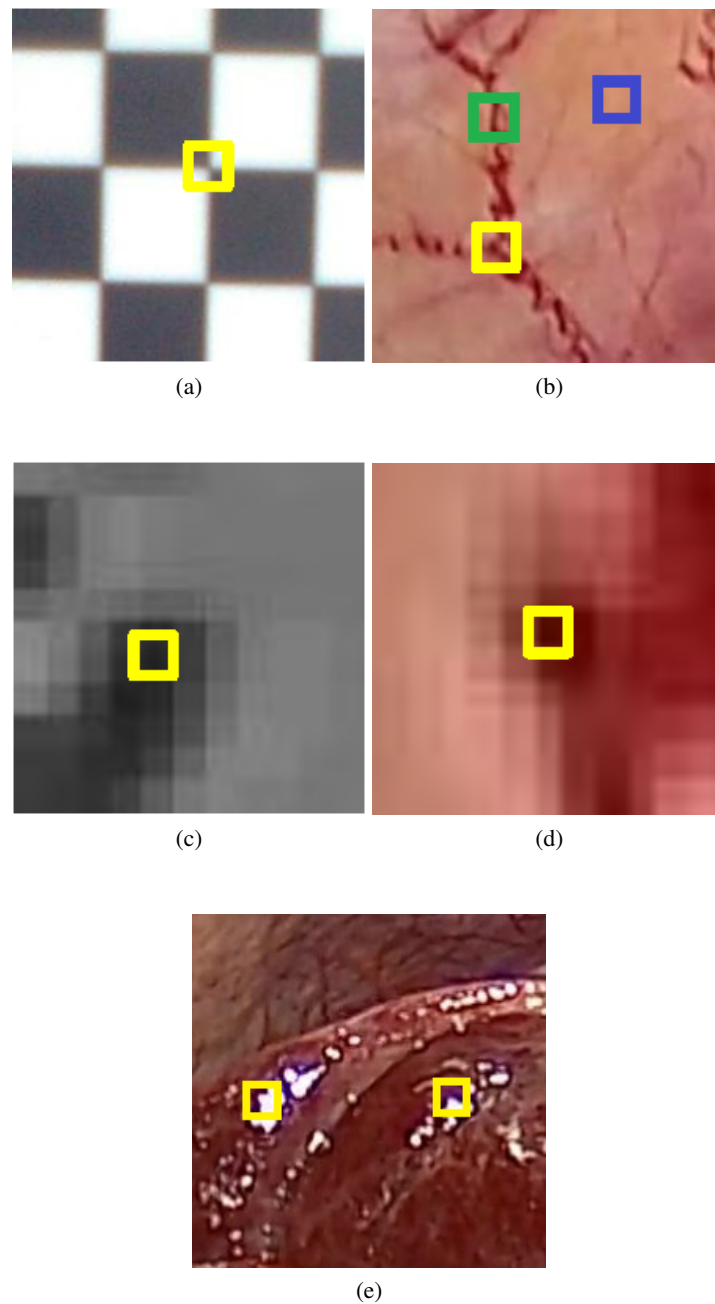


FIGURE 2.2 – Différent type de points d'intérêt (coin) dans le cas d'image réelle. (a) coin sur la jonction d'un damier, (b) coin sur la bifurcation d'un vaisseau sanguin, (c) coin sur une image en niveau de gris, (d) coin sur une image en couleur, (e) points spéculaires détecté comme coin. Dans l'image (b) les régions faiblement texturées sont représentées par le carré bleu (région homogène) et le carré vert (contour).

2. Approches intensité : à partir du calcul d'une fonction de gradient sur plusieurs directions, il est possible d'extraire directement les points d'intérêts sur l'image.
3. Approches à base de modèles : les points d'intérêts sont identifiés dans l'image par mise en correspondance de la fonction d'intensité avec un modèle théorique (préalablement connu) de cette fonction des points d'intérêts considérés.

Les approches basées sur l'intensité (gradient de l'image) sont les plus souvent utilisées dans le contexte des images en chirurgie laparoscopique du fait qu'elles ne reposent pas sur un modèle théorique, qui actuellement n'existe pas dans notre cadre ou à une détection des contours.

Un grand nombre de méthodes de détection de points ont été proposées dans la littérature. Pour guider le lecteur à comprendre les fondements physiques ainsi que les spécificités des détecteurs de points, nous présentons une étude détaillée des principaux détecteurs basés sur le calcul du gradient de l'image, qui seront par la suite utilisés dans nos expérimentations.

2.2.1 Détecteur de Harris

Le détecteur de Harris, proposé par Harris et Stephens [HS88], cherche à détecter des points en analysant les directions du gradient de l'image. Il est basé sur le calcul de la matrice d'auto-corrélation de son gradient, qui est souvent utilisé pour la détection des caractéristiques et la description des structures locales de l'image. Cette matrice décrit la distribution du gradient dans un voisinage local d'un point sur l'image. Soit la fonction de changement d'intensité :

$$E(x, y) = \sum_{u,v} W(u, v) [I(x + u, y + v) - I(u, v)]^2 \quad (2.1)$$

Où :

1. W spécifie la fenêtre/voisinage considérée (valeur inférieure à 1 à l'intérieure de la fenêtre et 0 à l'extérieur).
2. $I(u, v)$ est l'intensité au pixel (u, v) .
3. $E(x, y)$ représente la moyenne du changement d'intensité pour les déplacements (x, y) de la fenêtre.

En raison du caractère discret des changements de direction (rotation par 45°) de la fonction d'intensité, Harris et Stephens proposent d'effectuer le développement de Taylor de la fonction d'intensité au voisinage du pixel (u, v) :

$$E(x, y) = \sum_{u,v} W(u, v) \left[\left(x \frac{\partial I}{\partial x} + y \frac{\partial I}{\partial y} + O(x^2, y^2) \right)^2 \right] \quad (2.2)$$

En négligeant le terme $O(x^2, y^2)$ (valide pour les petits déplacements), la forme de l'énergie devient :

$$E(x, y) = Ax^2 + 2Cxy + By^2 \quad (2.3)$$

Avec :

- $A = \sum_{u,v} \left(\frac{\partial I}{\partial x} \right)^2 * W$
- $B = \sum_{u,v} \left(\frac{\partial I}{\partial y} \right)^2 * W$
- $C = \sum_{u,v} \left(\frac{\partial I}{\partial x} \frac{\partial I}{\partial y} \right) * W$

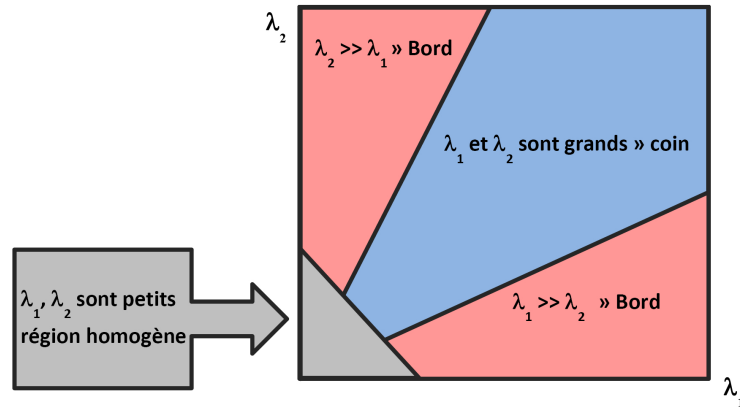


FIGURE 2.3 – Représentation de la réponse des valeurs propres pour la détermination des régions.

Les paramètres A , B et C de la matrice M impliquent les gradients de l'image qui sont souvent calculés en utilisant l'opérateur Sobel décrit ci-dessous :

$$S_x = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} \quad S_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \quad (2.4)$$

Harris et Stephen propose de lisser la réponse des gradients de l'image par convolution avec une fenêtre gaussienne :

$$W = e^{-(u^2+v^2)/2\sigma^2} \quad (2.5)$$

La gaussienne W , est calculée sur une fenêtre circulaire autour de chaque pixel, ce qui rend le calcul des gradients moins approximatif. Pour simplifier l'étude de la fonction $E(x, y)$, il est possible de la reformuler en une fonction quadratique :

$$E(x, y) = (x, y).M.(x, y)^T \quad (2.6)$$

$$\text{Avec : } M = \begin{bmatrix} A & C \\ C & B \end{bmatrix}$$

Cette matrice est une matrice symétrique 2×2 qui caractérise la moyenne des changements d'intensités dans toutes les directions. Il est possible de montrer que les valeurs propres de la matrice M nous renseignent sur les changements d'intensités principaux dans l'image (cf. Fig. 2.3). Une valeur propre élevée implique un fort gradient. La matrice M caractérise le comportement local de la fonction E , les valeurs propres $[\lambda_1, \lambda_2]$ de la matrice M renseignent sur les caractéristiques de la fenêtre considérée.

- Une valeur propre élevée = contour
- Deux valeurs propres élevées = coin

Les auteurs proposent de ne pas calculer les valeurs propres pour accélérer les performances de calcul et proposent l'opérateur suivant pour définir un coin dans une image :

$$R = \text{Det}(M) - k. \text{Trace}(M)^2 \quad (2.7)$$

Avec :

- $\text{Trace}(M) = \lambda_1 + \lambda_2 = A + B$
- $\text{Det}(M) = \lambda_1 \cdot \lambda_2 = AB - C^2$
- k compris entre 0.02 et 0.06.

L'intérêt majeur de ce critère est d'éviter le calcul explicite des valeurs propres. De manière simplifiée c'est d'interpréter géométriquement les courbures principales de la fonction d'auto-corrélation du gradient du signal. Une courbure importante dans les deux directions principales indique la présence d'un coin qui est retenu comme point d'intérêt. Le paramètre « k » est utilisé pour le calcul des réponses du détecteur de Harris. À aucun moment, Harris et Stephens ne mentionnent une quelconque valeur stratégique pour k . Certaines implémentations utilisent $k = 0.04$. Elle est partie intégrante de l'implémentation temps-réel exploitant le jeu d'instructions MMX d'INTEL [Par09]. Son rôle est essentiellement de faire la distinction entre les coins et les contours. Plus « k » est faible, plus la probabilité de confondre un point de bord pour un point de coin est élevé.

2.2.2 Détecteur de Shi-Tomasi

Le détecteur de coin de Shi-Tomasi [Tom91, Tom94] est basé sur le détecteur de coin de Harris. Cependant, une variation dans les critères de sélection a fait que ce détecteur soit plus performant que l'original, et qu'il soit plus utilisé pour la détection des coins [SMB00]. Le détecteur de coin de Harris opère avec un critère de sélection de coin. En effet, un score est calculé pour chaque pixel et si le score est au-dessus d'une certaine valeur, le pixel est marqué comme un coin. Par ailleurs, le score est calculé en utilisant deux valeurs propres et un paramètre k qui lui, influe considérablement sur la considération d'un coin ou non. Shi et Tomasi ont suggéré que la fonction de Harris (Eq. 2.7) doit être reconsidérée. Seules les valeurs propres devraient être utilisées pour définir si le pixel est un coin ou non. Pour Shi-Tomasi, il est calculé comme ceci : Dans un premier temps, ils recherchent une valeur seuil ϵ permettant de caractériser des points d'intérêt à partir des valeurs propres λ_1 et λ_2 de telle sorte que si $\min(\lambda_1, \lambda_2) > \epsilon$ alors le point sera considéré comme un coin. Pour le choix de ϵ , ils mesurent la valeur ϵ_0 sur une région avec un contraste uniforme (faiblement texturée) et ϵ_1 sur une région fortement texturées avec : $\epsilon_1 \gg \epsilon_0$. La valeur de ϵ est choisie à la moyenne des deux valeurs (ϵ_0, ϵ_1). Cette valeur influe sur le nombre de points détectés. Dans notre cas ϵ est adapté en fonction de nos expérimentations qui seront décrites dans les chapitres suivants. Pour calculer le minimum des valeurs propres d'une matrice M carrée 2×2 , la méthode classique est de calculer le polynôme caractéristique, et de chercher ses racines.

$$M - \lambda \cdot I_d = \begin{bmatrix} A - \lambda & C \\ C & B - \lambda \end{bmatrix} \quad (2.8)$$

Avec : I_d matrice d'identité

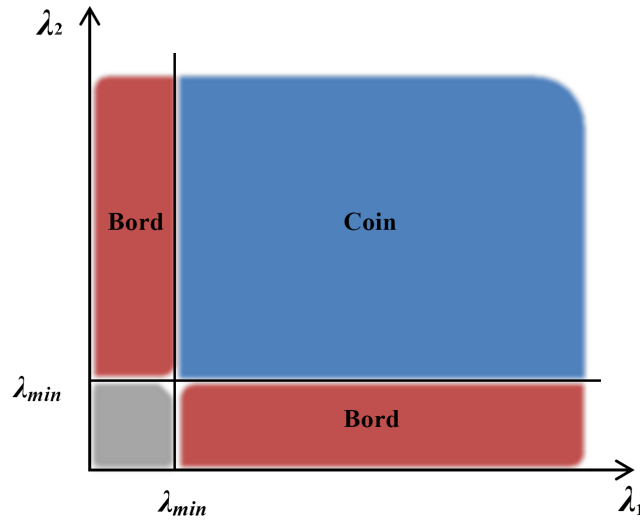
Le calcul du déterminant de cette matrice donne l'équation ci-dessous :

$$\lambda^2 - \lambda(A + B) + A \cdot B - C^2 = 0 \quad (2.9)$$

Il suffit de déterminer les solutions de cette équation. C'est une équation du second ordre avec deux solutions réelles. Dans le cas présent la matrice M , est une matrice symétrique diagonalisable, il suffit de déterminer une seule solution qui n'est que la plus petite valeur pour résoudre le système de ce cas.

$$X = (A + B)/2 - \sqrt{(A/2 + B/2)^2 - (A \cdot B - C^2)} \quad (2.10)$$

Pour rappel :

FIGURE 2.4 – Réponse de Shi-Thomasi suivant les valeurs propres et le seuil λ_{min} .

- $A = \sum_{u,v} \left(\frac{\partial I}{\partial x}\right)^2 * W$
- $B = \sum_{u,v} \left(\frac{\partial I}{\partial y}\right)^2 * W$
- $C = \sum_{u,v} \left(\frac{\partial I}{\partial x} \frac{\partial I}{\partial y}\right) * W$

Dans leur papier, Shi et Tomasi ont démontré expérimentalement que ce critère de score était bien meilleur. Ils ont évalué la détection de coins sur des images contenant des damiers ou des structures présentant des coins (tables, livres,..) . Si ϵ est plus grand qu'une certaine valeur prédéfinie, il peut être marqué comme un coin. Ainsi, un coin est simplement défini dans une région délimitée par une intersection entre deux régions (cf. Fig. 2.4).

2.2.3 Détecteur de Lowe : SIFT

L'algorithme SIFT (Scale Invariant Feature Transform) a été proposé par David Lowe [Low04], pour détecter et décrire des zones ou points d'intérêts dans une image. Il est connu pour son invariance à l'échelle, l'orientation et les changements d'illuminations. Grâce à ces descripteurs l'algorithme de SIFT permet de discriminer ces points détectés pour pouvoir identifier (mettre en correspondance) par la suite ces points d'intérêts dans d'autres images de la même scène. Cette partie sera discutée dans la section 2.3.2.

La première étape de l'algorithme permet la détection de différents points d'intérêts en tenant compte de leur facteur d'échelle. Le facteur d'échelle permet de modifier la résolution d'observation de ces points. En effet, un point peut être plus au moins franc/flou en fonction de l'échelle considérée. La modélisation d'une image à un facteur d'échelle σ est obtenue par convolution avec un filtre gaussien G de paramètre σ . Soit un point (x, y, σ) défini d'une part par ses coordonnées sur l'image (x, y) et d'autre part par son facteur d'échelle caractéristique (σ). On appelle représentation dans l'espace d'échelle le résultat de la convolution d'une image I par un filtre gaussien G de paramètre σ

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (2.11)$$

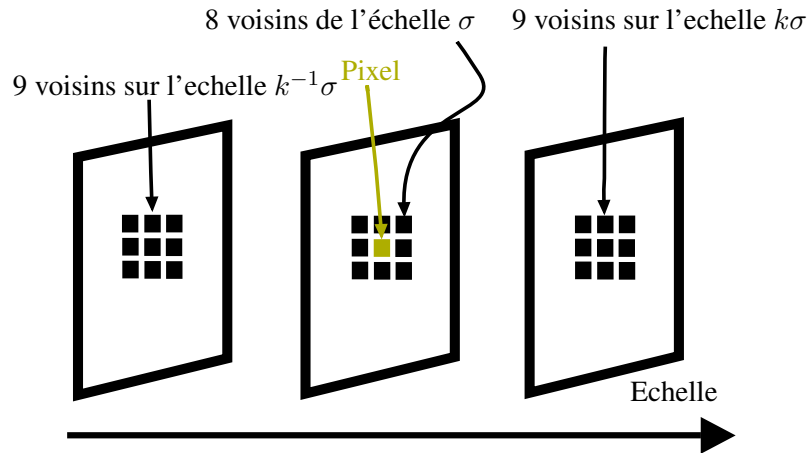


FIGURE 2.5 – Pyramide de différences de gaussiennes (DoG), exemple de recherche d’extrémums pour un pixel appartenant à la pyramide DoG. Recherche effectuée parmi les 26 pixels voisins appartenant aux échelles supérieure et inférieure ainsi qu’à l’échelle considérée.

La détection des points s’effectue à l’aide de la recherche d’extrémums d’une approximation de Laplacien. Cette approximation de laplacien peut s’obtenir comme la différence de deux images avec deux facteurs d’échelles différents [Lin99]. Cette approximation connue sous le nom de pyramide de différences de gaussiennes (DoG), est définie comme suit :

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (2.12)$$

Pour chaque pixel appartenant à une DoG, évaluée à une échelle, une comparaison est effectuée avec les huit pixels voisins de même échelle ainsi qu’avec les neuf pixels voisins appartenant respectivement à l’échelle voisine supérieure et inférieure (cf. Fig. 2.5). Visant à augmenter la précision des points détectés Lowe propose d’interpoler les extremums des valeurs discrètes de x, y et σ dans un espace à trois dimensions (x, y, σ) . Cette interpolation s’obtient par un développement de Taylor à l’ordre 2 de la fonction différences de gaussiennes $D(x, y, \sigma)$.

$$D(\mathbf{x}) = D + \frac{\partial D^T}{\partial \mathbf{x}} \mathbf{x} + \frac{1}{2} \mathbf{x}^T \frac{\partial^2 D}{\partial \mathbf{x}^2} \mathbf{x} \quad (2.13)$$

Les dérivées sont calculées au point candidat préalablement détecté et où $\mathbf{x} = (x, y, \sigma)^T$ représente le delta par rapport à la position de ce point. La position précise de l’extremum $\hat{\mathbf{x}}$ est déterminée en résolvant l’équation annulant la dérivée de la fonction 2.13 par rapport à \mathbf{x} on trouve ainsi :

$$\hat{\mathbf{x}} = - \frac{\partial^2 D}{\partial \mathbf{x}^2}^{-1} \frac{\partial D}{\partial \mathbf{x}} \quad (2.14)$$

Le delta $\hat{\mathbf{x}}$ est ajouté aux point candidat courant pour améliorer son estimation. Cependant un delta $\hat{\mathbf{x}}$ supérieur à 0,5 dans l’une des trois dimensions signifie que le point considéré est plus proche d’un des voisins dans l’espace des échelles discret. Dans ce cas, le point-clé candidat l’interpolation est réalisée à partir des nouvelles coordonnées. Lowe propose deux filtres afin d’éliminer les points candidats qui pourraient créer des ambiguïtés. En premier lieu, les points faiblement contrastés sont éliminés par un seuillage arbitraire. Ces points se caractérisent par un $D(\hat{\mathbf{x}})$ très faible. En second lieu, les points se

trouvant sur des contours sont éliminés. Un point sur un contour se caractérise par une grande courbe par rapport à sa courbure dans la direction orthogonale. La matrice Hessienne permet de calculer ses courbures. En effet, les réponses se trouvant sur des contours sont éliminées en considérant le Hessien de l'image DoG et d'un opérateur proche de celui de Harris.

$$\mathbf{H} = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \quad (2.15)$$

2.2.4 Détecteur SURF

Le détecteur SURF ("Speeded-Up Robust Features"), a été conçu comme une alternative efficace à SIFT, dans le but d'accélérer le processus d'extraction et de description des points.

Le détecteur SIFT se base sur une détection en utilisant une pyramide de différences de gaussiennes. En revanche, SURF utilise la matrice Hessienne pour localiser des points dans l'espace et la fonction de Laplace pour calculer leur échelle avec son propre calcul du gradient. Les points seront donc localisés là où le déterminant de la matrice Hessienne est maximal. Au lieu de s'appuyer sur les dérivées des gaussiennes, les dérivées partielles sont approximées par des fonctions box filters représentées dans la figure 2.11 qui peuvent être efficacement calculées en utilisant les images intégrales [VJ04]¹. En effet calculer les dérivées par les box filters ne requiert que 16 additions. De plus, ils n'ont pas besoin de construire une pyramide de gaussienne car ils peuvent appliquer de grande convolution pour un coût de calcul constant.

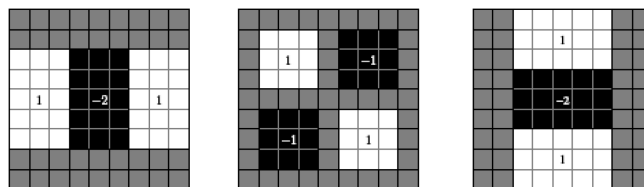


FIGURE 2.6 – Box filter utilisé pour approximer les dérivées D_{xx} , D_{xy} et D_{yy}

2.2.5 Détecteur FAST

Features from Accelerated Segment Test (FAST) [Ros06], est un algorithme de détection de point conçu pour les applications temps réel. L'algorithme opère en deux étapes : dans la première étape, un test de segment basé sur les luminosités relatives est appliqué à chaque pixel de l'image traitée. Un point est considéré comme coin (point d'intérêt) lorsqu'un nombre N de points voisins situés sur un cercle sont plus clairs ou plus foncés que le point central. Les auteurs proposent d'utiliser un cercle discret de 16 pixels et de rayon 3 pixels (cf. Fig. 2.7). Le nombre N est posé à 12. Pour optimiser le temps de calcul les tests sont hiérarchisés de façon que les pixels voisins sont considérés dans l'ordre nord, sud, est, ouest (respectivement pixel 1, 9, 5, 13 dans la figure 2.7) dont trois au moins doivent être clairs pour que le point soit considéré comme un coin. Si ce test n'est pas satisfait, les pixels

1. Une image intégrale (ou summed area table) est une représentation sous la forme d'une image numérique, permettant de calculer rapidement des sommes de valeurs dans des régions rectangulaires en seulement 4 additions.

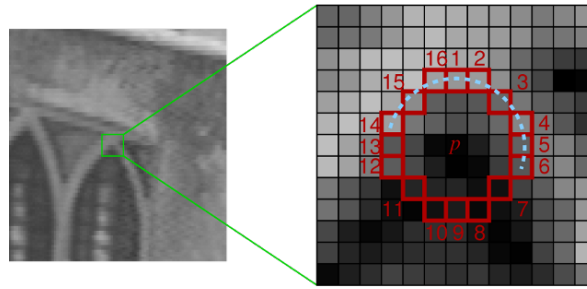


FIGURE 2.7 – Test de segment appliqué sur un rayon de 3 pixels autour du pixel p de l'image.

voisins restants ne sont pas vérifiés. La deuxième étape permet d'affiner et de limiter les multiples réponses du détecteur pour un même coin en supprimant les valeurs non maximales du score V à partir d'un seuil t :

$$V(p) = \max\left(\sum_{p_1 \in E_{\text{clair}}} |I(p) - I(p_1)| - t, \sum_{p_1 \in E_{\text{foncé}}} |I(p) - I(p_1)| - t\right) \quad (2.16)$$

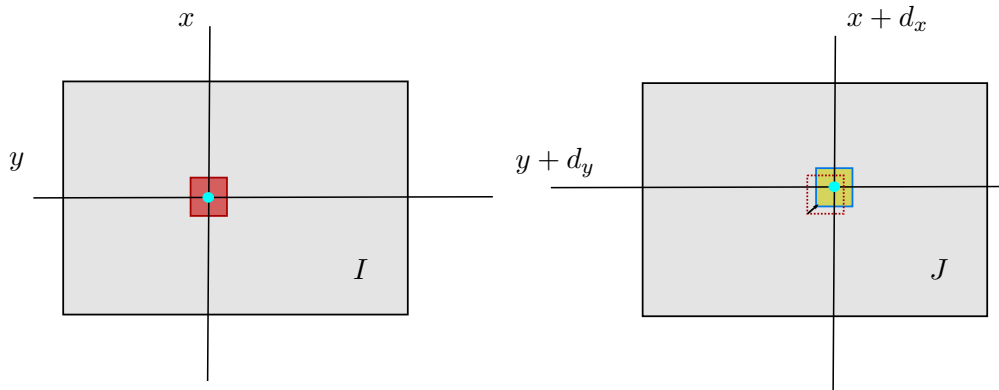
Où $(E_{\text{clair}}, E_{\text{foncé}})$ correspondent respectivement à l'ensemble des points plus claires (plus foncé) que le point candidat. Et $I(p_1)$ correspond à l'intensité du pixel candidat. Une version améliorée du détecteur FAST proposé par le même auteur [RPD10] remplace les tests statistiques par un arbre de décision construit à partir d'un apprentissage sur un ensemble d'image. Ce détecteur offre de bon résultats pour un temps de calcul très rapide. Néanmoins le détecteur FAST est très sensible au flou et au changement d'échelle. L'extension à l'invariance à l'échelle est assurée par le détecteur BRISK [LCS11], qui combine l'espace échelle et l'approche FAST.

2.3 Suivi de points d'intérêt

Après la détection de points vient ensuite l'étape de l'appariement ou du suivi de points, deux stratégies peuvent être employées : soit l'approche récursive, qui consiste à estimer le déplacement des points détectés à partir d'une prédiction du mouvement, soit l'approche descripteur, qui consiste à appairer les points détectés sur chaque images en utilisant leurs descripteurs.

2.3.1 Approche récursive

Les approches récursives sont régies par les techniques basées sur le calcul du flux optique, qui consiste à prédire le mouvement des objets à partir d'un champ de vecteur de vitesse estimé par la projection de l'objet 3D sur la scène 2D (image). Ces algorithmes se basent sur une suite de deux ou plusieurs images et repose sur l'hypothèse de la conservation de l'intensité lumineuse du point en mouvement et sur sa continuité spatiale du mouvement. Leur intérêt majeure repose sur un temps de calcul rapide.

FIGURE 2.8 – Calcul de la SSD sur une fenêtre prédéfinie pour un point au coordonnées (x, y) .

2.3.1.1 Estimation du flux optique par les méthodes de corrélation de fenêtres

L'objectif du flux optique est d'estimer le vecteur déplacement d'un point p d'une image à une autre avec l'hypothèse que la valeur de l'intensité lumineuse ne varie pas ou peu pour le point. Cette condition est implicite dans les méthodes de suivi par corrélation de fenêtres avec l'hypothèse que la zone d'intérêt dans la première image est présente aussi dans la deuxième image. Soit I et J deux images en niveau de gris. Les quantités $I(p) = (x, y)$ et $J(p) = (x, y)$ représentent la valeur en niveau de gris du point $p = [x, y]^T$ où x et y sont les coordonnées en pixel de l'image (la coordonnée du coin gauche supérieure correspond donc à $[0, 0]^T$). L'image I sera référencée comme première image et J la suivante. Soient, n_x et n_y la largeur et la hauteur de l'image, donc la coordonnée du coin droit inférieur correspond à $[n_x - 1, n_y - 1]^T$. Soit un point $u = [u_x, u_y]^T$ dans l'image I . Le but est de trouver la coordonnée du point $v = u + d = [u_x + d_x, u_y + d_y]^T$ tel que $I(u)$ et $J(v)$ sont similaires. Le vecteur $d = [d_x, d_y]^T$ correspond à la vitesse de l'image au point p , aussi connu sous flux optique au point p . Soient w_x et w_y deux entiers qui déterminent la taille de la fenêtre, on définit la vitesse d au point u_x, u_y de l'image comme un vecteur minimisant la fonction résiduel ci-dessous :

$$F(d) = \sum_{x=u_x-w_x}^{u_x+w_x} \sum_{y=u_y-w_y}^{u_y+w_y} (I(x, y) - J(x + d_x, y + d_y))^2 \quad (2.17)$$

Cette fonction aussi appelée la somme des différences au carrée (SSD) de deux blocs, voir figure 2.8. Les différents algorithmes existant [HS81, LK81, Adi85, BFB94, BBPW04, SES12], varient sur les méthodes de minimisation de la fonction 2.17.

2.3.1.2 Algorithme de Lucas kanade

L'algorithme de Lucas-Kanade [LK81, Tom91, BM04] et l'un des plus connus est utilisé pour le suivi de points qui se base sur du flux optique. Lucas-Kanade suppose que le déplacement d'un point de l'image entre deux instants consécutifs est petit et approximativement constant dans un voisinage du point p . L'équation du flux optique peut être supposée vraie pour tous les pixels dans une fenêtre centrée au point p et le but est de trouver le déplacement qui minimise la SSD (Eq. 2.17) entre les deux images. Le calcul de la dérive par rapport au déplacement ainsi que le développement de Taylor

à l'ordre de 1 de la fonction SSD nous donne l'équation suivante :

$$\frac{\partial F(d)}{\partial d} \approx -2 \sum_{x=u_x-w_x}^{u_x+w_x} \sum_{y=u_y-w_y}^{u_y+w_y} (I(x, y) - J(x, y) - [\frac{\partial J}{\partial x} \frac{\partial J}{\partial y}]d) [\frac{\partial J}{\partial x} \frac{\partial J}{\partial y}] \quad (2.18)$$

L'expression $I(x, y) - J(x, y)$ représente la dérivée de l'image dans le temps au point (x, y) , qu'on notera $\delta I(x, y)$. Les dérivées partielles de J en x et en y représente le gradient de l'image. De plus en considérant les petits mouvements (fenêtre de recherche étant plus grande par rapport au vecteur déplacement), la dérivée de l'image dans la fenêtre sur l'image J sera la même que sur l'image I . Pour tout (x, y) on obtient le gradient suivant :

$$I_x(x, y) = \frac{\partial I(x, y)}{\partial x} = \frac{I(x+1, y) - I(x-1, y)}{2} \quad (2.19)$$

$$I_y(x, y) = \frac{\partial I(x, y)}{\partial y} = \frac{I(x, y+1) - I(x, y-1)}{2} \quad (2.20)$$

En remplaçant les expressions ci-dessus dans l'équation 2.18 on obtient alors :

$$\frac{1}{2} \frac{\partial F(d)}{\partial d} \approx \sum_{x=u_x-w_x}^{u_x+w_x} \sum_{y=u_y-w_y}^{u_y+w_y} ([I_x I_y]d - \delta I) [I_x I_y] \quad (2.21)$$

Une simplification de l'équation 2.21.

$$\frac{1}{2} \left[\frac{\partial F(d)}{\partial d} \right]^T \approx Gd - b \quad (2.22)$$

Avec :

$$G = \sum_{x=u_x-w_x}^{u_x+w_x} \sum_{y=u_y-w_y}^{u_y+w_y} \begin{pmatrix} \delta I_x^2 & I_{xy} \\ I_{xy} & I_y^2 \end{pmatrix} \quad (2.23)$$

$$b = \sum_{x=u_x-w_x}^{u_x+w_x} \sum_{y=u_y-w_y}^{u_y+w_y} \begin{pmatrix} \partial \delta I I_x \\ \partial \delta I I_y \end{pmatrix} \quad (2.24)$$

L'objectif est d'annuler la dérivée de l'erreur afin de la minimiser pour trouver le vecteur déplacement.

$$d = G^{-1}b \quad (2.25)$$

Le calcul précédent représente un unique passage de l'algorithme or le résultat final est affiné en effectuant plusieurs passages, jusqu'à ce que la norme du déplacement soit inférieure à une valeur seuil défini par l'utilisateur. De plus afin d'améliorer l'aspect de l'estimation dans le cas d'un grand déplacement, une approche pyramidale à l'implémentation de Lucas-Kanade fut proposé [Bou01, Bak04], qui consiste à estimer le mouvement à un niveau de résolution plus bas (I^N), qui est ensuite adapté à un niveau de résolution plus haut (I^{N-1}), puis ré-estimées à ce niveau (cf. Fig. 2.9). Une amélioration récente de l'algorithme pyramidale de Lucas-Kanade proposée par Tobias, et al. [SES12] consiste à adapter la taille de la fenêtre de recherche pour augmenter la robustesse du suivi. Ils rapportent un gain en robustesse pour un temps de calcul légèrement supérieur à l'algorithme de Lucas-Kanade. En effet le calcul du flux optique repose sur l'estimation du déplacement qui minimise la SSD (Eq. 2.17) qui elle même repose sur une fenêtre de recherche centrée autour du point. Adapter la taille de fenêtre permet donc d'affiner l'estimation du déplacement.

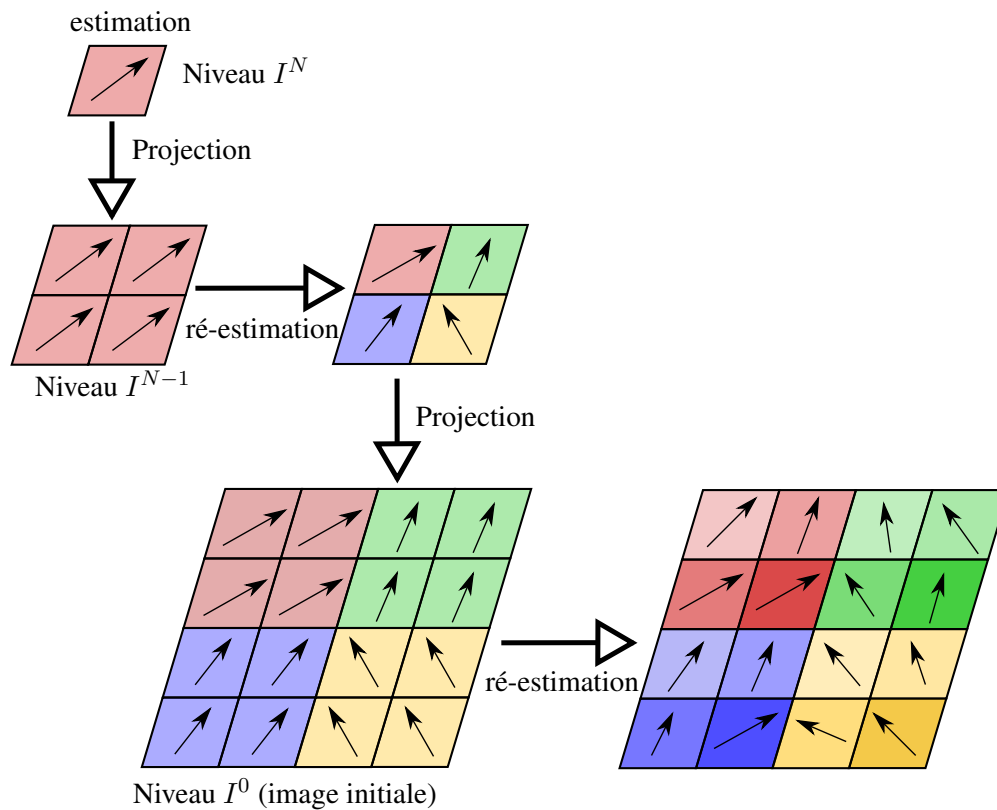


FIGURE 2.9 – Pyramide multi-résolution pour l'estimation du mouvement avec l'algorithme de Lucas-Kanade.

Limites Dans l'équation 2.25 G n'est inversible que si le gradient de I est défini, c'est-à-dire que le contraste autour du point dans la fenêtre de recherche doit être suffisamment élevée. Néanmoins un contraste élevé peut être aussi observé dans le cas où un point est défini sur un contour et dans ce cas l'estimation du mouvement peut être fautive car le point peut glisser sur le long du contour. Ce cas appelé problème d'ouverture, est dû au fait que la composante normale du déplacement est orientée dans la direction du gradient spatial de l'intensité au point considéré. Le problème d'ouverture est illustré dans la figure 2.10. Il n'est pas possible de déterminer le mouvement du point p_1 car le point peut se mouvoir dans la direction verticale supérieure (comme en réalité) ou dans la direction normale au bord du point. Alors que le mouvement du point p_2 peut être déterminé correctement, parce que le gradient d'intensité dans la fenêtre de recherche a deux composantes normales entre elles. Par conséquent, pour éviter le problème d'ouverture, il est nécessaire de suivre des points qui contiennent suffisamment d'informations représentées par des variations locales du gradient (niveaux de gris). Dans [Tom91] ils montrent que l'algorithme de Lucas-Kanade est plus efficace pour suivre des points d'intérêt (coins) dans l'image. En effet, un point d'intérêt correspond généralement à un point avec un contraste élevé dans au moins deux directions.

Par ailleurs, du fait de l'estimation récursive dans le suivi de points, ce type d'approche souffre généralement du problème de dérive de points au cours du suivi (éloignement de la position réelle du point déterminé lors de la détection). De plus, du fait que l'estimation du mouvement se fait à partir d'une projection 2D, le croisement ou la disparition des objets (occlusions) dans la scène, cause un

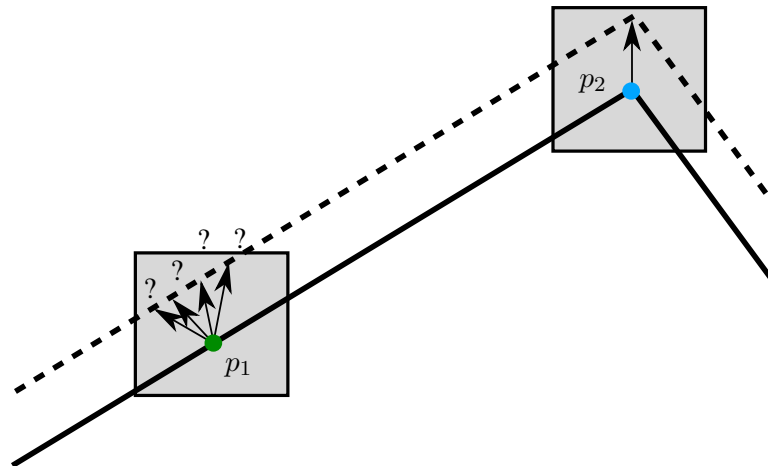


FIGURE 2.10 – Illustration du problème d'ouverture. En l'absence de contraintes supplémentaires le point p_1 peut induire plusieurs estimations. Le mouvement du point p_2 est déterminé correctement pour raison que le gradient est défini avec deux composantes normales.

échec de suivi dans l'algorithme de Lucas-Kanade. Les occlusions rendent difficile l'estimation de mouvement apparent dans la séquence vidéo. En effet, les techniques d'estimation de mouvement entre deux images prises à deux instants t et $t + 1$ d'une séquence vidéo sont basées sur les méthodes de corrélation (ressemblances) entre ces deux images. Si des zones apparaissent ou disparaissent entre les deux images, les ressemblances diminuent et l'estimation échoue dans ces zones.

2.3.2 Approche descripteur

Les approches basées descripteurs ou basées sur la détection de points, reposent sur un appariement (matching) de points détectés dans au moins deux images en utilisant les descripteurs calculés pour chacun des points détectés. Parce que les descripteurs locaux sont en générales invariants aux rotations, aux changements d'échelles et aux effets photométriques, les approches descripteurs sont en mesure de tolérer des variations raisonnables de point de vue et d'illumination à travers les points de vue de l'objet. Pour guider le lecteur à comprendre le principe du calcul des descripteurs, nous présentons les descripteurs SIFT [Low04] et SURF [Bay08]. Le descripteur SIFT est l'un des plus connus et plus complets. La plupart des approches proposées dans la littérature partent de l'algorithme SIFT [Bay08, RRKB11, CLO⁺11, LCS11, AOV12] et proposent des stratégies pour accélérer l'extraction de région d'intérêt et/ou la description. Par la suite nous décrivons le principe de l'appariement de points entre deux images qui repose sur les descripteurs des points.

2.3.2.1 Descripteur SIFT

La phase de description consiste à assigner une orientation ainsi que le calcul d'un descripteur aux points détectés. L'orientation $\theta(x, y)$ et l'amplitude (sa norme) $m(x, y)$ associées aux points SIFT sont déterminées à partir du voisinage local du point (x_0, y_0) dont le paramètre est le plus proche du facteur d'échelle du point (σ_0). De cette façon, le calcul est invariant à l'échelle. Pour les points

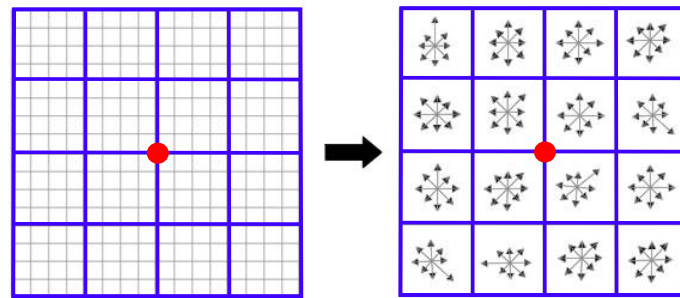


FIGURE 2.11 – Descripteur de SIFT, (a) en bleu région de 4×4 de dimension égale subdivisée en des sous régions de 4×4 pixels, (b) calcul de l'histogramme des orientations selon les 8 directions pour chaque région.

(x, y) retenus, l'orientation $\theta(x, y)$ et sa norme $m(x, y)$ sont calculées à partir de la représentation espace-échelle et à l'aide de la fonction $L(x, y, \sigma_0)$.

$$m(x, y) = \sqrt{\left(L(x+1, y, \sigma_0) - L(x-1, y, \sigma_0)\right)^2 + \left(L(x, y+1, \sigma_0) - L(x, y-1, \sigma_0)\right)^2} \quad (2.26)$$

$$\theta(x, y) = \tan^{-1} \left(\frac{L(x, y+1, \sigma_0) - L(x, y-1, \sigma_0)}{L(x+1, y, \sigma_0) - L(x-1, y, \sigma_0)} \right) \quad (2.27)$$

Du fait que les descripteurs sont calculés relativement par rapport à ces orientations, ceci garantit donc une invariance par rapport à la rotation.

Une fois les points SIFT associés à des facteurs d'échelles et à des orientations, arrive l'étape de calcul des vecteurs descripteurs, traduisant numériquement les caractéristiques visuels de ces points. Les descripteurs sont construits en utilisant le gradient de l'image avec le facteur d'échelle correspondant, estimé au voisinage de chaque point. Un filtre gaussien de paramètre σ est appliqué sur chaque point dont sa valeur correspond au paramètre d'échelle associé au point. De cette manière, la contribution du gradient des pixels dans le voisinage est pondérée.

Lowé propose de considérer autour de ce point une fenêtre centrée autour du point de taille $(2\sigma \times 2\sigma)$ et orienté selon l'angle $\theta(x, y)$. La fenêtre est subdivisée en 4×4 régions de dimension égale. Chaque région est subdivisée elle aussi en des sous régions de 4×4 pixels chacune. Pour chacune de ces sous-régions, est déterminées l'histogramme des orientations. Finalement, les sous-régions sont regroupées pour former 16 régions. À chacune d'elles est associé l'histogramme des orientations (huit directions) calculé à partir de la direction du gradient des sous-régions (cf. Fig. 2.11) pour finalement obtenir le descripteur SIFT à 128 dimensions. Finalement il propose de normaliser le vecteur descripteur dans l'intervalle $[0,1]$ pour garantir l'invariance au changement de contraste. Dans le cas où chaque valeur du pixel de l'image est multiplié par une constante, le gradient sera multiplié par la même constante, de ce fait le changement de contraste sera annulé par le vecteur normalisé. Le calcul du gradient ne sera pas affecté non plus par les changements de luminosités (addition d'une constante à tous les pixels), car il est calculé à partir des différences de pixels (DoG).

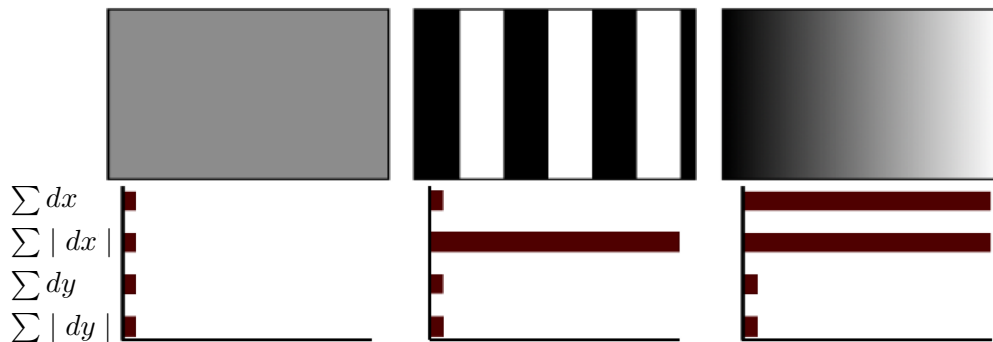


FIGURE 2.12 – Le vecteur intensité du descripteur SURF. A gauche : cas d'une région homogène, toutes les valeurs sont relativement faibles. Au milieu : En présence de fréquences de changement d'intensité dans la direction x , la valeur de $\sum |dx|$ est élevé. A droite : Si l'intensité augmente progressivement dans la direction x , les deux valeurs dx et $\sum |dx|$ sont élevés.

2.3.3 Descripteur SURF

Comme pour le descripteur SIFT, SURF décrit l'intensité des pixels dans un voisinage autour de chaque point. Cependant la réponse du gradient en x et y est calculée en utilisant les "ondelettes de Haar". Il est posé, dx et dy les réponses d'ondelettes selon x et y . Afin de décrire l'information sur les changements d'intensités autour de chaque point, ils extraient la somme des valeurs absolues des réponses, $|dx|$ et $|dy|$. Par conséquent, chaque sous région est décrite par un vecteur d'intensité à 4 dimensions $v = (\sum dx, \sum dy, \sum |dx|, \sum |dy|)$. À partir de ces valeurs, l'orientation dominante de chaque point ainsi que le descripteur est calculé sur une région de 4×4 autour du point pour obtenir un descripteur de 64 dimensions. Malgré cette simplification, SURF assure une robustesse comparable à SIFT tout en offrant un calcul plus rapide [Bay08, JG09]. La figure 2.12 montre les propriétés des descripteurs pour trois différents modèles d'intensités de l'image. Tout comme le descripteur SIFT le vecteur descripteur est normalisé à l'échelle unitaire pour garantir son invariance au changement de contraste. De plus le calcul du gradient par les ondelettes de Haar garantit une invariance à l'illumination dans le cas où une constante est ajoutée à chaque pixel de l'image.

Les plus récents descripteurs (BRIEF [CLSF10], BRISK [LCS11], FREAK [AOV12]) tendent à améliorer le coût en temps de calcul, par des simplifications sur la description des caractéristiques locales des points. Ces descripteurs reposent sur une phase de description des points détectés en utilisant un descripteur binaire [CLSF10]. Des études [Sch, CLSF10, AOV12] ont montré des performances comparables à celle de SIFT et SURF avec un gain en temps de calcul.

2.3.3.1 Principe de l'appariement des points

Soient deux images prises à des instants différents t_1 et t_2 de la même scène. Après une étape de détection et de description des caractéristiques locales des points (Sec. 2.2.3) dans les deux images, il est obtenu deux ensembles S_{t_1} et S_{t_2} . L'idée est de rechercher parmi les sets S_{t_1} et S_{t_2} les points similaires dans les deux images qui possède une signature plus au moins identique de leurs descripteur. Une signature identique entre deux descripteurs est représentée par une distance proche de zéro. Cette distance représente donc une mesure de similarité des caractéristiques locales (intensité, échelle,

orientation) qui sont concaténées dans le vecteur descripteur. En fonction du descripteur utilisé, la valeur physique de la distance varie. Par exemple pour les descripteur SIFT [Low04] et SURF [Bay08] il s'agit de trouver la plus petite distance Euclidienne entre les descripteurs, alors que les plus récents descripteurs (BRIEF [CLSF10], BRISK [LCS11], FREAK [AOV12]), reposant sur des descripteurs binaires, utilisent la distance de Hamming afin d'accélérer le calcul de distance et donc le processus d'appariement de points.

2.3.3.2 Étude des appariements de points

La solution naïve pour identifier les appariements de points est de parcourir tous les descripteurs de l'ensemble S_{t_2} et de comparer leurs distances par rapport aux descripteurs de l'ensemble S_t . En fonction d'un certain seuil les candidats sont appariés. Une telle analyse nécessite un temps de calcul très important. Des algorithmes efficaces reposant sur la méthode des k plus proches voisins (k-nearest neighbor) sont employés pour minimiser le temps de calcul. Dans la figure 2.13, nous illustrons les différents cas de figures observés lors d'un appariement de points en utilisant les approches descripteurs. Pour résumer les différents cas sont les suivants :

- Un point détecté dans les deux images est apparié.
- Un point détecté dans les deux images mais qui n'est pas apparié.
- Un point détecté dans uniquement l'image t_1 ou t_2 qui n'est apparié à aucun point.
- Un point détecté dans l'image t_1 apparié avec un mauvais point dans l'image t_2 et vice versa (appariement ambigu ou outlier).

Limites L'intérêt principal des approches descripteur pour le suivi de points est qu'ils sont robustes aux occlusions du fait de leurs indépendances vis à vis de la continuité temporelle. Néanmoins, elles souffrent du manque de persistance temporelle dans le suivi de points. En effet, la robustesse du suivi repose sur la répétabilité des points détectés d'une image à une autre. Alors que la phase de détection d'une image à une autre renvoie des réponses différentes avec des sets différents de points détectés.

Il est à noter que l'appariement de points en utilisant les méthode descripteurs produisent souvent des appariements ambigus (cf. Fig. 2.14a). Ces appariements causent des erreurs sur l'estimation de la déformation des organes. Une étape nécessaire et fréquente est de suivre le processus d'appariement par des tests de cohérence géométrique et spatiale entre les points d'intérêts, afin de renforcer la robustesse et éliminer les appariements ambigus (cf. Fig. 2.14b). Ces tests reposent essentiellement sur des filtres d'appariement en fonction de la position du point (coordonnées, échelle, rotation).

2.4 Bilan

Dans ce chapitre, notre intérêt s'est porté sur la description des méthodes de détection et de suivi de points. Nous avons décrit les principaux détecteurs ainsi que les deux approches utilisées pour le suivi de points dans les images endoscopiques, qui seront par la suite utilisé dans nos expérimentations. Il est à noter que les applications visant à aider le déroulement du geste chirurgical tels que la réalité augmentée, reconstruction 3D, etc repose sur la dualité d'une détection et d'un suivi robuste et précis. D'une part, il s'agit d'identifier des points de qualité, fiable et facile à identifier, et d'une part il s'agit de suivre et/ou d'apparier ces points par des techniques robustes et de façon précise. Toutefois, un point important quant au suivi de points est celui du temps de calcul nécessaire pour accomplir cette

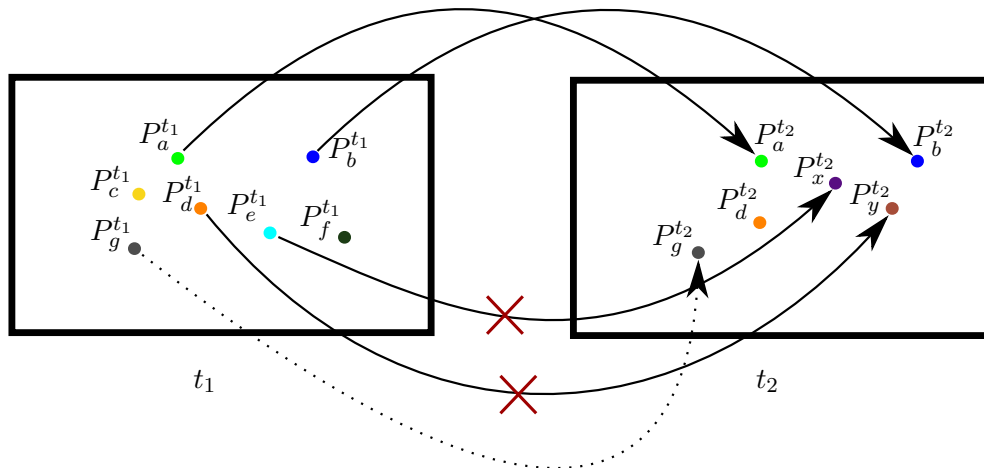
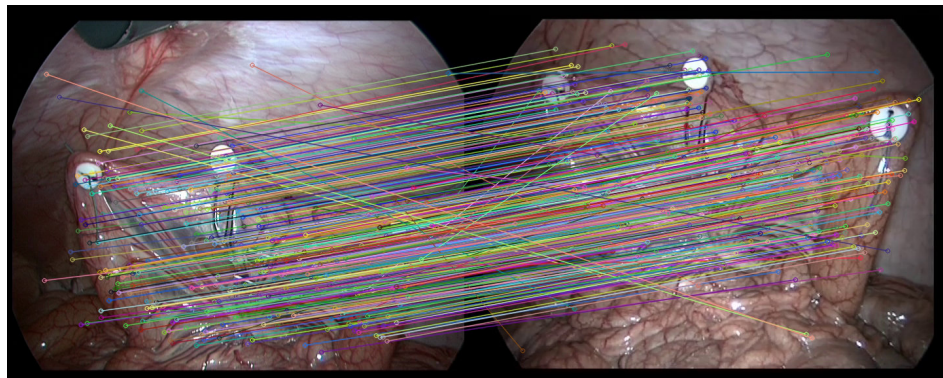


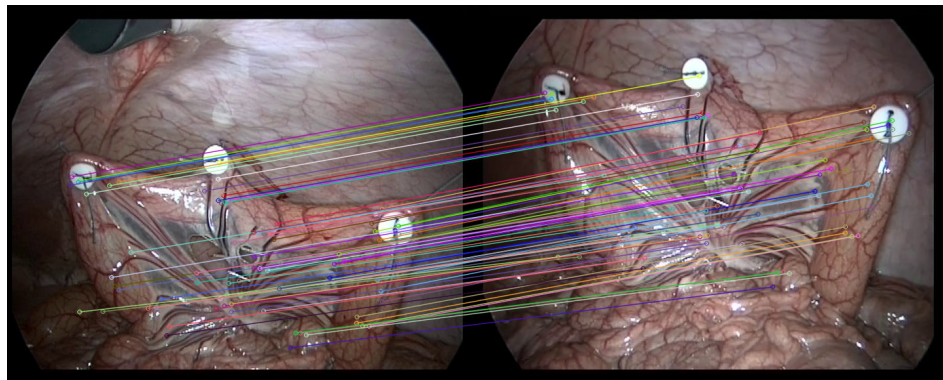
FIGURE 2.13 – Illustration des différents cas de figure lors d'un appariement de points. La paire $(P_a^{t_1}, P_a^{t_2})$ et $(P_b^{t_1}, P_b^{t_2})$ correspondent à un bon appariement. La paire $(P_d^{t_1}, P_y^{t_2})$ et $(P_e^{t_1}, P_x^{t_2})$ correspondent à un mauvais appariement. La paire $(P_g^{t_1}, P_g^{t_2})$ et $(P_d^{t_1}, P_d^{t_2})$ n'est pas appariée alors que ce sont les mêmes points détectés.

tâche. Il est nécessaire dans certaines applications notamment en réalité augmentée de pouvoir garantir un suivi en temps réel, qui n'est pas forcément garanti quand des techniques de filtrages multiples sont nécessaires pour extraire les mauvais appariements.

Ceci relève donc l'importance d'évaluer les performances des méthodes de détection et de suivi. Dans le chapitre suivant nous discuterons des méthodes d'évaluation des performances de suivi points, plus particulièrement dans le cadre des images en chirurgie laparoscopique. Nous verrons les lacunes des méthodes d'évaluations ainsi que les difficultés rencontrées dans ce sujet du fait de l'absence d'une vérité terrain. Aucune stratégie exhaustive et quantitative, permettant l'évaluation des algorithmes de suivi de points dans les images en chirurgie laparoscopique n'a été proposée. C'est pourquoi dans le cadre de cette thèse, nous nous intéressons plus particulièrement au développement d'un framework de validation capable de comparer différents algorithmes en matière de performance de suivi.



(a)

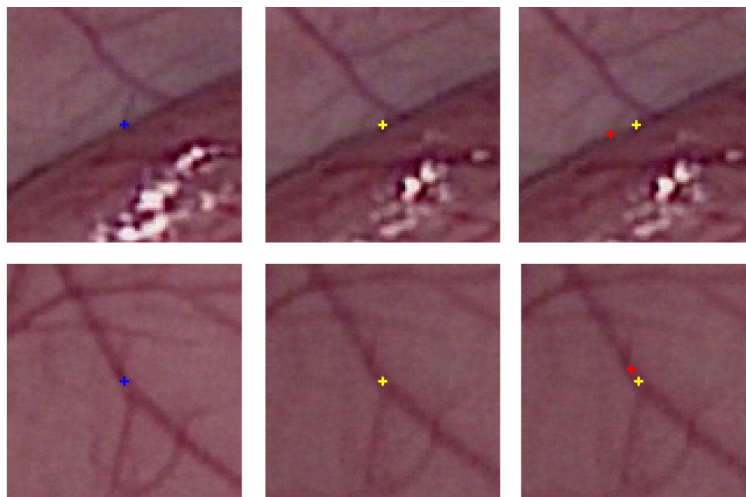


(b)

FIGURE 2.14 – Appariements des points dans le cas des images en chirurgie laparoscopique. (a) Appariements avant filtrage, (b) Appariements après filtrage.

ÉVALUATION DES PERFORMANCES DE SUIVI

Plusieurs applications en vision par ordinateur reposent sur un suivi de points robuste et précis, il est donc primordial d'évaluer les performances de suivi de points d'une manière exhaustive et quantitative. Par contre, l'absence d'une vérité terrain dans le cas des images en chirurgie laparoscopique, rend l'évaluation difficile.



Dans les chapitres précédents, nous avons rappelé l'importance du suivi de points pour les applications en chirurgie laparoscopique ainsi que les fondements de base des différentes méthodes existantes de détection et de suivi les plus utilisées. Nous avons mis en évidence l'importance de l'évaluation des performances de suivi, qui reste une mesure de fiabilité pour les applications proposées aux chirurgiens en laparoscopie. Dans ce chapitre, nous discuterons des techniques utilisées pour évaluer le suivi de points en chirurgie laparoscopique, ainsi que des méthodologies utilisées dans la littérature afin de comparer les différents algorithmes de détection et de suivi. Une analyse mettra en évidence les limitations de ces méthodes et leur biais, rendant difficile une comparaison rigoureuse.

Nous verrons par la suite comment à partir des méthodes de détection des erreurs de suivi nous proposons une technique qui permet d'obtenir une vérité terrain automatique. De plus nous proposons une méthodologie d'évaluation innovante et rigoureuse qui permet de comparer les algorithmes de détection et de suivi d'une manière exhaustive et quantitative.

3.1 Évaluation du suivi

En fonction du type d'approche de suivi (récursive ou basée descripteur) les mesures de performances s'appuient généralement sur la précision et la robustesse des points suivis. Dans les deux cas, l'absence de vérité terrain est un problème majeur pour l'évaluation du suivi dans les images en chirurgie laparoscopique. Pour évaluer la performance de suivi dans le contexte des images en chirurgie laparoscopique, les travaux précédents se basent sur deux approches de validation.

3.1.1 Approches directes

Pour les approches directes, les mesures d'évaluation s'effectuent directement sur les points suivis durant la séquence. La technique la plus classique repose sur une évaluation visuelle du suivi, en utilisant une annotation manuelle (cf. Fig. 3.1) [MY08, AGG⁺09, Yip12, MY11, GVSY12]. Elle consiste généralement à sélectionner des points à partir d'un ensemble de points détectés et de vérifier dans les images à des instants distants, si le suivi est correct. Toutefois, elle est subjective, fastidieuse et doit être répétée pour chaque nouvelle séquence. Dans [GVSY12] ils proposent d'évaluer leur algorithme de suivi prévu pour le suivi des tissus en chirurgie laparoscopique sur des scènes extérieures du monde réel avec une vérité terrain connue. Les points d'intérêts sont reliés par une homographie, car ils sont disposés sur des structures planes. Ces points sont utilisés pour évaluer les appariements. Cependant, ce type d'images est loin de posséder les mêmes caractéristiques que les images en chirurgie laparoscopique. Il est à noter qu'aucune base de données pour la validation du suivi n'est encore disponible pour le cas des images en chirurgie laparoscopique.

Une autre approche proposée par Mountney, et al. [MY11] dans le contexte de la chirurgie cardiaque, consiste à générer une scène virtuelle à partir d'un maillage 3D texturé avec une image endoscopique représentant la surface d'un cœur. Le maillage est déformé pour simuler les mouvements périodiques cardiaques et respiratoires, de plus, un bruit gaussien est rajouté à l'image pour simuler les bruits du capteur de la caméra. Une séquence de 4000 images est obtenue, pour chaque image la position de 100 points d'intérêts est connue. Ils comparent le suivi de ces points par rapport à la position d'origine. Néanmoins les conditions d'illuminations (réfraction de la lumière sur les tissus, occlusions ambiantes...) et de distorsion de la caméra n'ont pas été incluses. En outre, cette évaluation dépend fortement de l'image utilisée pour texturer le maillage.

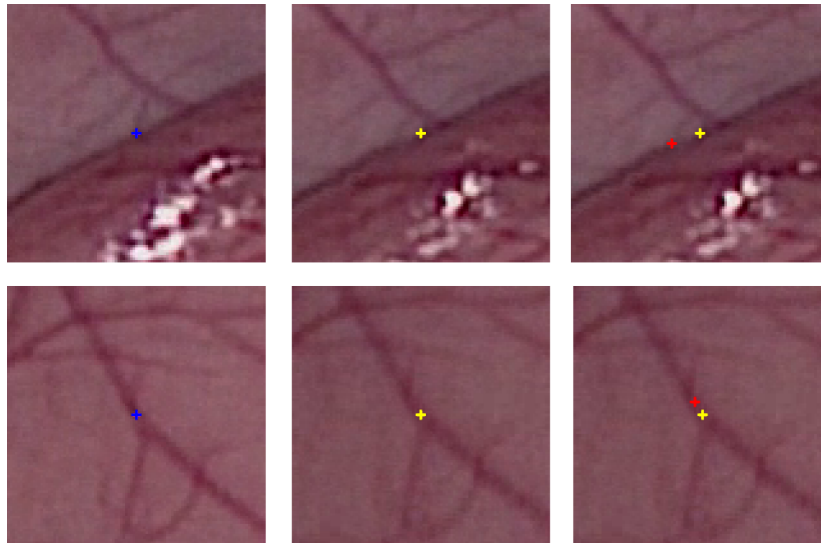


FIGURE 3.1 – Annotation manuelle pour la validation visuelle du suivi de points : à gauche, point détecté à l’image t (croix bleu). Au milieu et à droite, estimation du suivi de point à l’image $t + 1$ (croix jaune). À droite, position vérité terrain du point à l’image $t + 1$ (croix rouge).

3.1.2 Approches indirectes

Pour les approches indirectes, l’évaluation du suivi s’appuie sur des mesures ignorant la réelle position des points. Plusieurs techniques ont été proposées pour évaluer la performance du suivi. Dans l’article proposé par El-hawary [EP11] afin de mesurer la dérive des points suivis au fil du temps, ils calculent la position moyenne des points suivis à chaque période d’une pulsation cardiaque. Un modèle du cœur avec une pompe a été utilisé pour simuler les pulsations cardiaques. En supposant un mouvement périodique du modèle et connaissant le mouvement de la caméra, la position moyenne des points suivis à chaque période de pulsation est calculée et comparée à leur position initiale moyenne connue. La moyenne et l’écart-type de la différence de moyenne des positions à chaque période, indiquent de combien les points suivis ont dérivé de leur position initiale. Cette mesure est spécifique pour le cas où le mouvement de l’objet est supposé périodique.

Dans [Hu 10] afin de générer une vérité terrain, ils utilisent le système AlignRT par VisionRT pour reconstituer une surface partielle du modèle cœur. Le système AlignRT repose sur de la triangulation de points à partir d’une caméra stéréoscopique et d’un projecteur de lumière afin d’améliorer la performance de correspondance. À partir de leur méthode de suivi de points, ils reconstruisent un modèle 3D du modèle, et puis ils comparent l’erreur d’alignement entre les deux modèles. Suivant la même approche Yip, et al. [Yip12] utilisent des repères en métal attachés à la surface de l’organe d’intérêt pour représenter des emplacements dans le tissu qui peuvent être segmentés clairement dans les images. Ces emplacements seront utilisés pour reconstruire un modèle 3D de la scène qui représente la vérité terrain. À partir du suivi de points en ignorant les repères en acier, ils reconstruisent un deuxième modèle 3D de la même scène. Par la suite, ils calculent l’erreur d’alignement entre les deux modèles. Les évaluations ont été réalisées sur des données animales.

Collins, et al.[CPB⁺13] utilisent des marqueurs artificiels pour évaluer le suivi sur des données humaines. Ils marquent la surface de l’utérus avec un instrument de cautérisation pour obtenir un

ensemble de petites régions. Du fait de la cautérisation, ces régions sont fortement texturées ce qui facilite le suivi. Par ailleurs, ils estiment que les points identifiés autour de ces régions suivent la même trajectoire. L'hypothèse repose sur le fait que l'utérus est un organe qui n'est pas affecté par les déformations dues à la respiration ou au rythme cardiaque, ce qui n'est pas vrai pour un grand nombre d'organes (cœur, foie, estomac, rate, ...).

3.2 Détection des erreurs de suivi pour les approches récursives

Pour les algorithmes de suivi récursifs, une simple méthode permet la détection d'un échec de suivi entre 2 images successives dans le cas des mouvements rapides ou des occlusions. Un mouvement rapide ou une occlusion, induit à un changement de la texture entre l'image t et $t+1$ qui se traduit par un changement des niveaux de gris. Bouguet [Bou01] propose de comparer le point suivi entre l'image t et $t+1$ en utilisant la somme de différence des carrés (SSD Eq. 2.17) sur la fenêtre de recherche autour du point. Si ce résidu est au dessus d'un certain seuil fixé par l'utilisateur, le point est déclaré comme étant un point perdu ou non robuste. Néanmoins, cette méthode ne permet pas la détection de la dérive des points au cours du suivi. Un point d'intérêt peut en effet dériver de sa position originale du fait des changements d'illuminations et de contraste dans l'image.

Une méthode récente qui permet de détecter la dérive des points au cours du suivi pour les méthodes récursives a été proposée dans [Kal10]. Elle est basée sur le calcul de l'erreur de suivi avant-arrière de points sélectionnés à partir de l'image initiale. Les points sont suivis en avant dans le temps sur une séquence d'images successives puis la vidéo est jouée dans le sens inverse et les points sont suivis en arrière jusqu'au début de la vidéo. Dans le cas d'une occlusion ou d'une dérive, le point est supposé ne pas revenir à sa position d'origine. Nous soulignons que cette méthode peut également fournir une vérité terrain artificielle pour les méthodes de suivi récursives. En premier lieu, elle permet de connaître la position où chaque point sélectionné devrait arriver sur la dernière image de la séquence avant-arrière (en fait l'image initiale). En effet si un point est parfaitement suivi dans la séquence avant-arrière, il devrait revenir à son emplacement sur la première image. En second lieu, elle permet d'évaluer la précision du suivi en comparant les deux trajectoires avant-arrière.

Cependant, le suivi avant-arrière entre $t \rightarrow t+1$ et $t+1 \rightarrow t$ proposé par Kalal [Kal10] entraîne une sous-estimation de l'erreur du suivi. En effet, les méthodes récursives reposent sur une fenêtre de recherche autour du point, qui est presque la même dans $t \rightarrow t+1$ et $t+1 \rightarrow t$. Ceci entraîne une corrélation de l'estimation du suivi avant-arrière.

Dans ce chapitre, nous proposons tout d'abord une version améliorée du suivi avant-arrière qui permet de pallier au problème de la corrélation susmentionnée et qui permet donc une estimation plus robuste de l'erreur de suivi. L'un des principaux avantages de cette méthode, est de réduire la complexité de l'évaluation des méthodes de suivi récursives ce qui permet de contourner les mesures indirectes et/ou l'annotation manuelle des points sur les images afin de valider les performances de suivi.

3.2.1 Suivi avant-arrière basé sur les images paires-impaires

Notre approche est basée sur l'hypothèse qu'un suivi doit être indépendant de la direction de l'écoulement du temps et que la trajectoire de validation doit être dé-corrélée de la trajectoire du suivi. Cette dé-corrélation est effectuée en prenant les images paires pour le suivi avant et les images

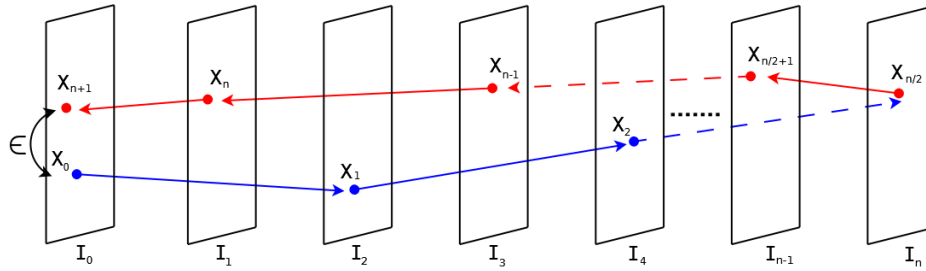


FIGURE 3.2 – Stratégie de suivi basée sur le suivi avant en utilisant les images paires (bleu) et le suivi arrière en utilisant les images impaires (rouge) pour calculer la dérive.

impaires pour le suivi arrière. Soit une séquence de $n + 1$ images $(I_0, I_1, \dots, I_{n-1}, I_n)$ et soit $F = (I_0, I_2, \dots, I_{n-2}, I_n)$ la séquence d'images pour le suivi avant, ce qui correspond à des images paires acquises par une caméra endoscopique. $B = (I_{n-1}, I_{n-3}, \dots, I_3, I_1, \oplus I_0)$ correspondent aux images impaires pour le suivi en arrière. La séquence arrière est donc utilisée comme un moyen de validation des performances de suivi. La figure 3.2 résume notre stratégie du suivi avant-arrière, où X_n correspond à la position d'un point à partir d'un ensemble de points détectés à la n ième image de la séquence avant-arrière.

Cette astuce peut être comparée à un bras robotisé tenant une caméra. Imaginez que nous enregistrons une vidéo avec une trajectoire de A à B avec une vitesse S . L'utilisation du suivi avant-arrière sur des images paires impaires, permet de simuler que le bras robotisé effectue la trajectoire de A à B puis de B à A le tout avec une vitesse $2S$. La figure 3.3 montre le mouvement de la caméra à travers les séquences d'images pour le cas du suivi avant-arrière sur des images successives et dans le cas des images paires impaires. Le suivi avant-arrière nous permet de contourner l'utilisation d'un bras robotisé, qui peut être gênant et coûteux durant les interventions pour la validation des performances de suivi.

Différentes distances peuvent être définies entre les deux trajectoires. Dans notre cadre d'évaluation, nous proposons deux mesures : la première évalue l'erreur finale du suivi et donc la qualité du suivi. La seconde vérifie que le phénomène d'hystérésis ne se produit pas pendant le suivi et donc que les deux trajectoires restent cohérentes.

3.2.1.1 Qualité du suivi

Pour chaque point, nous mesurons la distance euclidienne $\epsilon = \|X_0 - X_{n+1}\|^2$, qui correspond à la différence de position initiale et finale du point dans la séquence avant-arrière. Cette distance représente la dérive du point au cours de l'étape de suivi. Dans le cas d'un suivi parfait, ϵ doit être inférieure à 1 pixel. À partir du calcul de ϵ pour chaque point, nous calculons d'abord le pourcentage de points, où ϵ est inférieur à un certain seuil. Ce pourcentage évalue la robustesse du suivi. Le seuil exprimé en pixels est choisi pour correspondre à la plus petite épaisseur d'un vaisseau sanguin dans l'image (généralement 5 pixels pour les images avec une résolution de 1920×1080).

Cependant, cette mesure n'indique pas la précision des points suivis. En effet, deux algorithmes peuvent fournir le même pourcentage de points robustes, mais, l'histogramme des distances de ϵ peut être plus faible pour l'un des algorithmes. Ainsi, nous proposons également de fournir la moyenne des

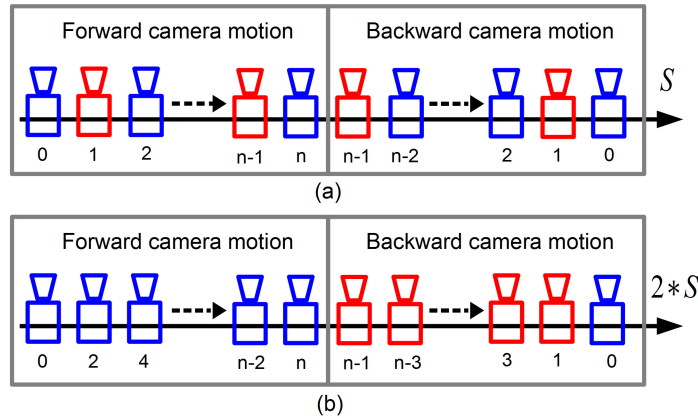


FIGURE 3.3 – Mouvement de la caméra, a) lors d'un suivi avant-arrière d'images successives. b) lors d'un suivi avant-arrière basé sur les images paires-impaires.

distances des points $\bar{\epsilon}$ calculée à partir de l'histogramme des distances ϵ , ce qui indique la précision du suivi.

3.2.1.2 Détection d'hystérésis

Mesurer la qualité finale du suivi n'indique aucunement la cohérence des trajectoires avant et arrière. En effet, dans le cas où ϵ est proche de 0 et que les trajectoires de suivi avant et arrière semblent être différentes, ceci correspond à une hystérésis. Dans le but de détecter ce comportement et d'évaluer la cohérence des trajectoires du suivi avant-arrière basé sur les images paires-impaires nous proposons la stratégie suivante : soit X_{2n} la position du point suivi en avant dans l'image I_{2n} . Soient X_{2n+1} , X_{2n-1} les positions du point suivi en arrière respectivement dans l'image I_{2n+1} et I_{2n-1} . Pour les points robustes dont ϵ est inférieur à un certain seuil, nous calculons la distance d qui correspond à la projection de X_{2n} sur le segment $[X_{2n-1}, X_{2n+1}]$ dans le but de vérifier que les trajectoires avant et arrière passent par une position proche (voir figure3.4).

$$d = \sqrt{\|X_{2n-1}X_{2n}\|^2 - \left(\frac{\overrightarrow{X_{2n-1}X_{2n+1}} \cdot \overrightarrow{X_{2n-1}X_{2n}}}{\|X_{2n-1}X_{2n+1}\|} \right)^2} \quad (3.1)$$

Projeter X_{2n} sur le segment $[X_{2n-1}, X_{2n+1}]$ permet de s'affranchir de l'hypothèse d'une accélération constante entre les 3 images. Nous supposons uniquement que le déplacement ne change pas de direction d'une manière significative entre les 3 images. Si $d = 0$ cela signifie qu'il n'y a pas de changement de direction du point. Soit \bar{d} le ratio des déviations exprimé par : $\bar{d} = \frac{d}{\|X_{2n-1}X_{2n+1}\|}$. Le ratio \bar{d} normalise l'erreur par rapport à la vitesse du point et par conséquent il est indépendant de l'accélération. Nous supposons que la direction du point ne change pas entre 3 images successives si $\bar{d} < 0.5$. Nous soulignons que pour les séquences vidéos utilisées dans notre évaluation, aucun cas d'hystérésis n'a été reporté.

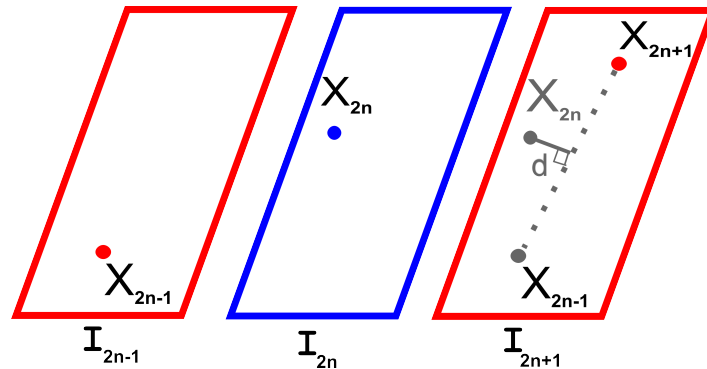


FIGURE 3.4 – Projection et calcul du vecteur d .

3.3 Méthodologie de comparaison des différents algorithmes de détection et de suivi

Comme vu dans le chapitre 2, le suivi de points repose sur des algorithmes de détection et de suivi. Les deux étapes sont conjointement liées pour assurer une estimation robuste et dense de la déformation des tissus. Jusqu'à présent nous avons seulement introduit la partie réalisée pour évaluer le suivi indépendamment de la détection de points.

Pour les approches récursives, le nombre de points détectés à la phase initiale influe sur la robustesse du suivi. Pour un nombre N de points détectés il s'agit idéalement d'avoir le maximum de points robustes dont la dérive est en dessous d'un certain seuil.

Pour les approches basées descripteurs, la robustesse du suivi est influencée par le nombre de points détectés dans la première image et le nombre des appariements corrects entre les deux images.

Le nombre de points à la détection dépend du détecteur et du seuil utilisé. Modifier les seuils des détecteurs permet en effet de modifier le nombre de points détectés. Toutefois il est préférable, d'avoir un plus grand nombre de points détectés et qui sont robustes.

3.3.1 État de l'existant

En général, les travaux courants proposant des méthodes pour le suivi de points en chirurgie laparoscopique [MY11, GVSY12, Yip12] comparent différents algorithmes à leur méthode proposée. La stratégie de comparaison repose principalement sur l'évaluation de la robustesse et de la précision du suivi en fonction du nombre de points détectés. Par ailleurs, les temps de calcul nécessaires pour le suivi de points font aussi partie du processus de l'évaluation. Ces mesures représentent une performance qualitative du suivi du fait qu'ils dépendent généralement des systèmes utilisés (PC et matériels) ainsi que de l'environnement et de l'implémentation du code (Windows/Linux, C/C++/Matlab, CPU/GPU,....).

Mountney, et al. [MY11] évaluent leur méthode de suivi en se basant sur un nombre de points défini à 100. L'évaluation n'a pas été élargie pour un nombre plus important, car leur méthode d'évaluation du suivi repose sur la position vérité terrain de 100 points. Yip, et al. [Yip12] rapportent le pourcentage de points robustes en fonction du nombre de points détectés pour différents types de détecteurs.

Utilisant les paramètres par défaut des détecteurs, le nombre de points détectés étant différent pour chaque détecteur, de ce fait la comparaison de la robustesse des points est biaisée.

Giannarou, et al. [GVSY12] adaptent les seuils des détecteurs afin de détecter le même nombre de points défini à 1600 points. L'évaluation du suivi a été effectuée donc en utilisant le même nombre de points détectés. Néanmoins, une étude plus élargie pour plusieurs nombres de points permet en effet d'évaluer le suivi sur un domaine plus large exploitant les seuils des détecteurs.

À titre d'exemple, soient deux détecteurs (A) et (B) dont les seuils ont été adaptés pour détecter 1000 et ensuite 2000 points chacun. En utilisant le même algorithme de suivi pour les 1000 points, le détecteur (A) fournit plus de points robustes que le détecteur (B). Cependant pour 2000 points la tendance peut s'inverser, et de ce fait le détecteur (B) fournit plus de points robustes que le détecteur (A). Ce cas de figure sera discuté dans la section évaluation. Le choix des valeurs des paramètres de détection doit donc être étudié pour différent nombre de points.

Toutefois, la distribution et la densité des points détectés dans la première image sont également des facteurs importants et doivent être évaluées, avant de mesurer la robustesse et la précision du suivi de points. Une répartition uniforme sur toute l'image permet en effet une récupération globale du tissu et de la cavité abdominale. Si les distributions sont trop différentes entre les détecteurs, la comparaison n'a pas de sens. Par exemple, soient deux détecteurs (A) et (B) dont les seuils ont été adaptés pour détecter le même nombre de points. Les points détectés avec le détecteur (A) peuvent être concentrés dans une petite région de l'image qui sera précisément suivie, alors que les points du détecteur (B) sont répartis dans toute l'image. Jusqu'à présent, il n'existe pas de méthodologie claire et reconnue afin d'évaluer et comparer différentes méthodes. Pour cela nous proposons une méthodologie claire et exhaustive pour évaluer et comparer différents algorithmes de détection et de suivi.

3.3.2 Mise au point de la comparaison des techniques de suivi de points

Dans l'objectif de fournir une évaluation quantitative et exhaustive des méthodes de suivi de points. Nous proposons que les seuils des détecteurs soient systématiquement adaptés afin de détecter le même nombre de points avant de comparer les performances de suivi et ceci pour différents nombres de points. L'évaluation du suivi pour un nombre différent de points permet ainsi d'évaluer la performance du détecteur pour des seuils différents.

La figure 3.5 montre la méthodologie à suivre pour comparer différentes techniques de suivi qui combine deux détecteurs (A, B) et deux algorithmes de suivi (C, D).

3.3.2.1 Distribution

Pour évaluer la distribution de points, l'idée la plus simple revient à comparer visuellement les points détectés. Néanmoins, dans le cas où plusieurs détecteurs sont évalués celle-ci devient fastidieuse. Pour résoudre ce problème, nous proposons l'idée suivante. Nous calculons la transformée de distance d'une image binaire [MJQR03], où les points détectés correspondent à la valeur 1 et le reste à 0 (cf. Fig. 3.6). Le résultat de la transformée est une image dans laquelle les intensités correspondent à la distance entre chaque pixel de l'image et les points détectés les plus proches. La figure 3.7 montre un exemple du calcul de la transformée en distance des points détectés à l'aide du détecteur SURF.

À partir de l'histogramme de l'image résultante de la transformation, nous calculons la moyenne ainsi que l'écart-type des distances. La statistique de la moyenne, correspond à la distance d'un point

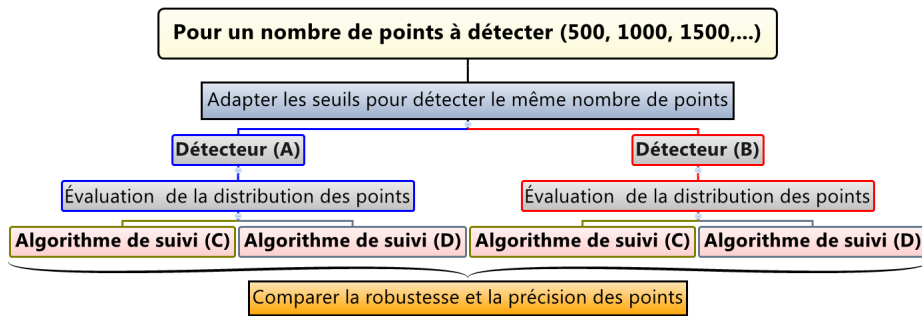


FIGURE 3.5 – Mise au point de la comparaison des techniques de suivi de points.

0	0	0	0	0
0	1	0	0	0
0	0	0	0	1
0	1	0	0	0
0	0	1	0	0

$2^{1/2}$	1	$2^{1/2}$	$5^{1/2}$	2
1	0	1	$2^{1/2}$	1
$2^{1/2}$	1	$2^{1/2}$	1	0
1	0	1	$2^{1/2}$	1
$2^{1/2}$	1	0	1	2

FIGURE 3.6 – Calcul de la transformée en distance d’une image binaire.

choisi aléatoirement de l’image par rapport au point le plus proche du point détecté. Plus la distance moyenne est petite, plus les points sont proches les uns des autres et plus l’écart-type est petit, plus les points sont équidistants dans l’image.

3.4 Évaluations

Dans cette section, nous discuterons de la pertinence de notre méthodologie d’évaluation pour le suivi de points. Premièrement, nous montrons l’influence des seuils des détecteurs sur la qualité du suivi. Deuxièmement, nous relevons l’importance de l’évaluation de la qualité du suivi pour un nombre de points différents et ceci toujours en adaptant les seuils des détecteurs lorsqu’il s’agit de comparer plusieurs algorithmes. Troisièmement, nous montrons l’importance de l’évaluation de la distribution des points avant d’évaluer la qualité du suivi. Finalement, nous montrons que pour les évaluations réalisées sur nos séquences aucun cas d’hystérésis n’a été reporté.

3.4.1 Influence des seuils des détecteurs

Nous proposons d’évaluer quatre différents détecteurs GFTT (G) [Tom94], SURF (S) [Bay08], FAST (F) [Ros06], BRISK (B) [LCS11] pour différents seuils de détection, sur une séquence *in-vivo*

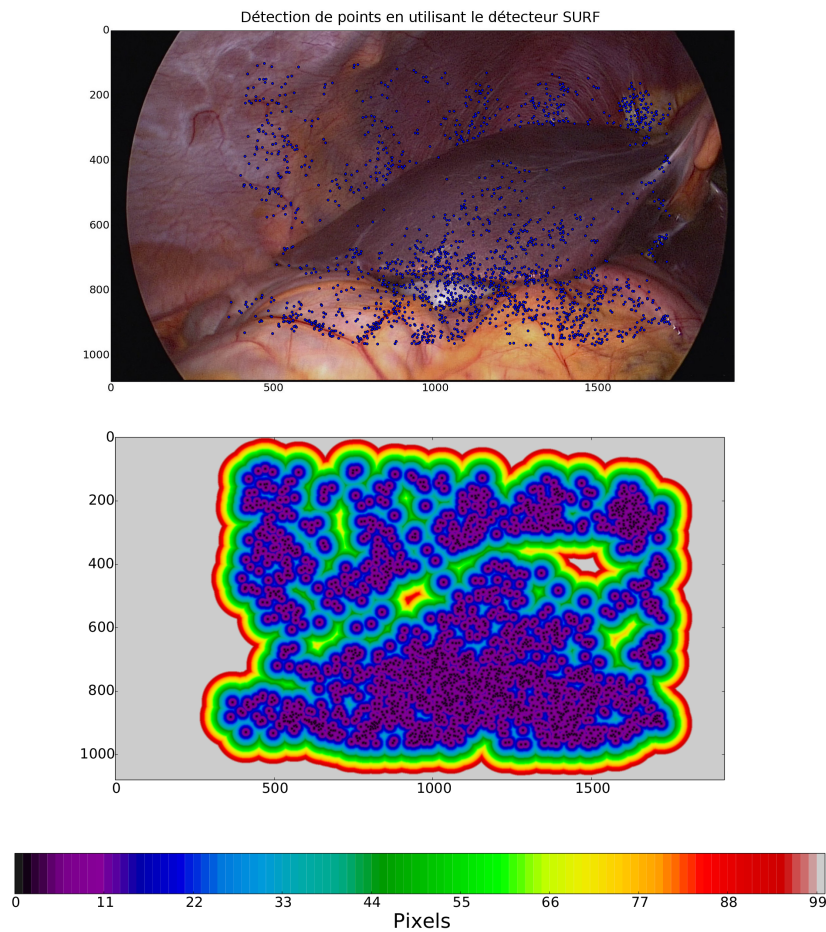


FIGURE 3.7 – La transformée en distance des points détectées à l'aide du détecteur SURF.

exploratrice visualisant les intestins du cochon C.7. Dans la figure 3.8 nous montrons les résultats de la robustesse ainsi que la précision des points détectés. Ces points sont suivis avec l'algorithme de Lucas Kanade (LK) [Bou01] en utilisant notre approche de suivi avant-arrière. Un point est déclaré robuste si il revient à son point d'origine avec une erreur inférieure à 5 pixels. Cet écart correspond à la plus petite épaisseur d'un vaisseau sanguin dans l'image. En fonction des seuils choisis, non seulement le nombre de points à la détection varie, mais, aussi le pourcentage des points robustes. Idéalement, il s'agit d'avoir un grand nombre de points et qui sont en plus robustes. Le fait d'étudier l'influence des seuils des détecteurs, nous permet en effet de trouver le bon compromis entre : le nombre de points détectés et le nombre de points robustes pour chaque détecteur.

Baisser le seuil permet de détecter un plus grand nombre de points, mais, avec des performances de suivi faible, ceci s'explique par le fait que les points détectés sont caractérisés par un faible gradient. Augmenter le seuil revient donc à détecter des points avec un fort gradient assurant un bon suivi. Toutefois, au delà d'une certaine valeur, les points résultants de la détection correspondent le plus souvent à des points spéculaires. Un point spéculaire est caractérisé par un fort gradient (comme pour un bon point d'intérêt), mais qui provoque une erreur de suivi. Par exemple, la figure 3.8a, illustre que lorsque les seuils sont supérieurs à 0.02 pour le détecteur GFTT le nombre de points ainsi que

le pourcentage de points robustes baisse. Cela est dû au fait que plus de 10% des points détectés correspondent à des points spéculaires.

Nous fournissons, dans la figure 3.8e, l'histogramme des distances ϵ pour 500 points détectés avec les différents détecteurs. Par l'analyse des figures 3.8a, 3.8b et 3.8e, il est important de relever que pour 500 points détectés, le détecteur GFTT offre le meilleur pourcentage de points robustes avec 93% suivi de SURF avec 89% (cf. Fig. 3.8a et 3.8b). Néanmoins en terme de précision le détecteur SURF le surpasse avec 196 points qui sont au dessous de 1 pixel d'erreur. L'évaluation de la robustesse dépend d'un certain seuil choisi par l'utilisateur, faire varier ce seuil modifie donc les valeurs de l'évaluation de la robustesse. Il est donc important d'associer l'évaluation de la robustesse à celle de la précision pour fournir une information exhaustive sur la qualité du suivi.

3.4.2 Influence du nombre des points à la détection

Nous proposons d'évaluer les mêmes détecteurs GFTT (G) [Tom94], SURF (S) [Bay08], FAST (F) [Ros06], BRISK (B) [LCS11] sur une séquence *in-vivo* exploratrice visualisant un foie humain (séquence C.3). Les seuils ont été adaptés afin de détecter 500, 1000, 1500, 2000 et 2500 points. Pour chaque nombre, les points ont été suivi en utilisant deux algorithmes de suivi : Lucas Kanade (LK) [Bou01] et le Robust Local Optical Flow (RLOF) [SES12]. Ceci permet de fournir une évaluation des performances des détecteurs et des algorithmes de suivi. Dans la figure 3.9 nous fournissons le pourcentage de points robustes en fonction du nombre de points détectés pour chacun des détecteurs et aussi pour les deux algorithmes de suivi. Nous remarquons que le meilleur algorithme de détection qui fournit le pourcentage le plus élevé de points robustes, peut varier en fonction du nombre de points détectés. Comme illustré dans la figure 3.9a, le détecteur GFTT offre le meilleur pourcentage de points robustes pour 500 et 1000 points détectés alors que pour 2000 et 2500 le détecteur SURF donne le pourcentage le plus élevé. La tendance est neutre pour 1500 points. En ce qui concerne la précision nous fournissons dans le tableau 3.1 et 3.2 la moyenne et l'écart type des erreurs de suivi calculés à partir de l'histogramme des distances ϵ pour les différents détecteurs et les nombres de points détectés, ceci en utilisant respectivement l'algorithme de LK et RLOF. De plus, il est à noter que pour cette séquence l'algorithme de suivi RLOF (cf. Fig. 3.9b) fournis un meilleur pourcentage de points robustes comparé à l'algorithme de LK et ceci pour tous les nombres de points détectés ou des algorithmes de détection. Néanmoins en terme de précision les résultats varient entre l'algorithme de LK et RLOF.

3.4.3 Pertinence de l'évaluation de la distribution spatiale

L'importance de l'évaluation de la distribution spatiale est essentielle dans le cas où nous comparons plusieurs algorithmes de détection et de suivi. En effet, pour un même nombre de points détectés il est nécessaire de comparer leur distribution avant d'évaluer le suivi, car les points peuvent être concentrés dans une région qui sera précisément suivie. Comme illustré dans la figure 3.10 nous montrons l'évaluation de 4 détecteurs GFTT, SURF, FAST, BRISK et de l'algorithme de suivi LK sur une séquence *in-vivo* exploratrice visualisant un foie humain (séquence C.1). Il est à noter que pour 1000 points détectés les détecteurs SURF et GFTT fournissent le même pourcentage de points robustes et une précision quasi identique (cf. Tab. 3.3). Néanmoins, en termes de distribution le détecteur SURF fournit une meilleure distribution des points comparer au détecteur GFTT (cf. Fig. 3.10c et Fig. 3.10b). La moyenne ainsi que l'écart-type des distances calculées à partir de l'histogramme de l'image de la

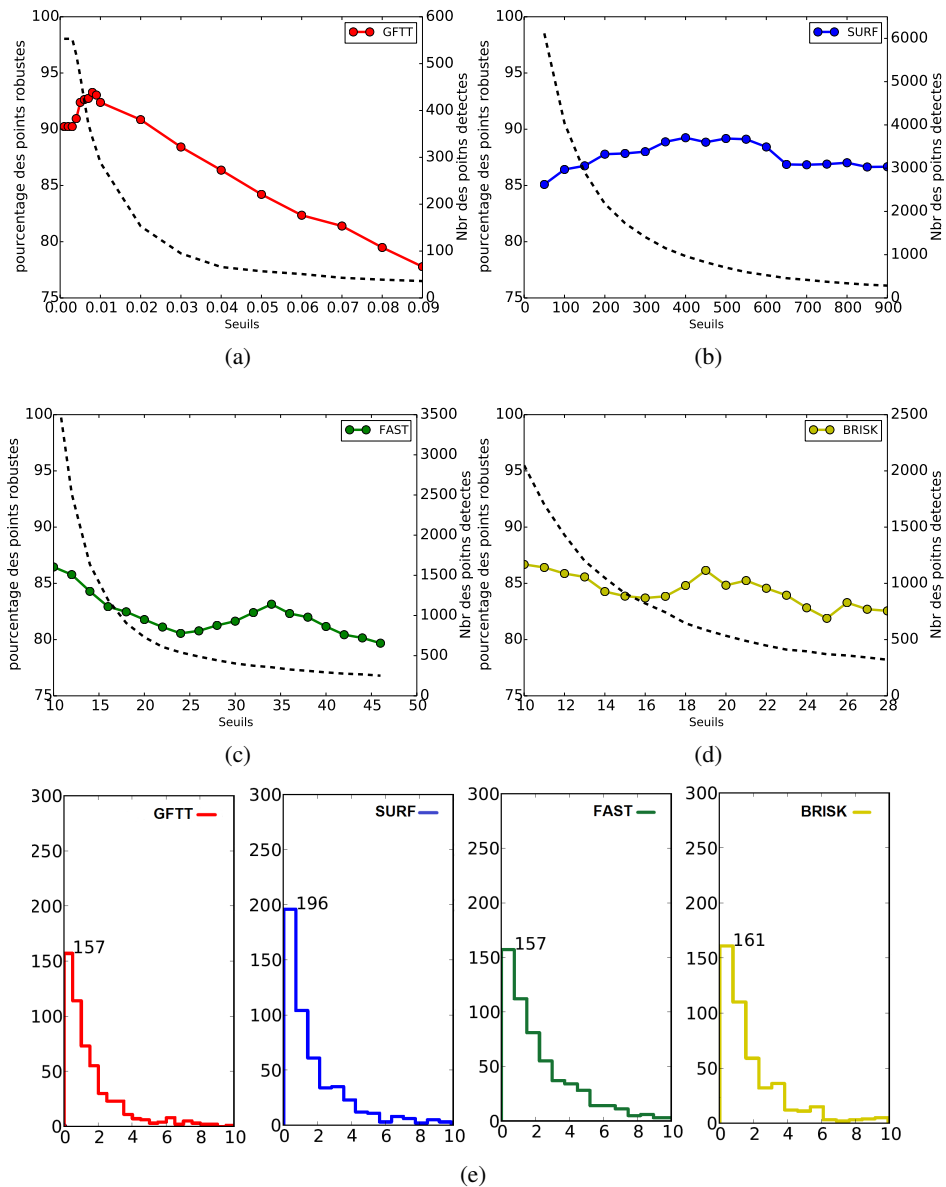


FIGURE 3.8 – Pourcentage des points robustes suivi avec l’algorithme de Lucas Kanade en fonction des détecteurs (a) GFTT, (b) SURF, (c) FAST, (d) BRISK. (e) Histogramme des distances ϵ pour 500 points détectés avec les différents détecteurs.

Nbr de points détectés	500	1000	1500	2000	2500
GFTT	2.41 ± 2.2	2.67 ± 2.23	2.86 ± 2.32	3.08 ± 2.41	3.18 ± 2.42
SURF	2.39 ± 2.21	2.6 ± 2.26	2.7 ± 2.21	2.89 ± 2.31	3.08 ± 2.41
FAST	2.94 ± 2.3	2.94 ± 2.4	3.01 ± 2.45	3.06 ± 2.47	3.13 ± 2.53
BRISK	2.76 ± 2.27	3.08 ± 2.45	3.17 ± 2.47	3.44 ± 2.59	3.49 ± 2.57

TABLE 3.1 – La moyenne et l'écart-type de l'erreur (en pixel) de suivi calculé à partir de l'histogramme des distances ϵ pour les différents nombres de points détectés à l'aide de GFTT, SURF, FAST et BRISK et suivi avec l'algorithme de Lucas Kanade.

Nbr de points détectés	500	1000	1500	2000	2500
GFTT	2.38 ± 2.15	2.63 ± 2.16	2.8 ± 2.28	2.98 ± 2.29	3.13 ± 2.36
SURF	2.7 ± 2.27	2.63 ± 2.14	2.85 ± 2.26	2.99 ± 2.29	3.04 ± 2.29
FAST	3.19 ± 2.31	3.03 ± 2.34	3.12 ± 2.39	3.08 ± 2.35	3.15 ± 2.44
BRISK	2.97 ± 2.28	3.09 ± 2.38	3.18 ± 2.43	3.4 ± 2.42	3.41 ± 2.47

TABLE 3.2 – Moyenne et écart-type de l'erreur de suivi calculé à partir de l'histogramme des distances ϵ pour les différents nombres de points détectés à l'aide de GFTT, SURF, FAST et BRISK et suivi avec l'algorithme Robust Local optical Flow.

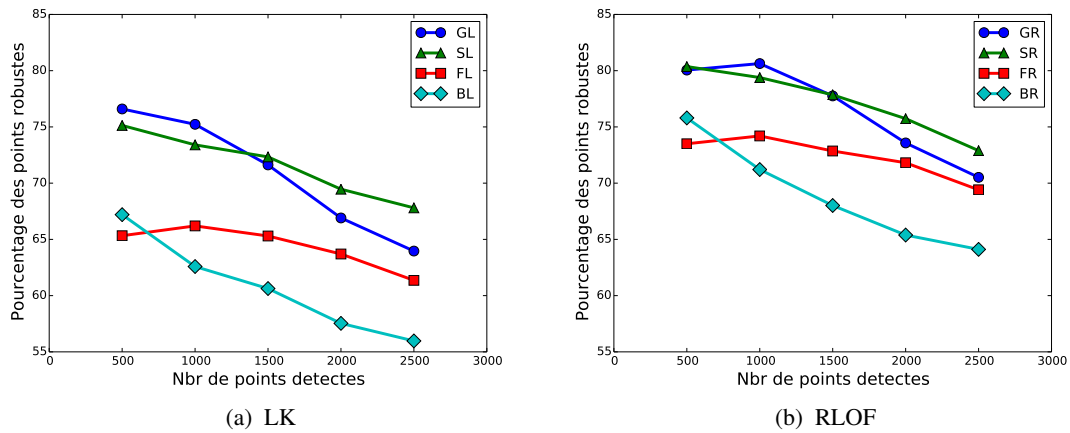


FIGURE 3.9 – Exemple de l'évaluation des détecteurs en utilisant deux algorithmes de suivi. (G) GFTT, (S) SURF, (F) FAST, (B) BRISK, (L) LK, (R) RLOF.

transformée en distance est de 111.8 ± 123.3 pour GFTT et 88.8 ± 95.7 pour SURF (cf. Tab. 3.4). Dans ce cas, le détecteur SURF offre de meilleures performances de suivi (distribution, robustesse et précision).

Dans la figure 3.11 nous soulignons un autre cas où les détecteurs SURF et GFTT offrent les mêmes pourcentages de points robustes. De plus, les distributions sont aussi identiques (cf. Fig. 3.11b et 3.11c), de ce fait l'évaluation de la qualité du suivi à tout son sens. De plus, nous pouvons confirmer que les deux algorithmes offrent les mêmes performances de suivi pour 1000 points détectés et pour la séquence courante.

Nbr de points détectés	500	1000	1500	2000	2500
GFTT	1.18 ± 1.38	1.67±1.79	1.97±2.	2.13±2.07	2.28±2.14
SURF	1.35 ± 1.56	1.66±1.81	2.85±2.26	2.06±1.98	2.27±2.12
FAST	1.68±1.95	2.04±2.17	2.28±2.29	2.42±2.36	2.53±2.39
BRISK	1.91±2.05	2.29±2.28	2.47±2.35	2.57±2.41	2.61±2.41

TABLE 3.3 – Moyenne et écart-type de l’erreur (en pixel) de suivi calculé à partir de l’histogramme des distances ϵ pour les différents nombres de points détectés à l’aide de GFTT, SURF, FAST et BRISK et suivi avec l’algorithme de LK.

Nbr de points détectés	500	1000	1500	2000	2500
GFTT	141.7 ± 128.5	118.8±122.7	111.8±123.3	108.2±123.8	103.1±123.5
SURF	120.4 ± 102.8	96.2 ± 94.5	88.8 ± 95.7	84.3 ± 94.8	80.9 ± 95.1
FAST	130.9 ± 132.6	103.5 ± 106.6	89.5 ± 97.7	84.8 ± 96.2	80.9 ± 94.7
BRISK	106.28 ± 100.2	98.1 ± 104.3	89.1 ± 99.7	84.1 ± 96.1	81.1 ± 96.1

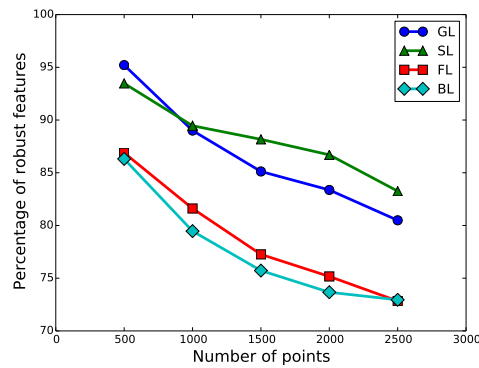
TABLE 3.4 – Moyenne et écart-type de la transformée de distance pour les différents nombres de points détectés à l’aide de GFTT, SURF, FAST et BRISK.

3.4.4 Évaluation du phénomène d’hystérésis

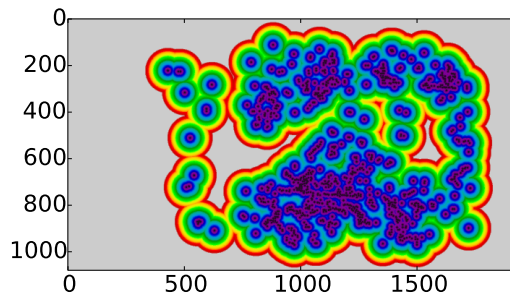
Notre méthodologie d’évaluation repose sur notre stratégie du suivi avant-arrière basé sur les images paires impaires. Celle-ci permet de fournir une vérité terrain artificielle et automatique pour chaque scène. Afin d’évaluer la cohérence de la trajectoire du suivi des points robustes, nous avons calculé le rapport des écarts de \bar{d} (comme expliqué dans la section 3.2.1.2) pour chacun des points sur 3 images successives ($I_{2n-1}, I_{2n}, I_{2n+1}$). La figure 3.12 montre l’histogramme des \bar{d} pour les points robustes d’une séquence exemple. Compte tenu de toutes les séquences utilisées pour nos évaluations, plus de 97% des points robustes sont en dessous de 0, 5. Seuls quelques cas sont au-dessus de 0, 5. Ces cas sont généralement signalés lorsque $\|X_{2n-1} - X_{2n+1}\|$ est de l’ordre de 1 pixel, ce qui se traduit par un faible déplacement de la scène et qui rend la valeur \bar{d} plus sensible. Notre mesure qui permet la détection du phénomène d’hystérésis, ne permet pas d’évaluer l’écart entre les trajectoires lors d’un faible déplacement du point. Nous supposons que les trajectoires restent cohérentes dans ces cas de figure. Finalement, aucun cas d’hystérésis n’a été signalé durant les évaluations pour nos séquences vidéos.

3.4.5 Synthèse

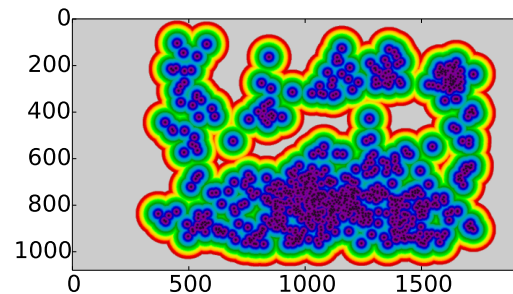
Nous avons montré à travers ce chapitre, l’importance de notre stratégie d’évaluation qui permet de garantir une comparaison exhaustive des algorithmes de détection et de suivi de points d’intérêt. Nous avons évalué les principaux algorithmes de détection et de suivi utilisés pour le suivi de points dans les images en chirurgie laparoscopique. Toutefois, notre stratégie peut aussi bien être utilisée sur des images présentant des prises de vues de scènes extérieures. Notre stratégie d’évaluation repose essentiellement sur le suivi avant-arrière qui permet de fournir une vérité terrain, artificielle et automatique pour chaque scène, et ceci sans utiliser de techniques indirectes ou de matériels supplémentaires. Nous avons montré qu’aucun hystérésis ne se produit durant nos évaluation, confirmant ainsi que les deux trajectoires restent cohérentes.



(a)



(b) GFTT

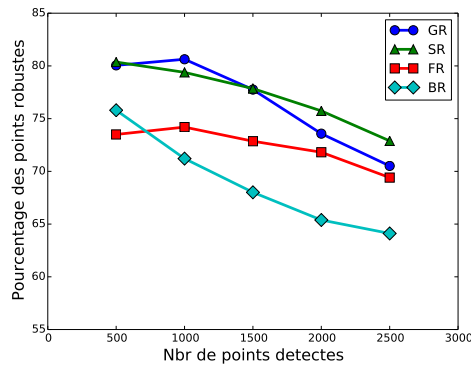


(c) SURF

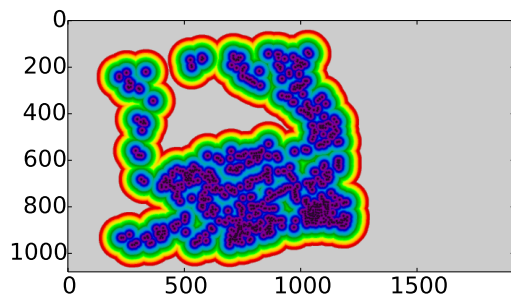
FIGURE 3.10 – Exemple de l'évaluation des détecteurs. (G) GFTT, (S) SURF, (F) FAST, (B) BRISK, (L) LK. (b, c) évaluation spatiale des points détectés avec les détecteurs GFTT et SURF respectivement.

Par ailleurs, changer le seuil pour détecter un plus grand nombre de points se traduit généralement par une baisse dans la qualité du suivi (robustesse et précision). C'est pourquoi, il est nécessaire d'adapter le seuil du détecteur afin de permettre une détection de points, qui seront par la suite suivis d'une façon robuste et précise. Ceci peut être facilement évalué grâce à notre stratégie du suivi avant-arrière qui permet de fournir une vérité terrain automatique. Nous avons noté que, baisser le seuil de détection, permet en effet de détecter plus de points, mais avec des performances de suivi faible. Ceci est principalement dû au fait que le gradient est plus faible autour de ces régions.

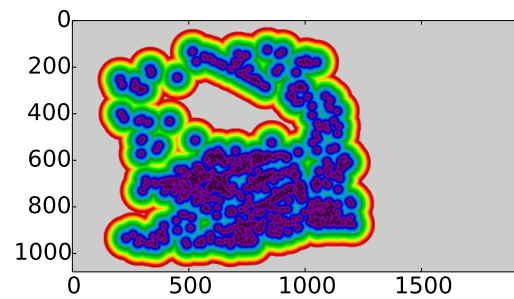
Nous verrons dans le chapitre suivant comment grâce à des techniques basées sur le filtrage d'image, nous améliorons les performances de suivi. De plus, nous décrivons une méthode de détection et de suivi basée sur les 3 canaux RGB de l'image dont le but est aussi d'améliorer les performances de suivi.



(a)



(b) GFTT



(c) SURF

FIGURE 3.11 – Exemple de l'évaluation de la distribution spatiale des points détectés par : (G) GFTT, (S) SURF, (F) FAST, (B) BRISK, (R) RLOF. Dans ce cas, la distribution entre les détecteurs GFTT et SURF est presque identique, de ce fait l'évaluation de la qualité du suivi à tout son sens.

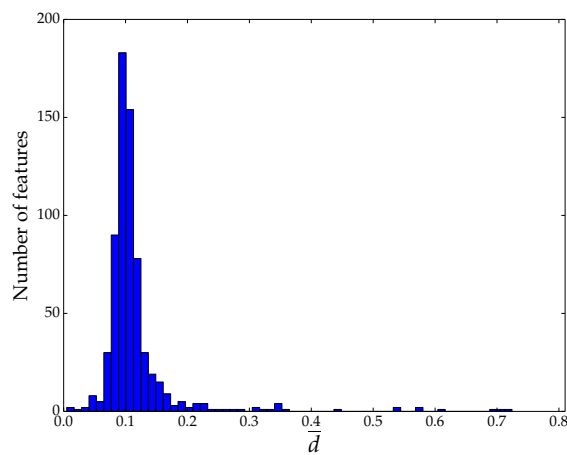
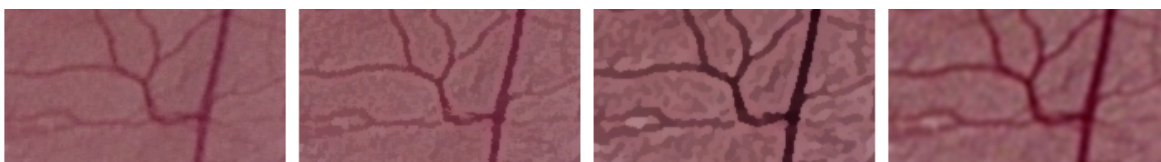


FIGURE 3.12 – L'histogramme du ratio des écarts \bar{d} en fonction du nombre de points.

AMÉLIORATION DES PERFORMANCES DE SUIVI

Plusieurs travaux sur le suivi de points dans les images en chirurgie laparoscopique ont été réalisés pour améliorer les performances de suivi. Ces travaux reposent généralement sur des techniques d'apprentissage ou de filtrage de points non robustes. Toutefois, ce type d'approche implique un temps de calcul supplémentaire qui rend le suivi en temps réel difficile. Nous proposons des techniques de pré-traitement en temps réel pour améliorer le contraste.



Le suivi de points d'intérêt est un problème difficile dans le cadre des images en chirurgie laparoscopique. Ceci est dû aux propriétés des images endoscopiques qui peuvent être mal contrastées, mal illuminées, à la présence de bruits et des réflexions spéculaires. Il est aussi nécessaire de prendre en compte l'occlusion provoquée par les instruments chirurgicaux, les artefacts d'image et des effets dynamiques comme la fumée de cautérisation ainsi que le saignement. Ces conditions peuvent causer l'échec du suivi de points. De plus, les caractéristiques des tissus tels que la présence des motifs similaires (répétitivité de la texture cf. Fig. 4.1) provoquent souvent des erreurs d'appariements (signature semblable des descripteurs pour des points différents).

À travers ce chapitre, nous discuterons des plus récentes techniques utilisées pour le suivi de points dans les images en chirurgie laparoscopique [MY11, GVSY12, Yip12]. Ces techniques ont mis au point plusieurs stratégies afin de garantir les meilleures performances de suivi. Ils se basent sur du suivi par flux optique (méthodes récursives), et/ou du suivi par détection (méthodes basées descripteurs). Une analyse mettra en évidence les limitations de ces méthodes.

Par la suite, nous présenterons nos travaux réalisés dans la thématique de l'amélioration des performances de suivi de points sur les images endoscopiques en chirurgie laparoscopique. Ces travaux reposent sur des solutions basées sur le traitement et l'exploitation de l'information de l'image dans le but d'améliorer les performances du suivi.

Premièrement, nous verrons que l'utilisation des 3 canaux R, G et B au lieu de l'image en niveau de gris améliore les performances de suivi. Deuxièmement, nous montrons qu'appliquer un pré-traitement en temps réel est important pour augmenter la robustesse et la précision du suivi dans les images endoscopiques. Troisièmement, grâce à notre méthodologie de validation et notre stratégie de suivi décrite dans le chapitre précédent, nous fournissons un mécanisme qui permet d'identifier le meilleur algorithme qui garantit la meilleure performance de suivi à partir d'un ensemble d'algorithmes en fonction de la scène et/ou du tissu observé.

4.1 Le suivi de points dans les images en chirurgie laparoscopique

Mountney, et al.[MY11], proposent un algorithme de suivi qui repose sur une phase d'apprentissage de bons points qui combine les techniques récursives et descripteurs. Ils proposent aussi de combiner les détecteurs Shi-Tomasi [Tom94] et SIFT [Low04] pour augmenter le nombre de points détectés. Les points sont initialement suivis avec l'algorithme de Lucas-Kanade pendant une courte

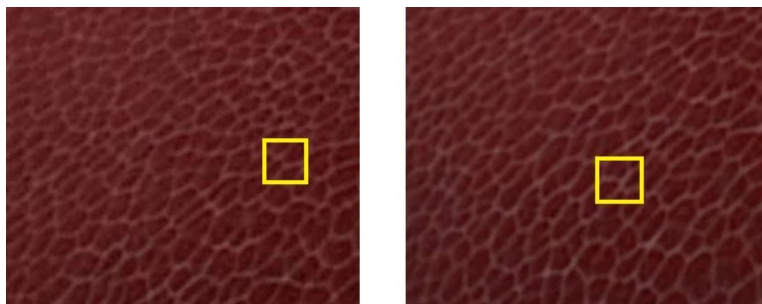


FIGURE 4.1 – Texture d'un foie montrant des motifs répétitifs. Il est difficile de localiser le même point au tour de son voisinage.

période, afin de générer un ensemble de données d'apprentissage qui serviront durant la phase de description des points. En effet, ils construisent un descripteur adapté à la séquence vidéo. Ce type d'approche est particulièrement bien adaptée dans le cas d'une déformation périodique des tissus, du fait que les données d'apprentissage sont générées à partir d'une courte durée qui se reproduit dans le temps.

Yip, et al.[Yip12] proposent de combiner le détecteur STAR [AK08] avec le descripteur BRIEF [CLSF10] dans le but de garantir un suivi en temps réel. Ils développent aussi une stratégie de mise en correspondance qui sauvegarde l'historique des points appariés et qui met à jour la liste des points afin de garantir un suivi sur le long terme. Une phase de test des cohérences géométriques et spatiales des points appariés d'une image à la suivante a été mise en place, afin de garantir un suivi robuste.

Giannarou, et al.[GVSY12] proposent de combiner un détecteur de région (Maximally Stable Extremal Regions, MSER) [MCUP04] avec une approche basée sur le détecteur SIFT [Low04] dans le but de fournir une localisation des points d'intérêt, à la fois sensible à l'échelle et aux transformations affines¹. Ils proposent aussi d'utiliser le filtre de Kalman pour prédire le mouvement, durant la phase d'appariement des points afin de vérifier le cas d'un échec de suivi.

Il est à noter que ces travaux reposent sur des étapes supplémentaires de filtrage et/ou de combinaisons de plusieurs techniques, qui sont complexes et coûteuses en temps de calcul. Dans le but de garantir un suivi robuste et précis du tissu pour limiter cet effet. Les algorithmes de détection et de suivi de points reposent sur l'utilisation d'une image en niveau de gris afin de réduire la complexité du calcul. Mountney, et al.[Mou07] suggèrent que l'information de la couleur ne semble pas apporter une amélioration significative quant au suivi de points. Cependant, aucune évaluation quantitative n'a été fournie. Il est évidemment important de fournir autant de points robustes que possible, afin de minimiser les étapes supplémentaires de filtrage ou de détection des mauvais points pendant le suivi [GCM11, Hu 10].

Améliorer l'identification des points robustes en utilisant des techniques de pré-traitement pour les images endoscopiques n'est pas commun. Dans [Wu04, GVSY12] ils utilisent l'égalisation d'histogramme pour améliorer la détection et le suivi de points. Ils stipulent que l'égalisation d'histogramme permet d'élever l'intensité du gradient pour éviter l'échec de suivi. Dans [Yip12], ils appliquent un lissage gaussien pour réduire la réponse au bruit. Néanmoins, aucun d'eux n'a effectué une évaluation quantitative de l'influence du pré-traitement appliqué sur les performances du suivi.

4.2 Détection et suivi simultané sur les canaux RGB

Les endoscopes actuels utilisés en chirurgie laparoscopique utilisent des capteurs Tri-CCD. Un capteur CCD repose sur le principe de la conversion de l'énergie lumineuse en un signal électrique (effet photoélectrique). Il est constitué de cellules appelées photosites réagissant à la lumière. Chaque photosite est recouvert d'un filtre coloré, ne laissant passer que les rayons d'une certaine longueur d'onde, donnant lieu, sur l'image finale, à un point (ou pixel). Ce filtre appelé filtre de Bayer, est composé d'une mosaïque de 4 filtres : 2 verts, 1 rouge et 1 bleu répartie sur 4 pixels (cf. Fig. 4.2). Une interpolation des couleurs est réalisée par la mesure des pixels adjacents, afin de récupérer les mesures manquantes.

Dans un Tri-CCD, chaque capteur est dédié à une couleur (rouge, vert et bleu) sur le même point de l'image. Afin de sélectionner les longueurs d'ondes projetées dans chaque capteur, un système de

1. En géométrie, une application affine est une transformation géométrique conservant le parallélisme.

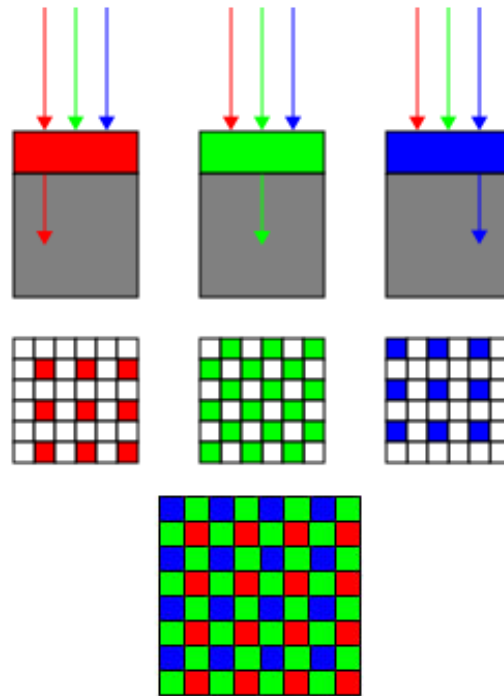


FIGURE 4.2 – Filtre de Bayer, placée devant un capteur photographique numérique afin de pouvoir enregistrer des images en couleur. Il est composé d'une mosaïque de 4 filtres : 2 verts, 1 rouge et 1 bleu répartie sur 4 pixels.

séparation de lumière basé sur un prisme dichroïque² est utilisé. L'image formée offre une meilleure résolution et sensibilité de la couleur (chaque pixel de l'image est représenté par une valeur pure de chaque canal).

Les tissus mous absorbent des longueurs d'ondes différentes [KF10], ce qui influe sur la dynamique de contraste pour chaque canal. Nous avons montré dans le chapitre 2, que la détection d'un point d'intérêt se base généralement sur l'intensité du gradient de celui-ci. L'exploitation de la totalité de l'information provenant de chaque canal (capteur) semble pertinente.

Cependant, les approches standard pour le suivi de points en imagerie laparoscopique, sont réalisées sur les images en niveaux de gris. L'image en niveau de gris (BW) n'est que la combinaison linéaire des 3 signaux (R, G, B), classiquement :

$$BW = 0.299 \times R + 0.587 \times G + 0.114 \times B \quad (4.1)$$

Par une simple analyse, nous mettons en évidence les 2 inconvénients suivants. D'une part, si un point est détecté sur le canal B par exemple, et pas sur les autres, il est alors probable que ce point ne sera pas détecté dans l'image BW, puisque l'intensité de gradient dans l'image BW sera pondéré sur tous les canaux alors que la contribution du canal B est trop faible. Ainsi, un point qui pourrait

2. Un prisme dichroïque est un assemblage complexe de prismes plus simples, qui permet de diviser et séparer un faisceau en trois faisceaux de trois couleurs, généralement rouge, vert et bleu qui compose le faisceau d'entrée.

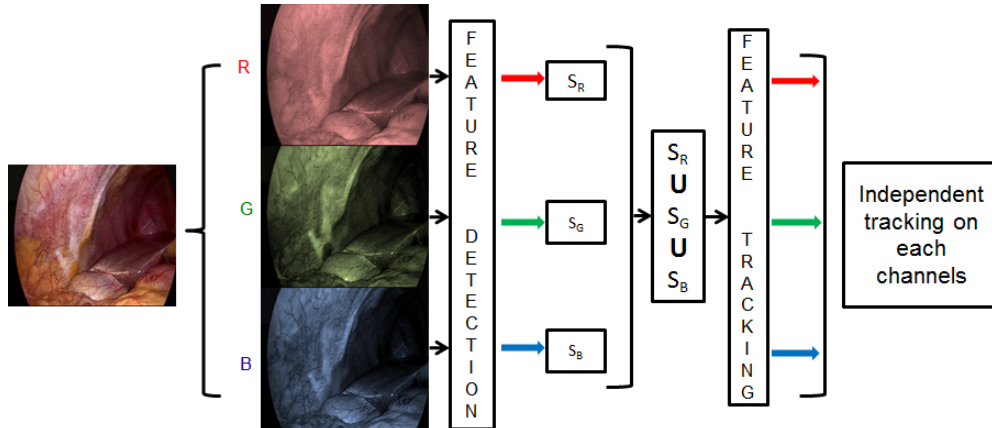


FIGURE 4.3 – Détection de points dans chaque canal et suivi simultané sur les 3 canaux.

être un bon élément à suivre dans un canal indépendant sera perdu. D'autre part, un point détecté dans l'image BW peut être suivi de façon plus efficace dans le canal qui présente le plus fort gradient.

Nous proposons donc [Sel], d'une part, une détection dans chaque canal afin d'augmenter le nombre de points détectés pour permettre un suivi dense de la scène et, d'autre part, un suivi simultané sur les 3 canaux pour augmenter la robustesse du suivi. La figure 4.3 montre un résumé de l'idée proposée, qui est décrite en détail ci-après.

4.2.1 Détection et sélection de points sur les canaux RGB

Soit S_x les coordonnées des points détectés dans le canal x . Premièrement, nous appliquons une détection de points d'intérêt sur chaque canal S_R , S_G , S_B et considérons l'union des trois ensembles $S_R \cup S_G \cup S_B$. Deuxièmement, nous fusionnons les points qui sont très proches. Typiquement, un point $M(x, y)$ et un point $U(u, v)$ sont considérés identiques si $(x + \tau > u > x - \tau)$ et $(y + \tau > v > y - \tau)$. Le nouveau point est la moyenne des deux points. τ est sélectionné en fonction de la résolution de l'image et la taille de la scène observée (foie, intestins, rate,...). Généralement nous utilisons des endoscopes avec une caméra full HD en filmant une scène de taille 30 cm, de ce fait il est raisonnable de choisir $\tau = 5$ ce qui correspond approximativement à 0.7 mm. Finalement nous notons S_{all}^0 le nouveau ensemble obtenu après la fusion des points : chaque point est suivi dans chaque canal simultanément. Il est à noter que les points appartenant à l'ensemble S_{all}^0 ne sont plus reliés à un canal de couleur spécifique (R, G ou B). En fait, nous considérons que chaque point peut être suivi d'une façon robuste dans tous les canaux. Ce choix peut sembler inapproprié, mais nous avons remarqué expérimentalement qu'en raison du changement des conditions d'illuminations et de contraste, un point robuste dans le canal R par exemple, peut devenir un point robuste dans le canal B ou/et G durant la séquence vidéo.

4.2.2 Suivi simultané de points sur les canaux RGB

Soit P^i un point de l'ensemble S_{all}^i à l'image i . Dans cette section nous expliquons notre stratégie pour calculer la position du point P^{i+1} dans l'image $i + 1$ à partir de sa position P^i . Premièrement,

nous estimons le déplacement du point P^i dans l'image $i + 1$ dans chaque canal R, G et B en utilisant un algorithme de suivi, donnant ainsi : $P_R^{i+1}, P_G^{i+1}, P_B^{i+1}$. Idéalement, dans le cas d'un suivi parfait, nous devons observer $P_R^{i+1} \simeq P_G^{i+1} \simeq P_B^{i+1}$.

De plus, puisque nous supposons que le mouvement de l'endoscope reste constant, le mouvement de P^i (ex : $\|P^i P^{i+1}\|$) doit avoir une amplitude proche du mouvement moyen de $P^{i-2}, P^{i-3}, \dots, P^{i-n}$ sur les images précédentes. Mathématiquement, la distance moyenne m^{i+1} des $P_R^{i+1}, P_G^{i+1}, P_B^{i+1}$ à leurs centre de gravité O^{i+1} doit donc être très petite (cf. Fig. 4.4 a), où la distance moyenne m^{i+1} est exprimée comme suit :

$$m^{i+1} = \frac{O^{i+1}P_R^{i+1} + O^{i+1}P_G^{i+1} + O^{i+1}P_B^{i+1}}{3} \quad (4.2)$$

Dans le cas où la distance entre ces 3 estimations est très petite et que le mouvement est cohérent, l'estimation du point P^{i+1} à l'image $i + 1$ est égale à $\frac{P_R^{i+1} + P_G^{i+1} + P_B^{i+1}}{3}$.

En général, le calcul de P^{i+1} dépendra de la dispersion de $P_R^{i+1}, P_G^{i+1}, P_B^{i+1}$ (représentée par m^{i+1}) et leur mouvement d_R^{i+1}, d_G^{i+1} et d_B^{i+1} , où le mouvement du point dans le canal x , est exprimé comme suit :

$$d_x^{i+1} = \|P_x^{i+1} P^i\| \quad (4.3)$$

La première étape consiste à filtrer les points P_x^{i+1} si leurs d_x^{i+1} sont supérieurs à un seuil $\lambda_{\bar{d}}$. Pour la deuxième étape, nous calculons la dispersion m^{i+1} des points restants et nous estimons P^{i+1} en fonction de m^{i+1} . La figure 4.4 a-f décrit les différents cas possibles, le cercle jaune représente le seuil accepté de la dispersion $\lambda_{\bar{m}}$.

Les deux seuils $\lambda_{\bar{d}}, \lambda_{\bar{m}}$ sont définis comme suit. Soit \bar{d} la moyenne des déplacements des points sélectionnés dans une ROI localisée au centre de l'image sur les 20 dernières images et $\sigma_{\bar{d}}$ l'écart-type de ces distances.

$$\bar{d} = \sum_{j=0}^{20} \frac{\|P^{i-j} P^{i+1-j}\|}{20} \quad (4.4)$$

Nous supposons que le suivi sur les 20 images reste cohérent sans aucune erreur et/ou dérive du fait que les conditions d'illumination ne changent pas au centre de l'image. De la même manière, \bar{m} est la moyenne de m^i sur les 20 images précédentes et $\sigma_{\bar{m}}$ son écart-type. Nous choisissons $\lambda_{\bar{d}} = \bar{d} + 3\sigma_{\bar{d}}$ et $\lambda_{\bar{m}} = \bar{m} + 3\sigma_{\bar{m}}$. \bar{d}, \bar{m} ainsi que leurs écart-type sont recalculés toutes les 20 images.

4.2.3 Évaluation de la détection et du suivi simultané sur les canaux RGB

Nous évaluons notre approche sur 4 séquences vidéos FHD (Full High Definition) (1920×1080). Deux d'entre elles contiennent un foie humain, C.1 C.3 et les deux autres contiennent des intestins de cochons C.7 C.8 (voir Annexe.C). Premièrement, nous évaluons le pourcentage de points supplémentaires détectés en utilisant tous les canaux RGB au lieu de l'image en niveau de gris uniquement avec les détecteurs GFTT [Tom94], SURF [Bay08] et SIFT [Low04]. Les seuils des détecteurs ont été choisis pour garantir le meilleur rapport points détectés et pourcentage de points robuste en utilisant l'image en niveau de gris, grâce à notre méthodologie d'évaluation décrite dans le chapitre 3. Deuxièmement, nous montrons que le suivi simultané sur les canaux RGB améliore le pourcentage de points robuste par rapport au cas d'un suivi sur l'image en niveau de gris seulement.

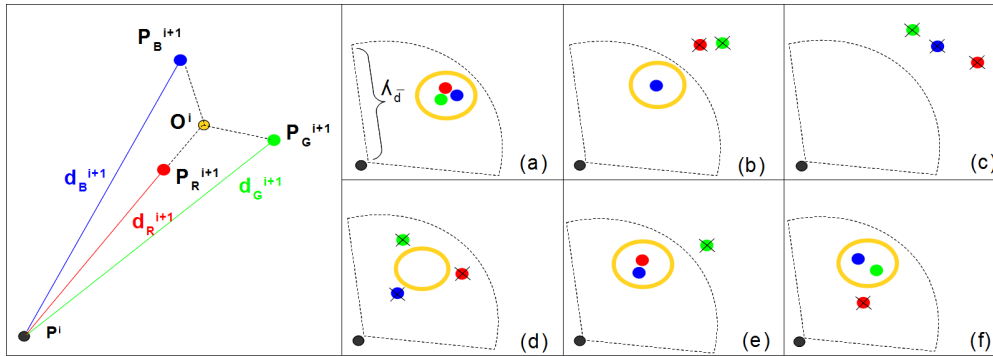


FIGURE 4.4 – Gauche : P_R^{i+1} , P_G^{i+1} , P_B^{i+1} sont les estimation de la position du point P^i à l’image $i+1$, et O^{i+1} leurs centre de gravité. Droite : les 6 figures décrivent les scénarios possibles durant le suivi simultané RGB. Pour chaque cas, nous filtrons l’estimation dans le canal x si d_x^{i+1} est supérieur à $\lambda_{\bar{d}}$ (cas b,c,e). Si m_i est inférieur à $\lambda_{\bar{m}}$, P^{i+1} est la moyenne des points restants (cas a,b,e). Si $m_i > \lambda_{\bar{m}}$ (cas d,f), nous vérifions si la distance entre les 2 points restants est inférieur à $\lambda_{\bar{m}}$ et nous calculons la position moyenne des deux points (cas f). Sinon, le suivi est arrêté pour le point (cas d).

Décteur	GFTT		SIFT		SURF	
	Intestins	Foie	Intestins	Foie	Intestins	Foie
Moyenne	46.63%	44.62%	34.84%	36.50%	42.27%	40.45%

TABLE 4.1 – Moyenne du pourcentage des points supplémentaires détectés en utilisant tous les canaux RGB de l’image au lieu de l’image en niveau de gris seulement.

4.2.3.1 L’analyse RGB simultanée pour augmenter le nombre de points détectés

Pour toutes les évaluations, l’implémentation fournie par la librairie OpenCV avec les paramètres suivants ont été utilisés pour les détecteurs. Pour GFTT, le seuil est défini à 0.02 avec une distance minimale entre les points égale à 25. Pour SIFT le seuil des coins es défini à 0.05 une valeur de 5 pour le seuil des bords. Pour SURF le seuil de la matrice hessienne est défini à 800.

Le tableau 4.1 montre le pourcentage moyen des nouveaux points détectés grâce à notre approche par rapport à une détection sur une image en niveau de gris pour les 4 séquences. Nous notons que la détection sur tous les canaux permet de fournir plus de 40% de nouveaux points pour GFTT et SURF et plus de 30% pour SIFT. La figure 4.5 fournit un exemple de notre approche de détection de points. Il est intéressant de noter que 99% des points détectés dans l’image BW sont également sélectionnés avec notre approche.

4.2.3.2 Évaluation de la robustesse du suivi de points en utilisant les canaux RGB

Afin d’évaluer les performances de notre approche pour le suivi de points par rapport à un suivi limité à l’image en niveau de gris, nous proposons la méthode suivante. Nous détectons les points dans chaque vidéo en utilisant notre approche RGB. L’ensemble des points est suivi simultanément sur tous les canaux RGB et sur l’image en niveau de gris seulement. Nous rappelons que les points

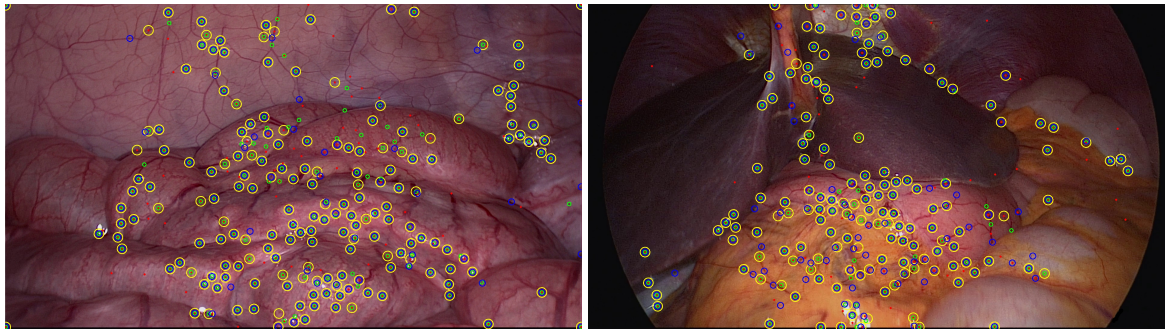


FIGURE 4.5 – Exemple de détection de points à l’aide du détecteur GFTT. Les cercles rouge/vert/bleu/jaune correspondent à une détection dans les canaux R/G/B/BW.

Séquence	Foie		Intestins	
	BW	RGB	BW	RGB
Approche de suivi				
Pourcentage points robustes	72.2%	80.3%	74.6%	89.5%

TABLE 4.2 – Pourcentage de points robustes suivis avec l’image en niveau de gris seulement (BW) et notre approche RGB, pour les deux séquences.

détectés grâce à notre approche RGB, ne sont pas forcément détectés en utilisant l’image en niveau de gris seulement, mais ils sont tout de même suivis dans l’image en niveau de gris afin de comparer le même ensemble de points suivi en utilisant les 3 canaux. L’algorithme de Lucas Kanade (LK) [Bak04] avec les paramètres suivants a été utilisé pour le suivi de points : une fenêtre de recherche de taille (25×25) et une pyramide de niveau 3.

Dans le tableau 4.2 nous fournissons le pourcentage de points robustes suivis avec l’image en niveau de gris seulement (BW) et en utilisant le suivi simultané sur les canaux RGB pour les séquences qui contiennent un foie humain et les séquences contenant les intestins du cochon. Les résultats correspondent à la moyenne des pourcentage de points robustes obtenu avec les 3 détecteurs. Nous notons que grâce à notre approche nous augmentons jusqu’à 15% le nombre de points robuste, et ceci pour différents détecteur. Un point essentiel à relever sur l’analyse RGB, et que le plus souvent, l’utilisation du canal vert seulement, donne de meilleures performances de suivi en comparaison à l’image en niveau de gris.

Par ailleurs, nous remarquons que sur les séquences contenant des intestins de cochon, nous avons obtenu un meilleur résultat pour l’amélioration des performances de suivi, ce qui s’explique par la nature des tissus qui sont plus vifs et claires, et qui se traduit par un contraste élargi dans le canal vert et bleu. Sur les séquences contenant le foie humain la caractéristique du tissu fait que sur les 3 canaux le contraste reste faible. C’est pourquoi l’idée d’ajuster le contraste afin d’optimiser la qualité du suivi nous semble pertinente.

4.3 Pré-traitement pour l'amélioration des performances de suivi

En raison du changement d'illumination (ombres, reflets) et de l'unique source de lumière disposée à l'extrémité de l'endoscope, les zones sombres dans les images endoscopiques peuvent être importantes. En effet, jusqu'à 20% des valeurs des pixels dans le champ de vue de la caméra (FoV) (carré rouge sur l'histogramme de la figure 4.6) ont moins de 1/3 de la luminance maximale de l'image, impliquant un histogramme sous-exposé. C'est pourquoi, il semble nécessaire d'ajuster de manière adaptative le contraste de l'image pour permettre, d'une part, la détection de points dans les zones sombres, et d'autre part, d'augmenter la qualité du suivi pour l'ensemble des points dans l'image afin de limiter les erreurs de suivi. Par ailleurs, du fait que les points d'intérêts dans les images endoscopiques sont le plus souvent identifiés sur des petites structures ou sillons (bifurcation de vaisseau sanguin, texture du foie, rate,...), il semble intéressant d'augmenter le gradient de ces petites structures à l'aide d'approches morphologiques.

Nous avons vu précédemment dans l'analyse RGB, qu'il est important d'utiliser les 3 canaux qui combinent l'image couleur au lieu de l'image en niveau de gris. Augmenter le gradient et adapter le contraste de l'image devront donc être eux aussi réalisés en tenant compte de l'ensemble de l'information de l'image. Plusieurs études portant sur l'amélioration de la dynamique de contraste et de la qualité des images couleurs, ont été proposées dans le domaine de la vision par ordinateur [Han11, Han04, BZX12]. Nous dérivons de ces approches, pour proposer des pré-traitements adaptés à la spécificité des images couleurs en chirurgie laparoscopique.

En premier lieu nous donnons une description détaillée des pré-traitements utilisés. En second lieu, nous évaluons l'influence de ces pré-traitements sur les performances de suivi de points d'intérêt.

4.3.1 Égalisation lissée

Afin d'améliorer le contraste des images, l'égalisation d'histogramme semble bien adaptée car elle permet d'élargir la dynamique des niveaux de gris à la fois dans les zones claires et des zones sombres. L'image résultante devient plus contrastée, ce qui augmente la différence d'intensité entre les points clairs et sombres de l'image. La méthode standard pour améliorer le contraste pour les images couleur, consiste à appliquer indépendamment l'égalisation d'histogramme à chaque canal d'une image RGB. Cette technique conduisant à une forte concentration de pixels lumineux, peut aussi induire des artefacts de couleur dû à l'encodage combiné de l'intensité et de la chromaticité dans les images RGB [Han11], ce qui peut entraîner des effets indésirables sur le suivi de points. Pour surmonter ce problème, nous convertissons l'image dans l'espace de couleur HSV (se référer à l'annexe.A) et nous appliquons indépendamment l'égalisation d'histogramme, à la saturation (S) et à la valeur (V). Nous gardons la chromaticité de l'image (Hue) inchangée. Nous recombinaisons l'image et nous reconvertissons l'image dans l'espace couleur RGB. Cette technique permet une meilleure distribution de la luminance sans affecter la teinte de l'image et créer des artefacts de couleurs, qui peuvent affecter les performances du suivi. Évidemment, l'égalisation d'histogramme est appliquée sur une ROI qui n'inclut pas les pixels noirs qui sont en dehors du champ de vue de la caméra (voir figure 4.6). Appliquer l'égalisation d'histogramme au canal V, rend les zones sombres plus contrastées. Pour ce qui est du canal S, l'égalisation d'histogramme rend les couleurs plus vives.

L'égalisation d'histogramme, permet de modifier l'image de sorte que son histogramme soit plus étiré, réalisant ainsi une amélioration du contraste, mais faisant en sorte que son gradient devient irrégulier (voir figure 4.7 a-b). Ces irrégularités peuvent créer des réponses multiples durant la phase

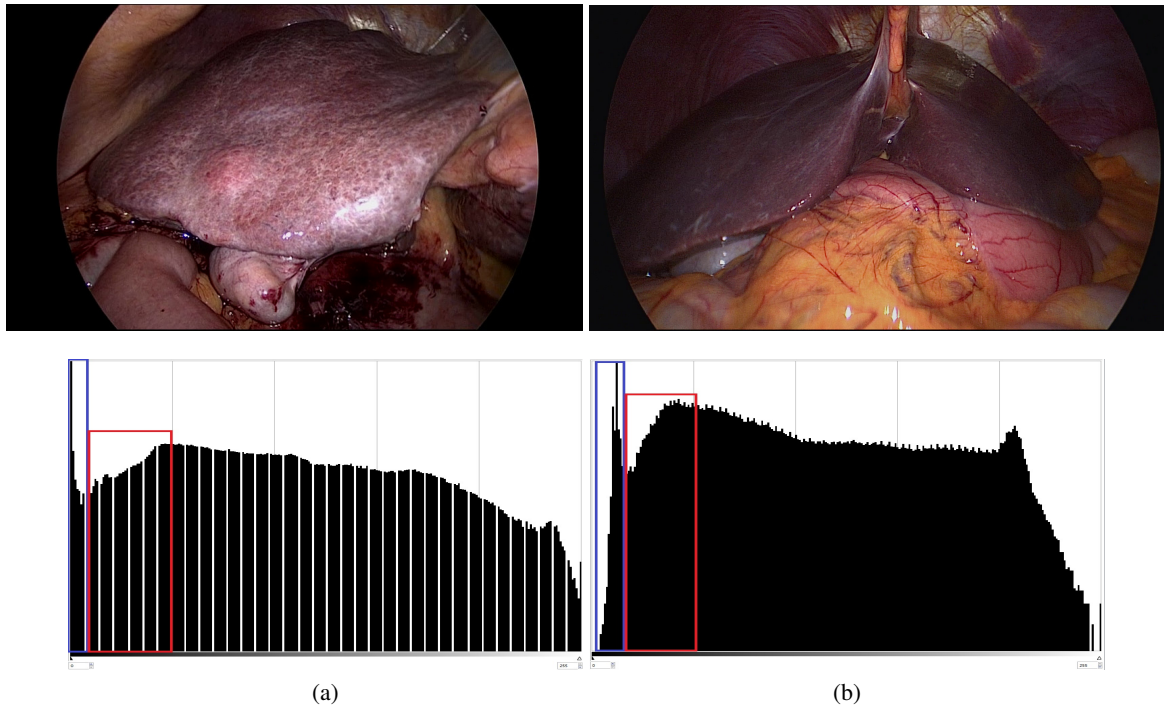
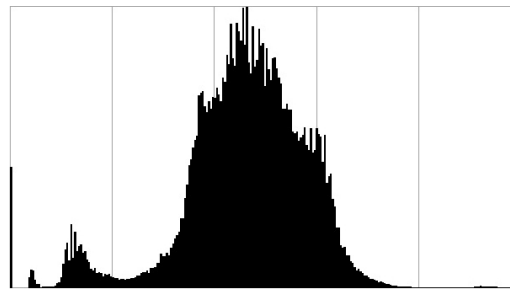


FIGURE 4.6 – Exemple d’images en chirurgie laparoscopique avec leurs histogrammes. Carré bleu : pixels noirs en dehors du champ de vue de la caméra. Carré rouge : pixels noirs à l’intérieur de champ de vue la caméra.

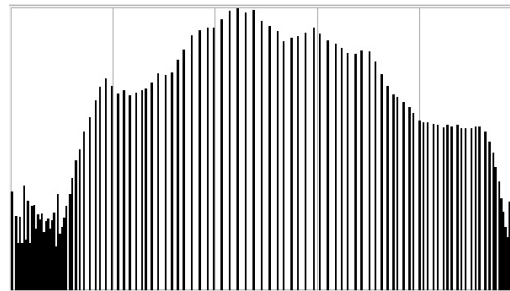
de détection. Afin d’obtenir une répartition uniforme de l’histogramme résultant et dans le but d’obtenir un gradient plus régulier, nous appliquons un filtre gaussien (G_σ) dans chaque canal S et V après l’égalisation d’histogramme, avec un $\sigma = 1,5$ (et un fenêtre de taille 5×5). Le σ a été adapté empiriquement par rapport à la résolution des images utilisées dans nos expérimentations, pour ne pas effacer les petites structures internes (vaisseaux sanguins). L’application d’un filtre gaussien à V (respectivement S) réduit la réponse du bruit et rend l’histogramme de l’image plus uniforme (respectivement étale la couleur des bords sombres comme les vaisseaux sanguins et des structures internes) (voir figure 4.7 b-c).

4.3.2 Chapeau haut de forme

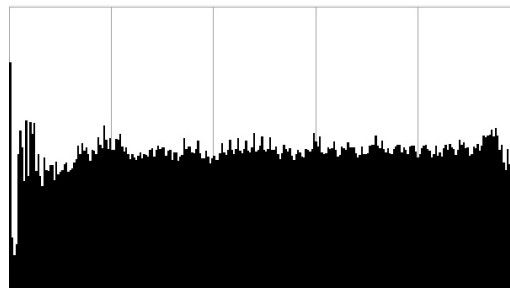
Augmenter le contraste des régions dans l’image peut être obtenu par application des opérateurs morphologiques sur l’image. En effet, ajouter la transformation en chapeau haut de forme blanc (white top-hat, WTH) à l’image originale augmente le contraste des objets clairs est soustraire le chapeau haut de forme noir (black top-hat, BTH) à l’image originale augmente le contraste des régions sombres (se référer à l’annexe.B). Soient ϕ et γ une fermeture et une ouverture morphologique, I et B une image et un élément structurant. Le chapeau haut de forme (top-hat) est défini comme suit :



(a) Histogramme d'une image originale



(b) Résultat d'une égalisation d'histogramme



(c) Résultat d'un filtrage gaussien l'histogramme (b)

FIGURE 4.7 – Influence de l'égalisation d'histogramme et du filtrage gaussien sur l'histogramme d'une image.

$$I^{th} = I + \underbrace{(I - \gamma_B I)}_{WTH_B} - \underbrace{(\phi_B I - I)}_{BTH_B} \quad (4.5)$$

L'élément structurant a été choisi de façon à augmenter le contraste des petites structures (vaisseaux sanguins). Par conséquent, l'élément structurant utilisé est une croix de taille 11×11 . Afin d'appliquer un opérateur morphologique sur les images couleurs, nous utilisons un ordre vectoriel qui permet d'ordonner l'intensité des couleurs pour calculer l'ouverture et la fermeture afin d'éviter les artefacts de couleur [Han04]. Nous appliquons à la fin du processus un filtrage gaussien avec un $\sigma = 1.5$ et une fenêtre de taille 5×5 afin de réduire le bruit.

4.3.3 Toggle

L'opérateur morphologique toggle est basé sur les érosions ε et les dilations δ de l'image. Il peut être utilisé pour segmenter ou augmenter les détails de l'image. Soit I et B respectivement une image et un élément structurant. Le toggle est défini comme suit :

$$T_g = \begin{cases} \delta_B I, & \text{si } \delta_B I - I < I - \varepsilon_B I \\ \varepsilon_B I, & \text{si } \delta_B I - I > I - \varepsilon_B I \\ I, & \text{sinon} \end{cases} \quad (4.6)$$

Dans le toggle, la valeur du pixel de l'image I est remplacée par la valeur du résultat de la dilatation ou de l'érosion de l'image, ce qui augmente les détails de l'image tels que les contours des petites structures. Nous convertissons l'image dans l'espace de couleur HLS (se référer à l'annexe.A) et nous appliquons le toggle opérateur uniquement sur le canal de la luminance (L) pour après revenir à l'espace de couleur RGB. Du fait que la luminance dans L encode l'information de la couleur des 3 couleurs primaires R, G, et B, en appliquant le toggle sur L nous évitons de produire des artefacts de couleur [Han04]. Nous appliquons par la suite un filtrage gaussien avec un $\sigma = 1.5$ et une fenêtre de taille 5×5 afin de réduire le bruit.

4.3.4 Toggle basé top-hat

L'opérateur morphologique top-hat augmente le contraste des régions, quant au toggle il permet d'augmenter les détails de l'image. Dans [BZX12] les auteurs proposent de combiner les deux opérateurs pour accomplir les deux fonctions. Soit I^{th} l'image pré-traitée avec du top-hat et B , $\varepsilon_B I$, δ respectivement l'élément structurant, une érosion et une dilatation. Le toggle basé top-hat est défini comme suit :

$$T_g^{th} = \begin{cases} \delta_B I^{th}, & \text{si } \delta_B I^{th} - I^{th} < I^{th} - \varepsilon_B I^{th} \\ \varepsilon_B I^{th}, & \text{si } \delta_B I^{th} - I^{th} > I^{th} - \varepsilon_B I^{th} \\ I^{th}, & \text{sinon} \end{cases} \quad (4.7)$$

Premièrement nous calculons le top-hat sur l'image I comme indiqué dans la Section 4.3.2. Deuxièmement, nous convertissons l'image résultante RGB vers l'espace de couleur HLS et nous calculons le toggle sur le canal L pour ensuite revenir dans l'espace de couleur RGB. La valeur du pixel de l'image résultante T_g^{th} est définie par le résultat de la dilatation ou de l'érosion (voir équation 4.7) de l'image pré-traitée avec le top-hat.

Pour évaluer l'influence des pré-traitements basés sur les opérateurs morphologiques, nous fournissons dans la figure 4.8 une comparaison visuelle de l'effet de chaque pré-traitement sur des images agrandies qui montre des vaisseaux sanguins des intestins de cochon et de la surface du foie humain. Il est clairement constaté une modification de la discontinuité et/ou du contraste des vaisseaux sanguins et de la surface du foie. De plus, nous fournissons dans la figure 4.9 un exemple des pré-traitements appliqués sur des séquences d'images en chirurgie laparoscopique.

4.3.5 Évaluation de l'influence d'un pré-traitement

Pour évaluer l'influence d'un pré-traitement sur les performances du suivi nous avons utilisé 6 séquences vidéos, trois correspondantes à une exploration abdominale chez un humain C.1C.3C.4 et trois correspondantes à une exploration abdominale chez un cochon C.6 C.5 C.7 (voir Annexe.C).

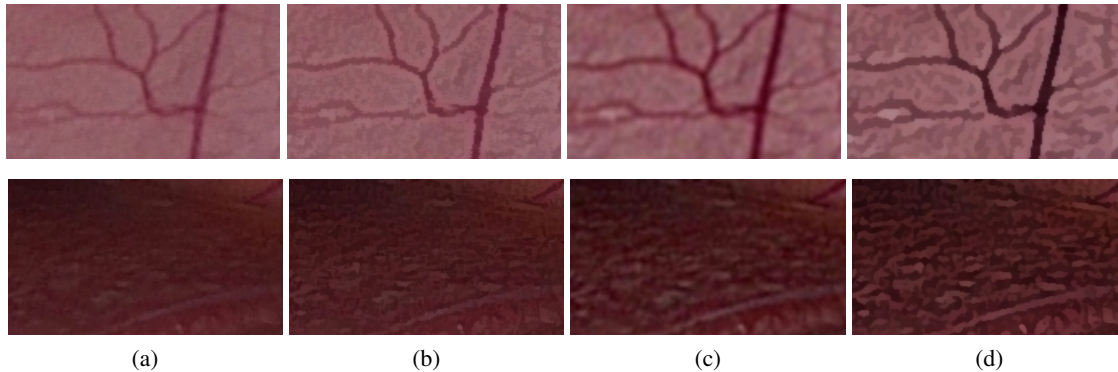


FIGURE 4.8 – Exemples d’applications des opérateurs morphologiques sur des séquences d’images en chirurgie laparoscopique. La première ligne montre les résultats sur les vaisseaux sanguins des intestins de cochon. La seconde ligne montre les résultats sur la surface d’un foie humain. (a) Image originale, (b) Toggle, (c) Top-Hat, (e) Toggle basé Top-Hat.

Dans cette évaluation, nous proposons d’utiliser les détecteurs : Shi-Tomasi (dit GFTT : good features to track) [Tom94], SURF [Bay08], FAST [Ros06] et BRISK [LCS11]. Les paramètres suivants, ont été utilisés pour toutes les séquences : GFTT les seuils ont été définis de la sorte : 0,001 à 0,01 avec un pas de 0,001 et de 0,01 à 0,09 avec un pas de 0,01. Pour SURF : les seuils de la matrice hessienne de 50 à 900 avec un pas de 50. Pour BRISK : le seuil de détection, varie de 10 à 28 avec un pas de 1. Pour FAST : le masque par défaut 9-16 a été utilisé, et un seuil de détection qui varie de 10 à 46 avec un pas de 2. Ceci permet en effet, une évaluation pour une grande variation des nombres de points détectés. Nous proposons d’évaluer l’influence du top-hat et de l’égalisation lissée comme pré-traitement lors de la phase du suivi et de détection de points.

4.3.5.1 Évaluation de l’influence du pré-traitement durant la phase du suivi de points

Nous proposons d’évaluer dans quelle mesure le processus de pré-traitement améliore la robustesse et la précision de suivi de point (quelle que soit la méthode de détection utilisée). En fait, le nombre et la robustesse des points lors de l’étape de suivi dépend généralement d’un certain seuil que l’utilisateur choisit à l’étape de détection. Un seuil élevé devrait détecter un grand pourcentage de points robustes, alors qu’un seuil bas devrait trouver beaucoup plus de points, mais dont un grand pourcentage avec de faibles performances de suivi. Ainsi, pour démontrer l’avantage d’un pré-traitement, il convient d’améliorer le pourcentage de points robustes quel que soit le seuil utilisé. Par conséquent, nous proposons la méthode suivante. Pour un seuil donné, nous détectons les points à suivre dans la première image sans pré-traitement. Nous suivons l’ensemble de points sur la séquence avant/arrière avec et sans pré-traitement. Nous comparons le pourcentage de points robustes et l’histogramme des ϵ (comme indiqué.3.2.1.1) avec et sans pré-traitement. Ce suivi est répété pour différentes valeurs de seuil qui couvrent la plage des seuils des détecteurs. Par conséquent, nous obtenons une comparaison de la robustesse et de la précision en fonction du seuil. Dans la figure 4.10, nous montrons un exemple du résultat de l’évaluation de l’influence des pré-traitements d’une séquence exemple (séquence C.7). Les figures 4.10a, 4.10b montrent le pourcentage de points robustes avec une dérive de moins de 5 Pixels (5 pixels correspondent à la largeur la plus fine d’un vaisseau). Nous remarquons qu’appliquer un pré-traitement améliore la robustesse du suivi, quel que soit le seuil de détection et/ou le détecteur utilisé.

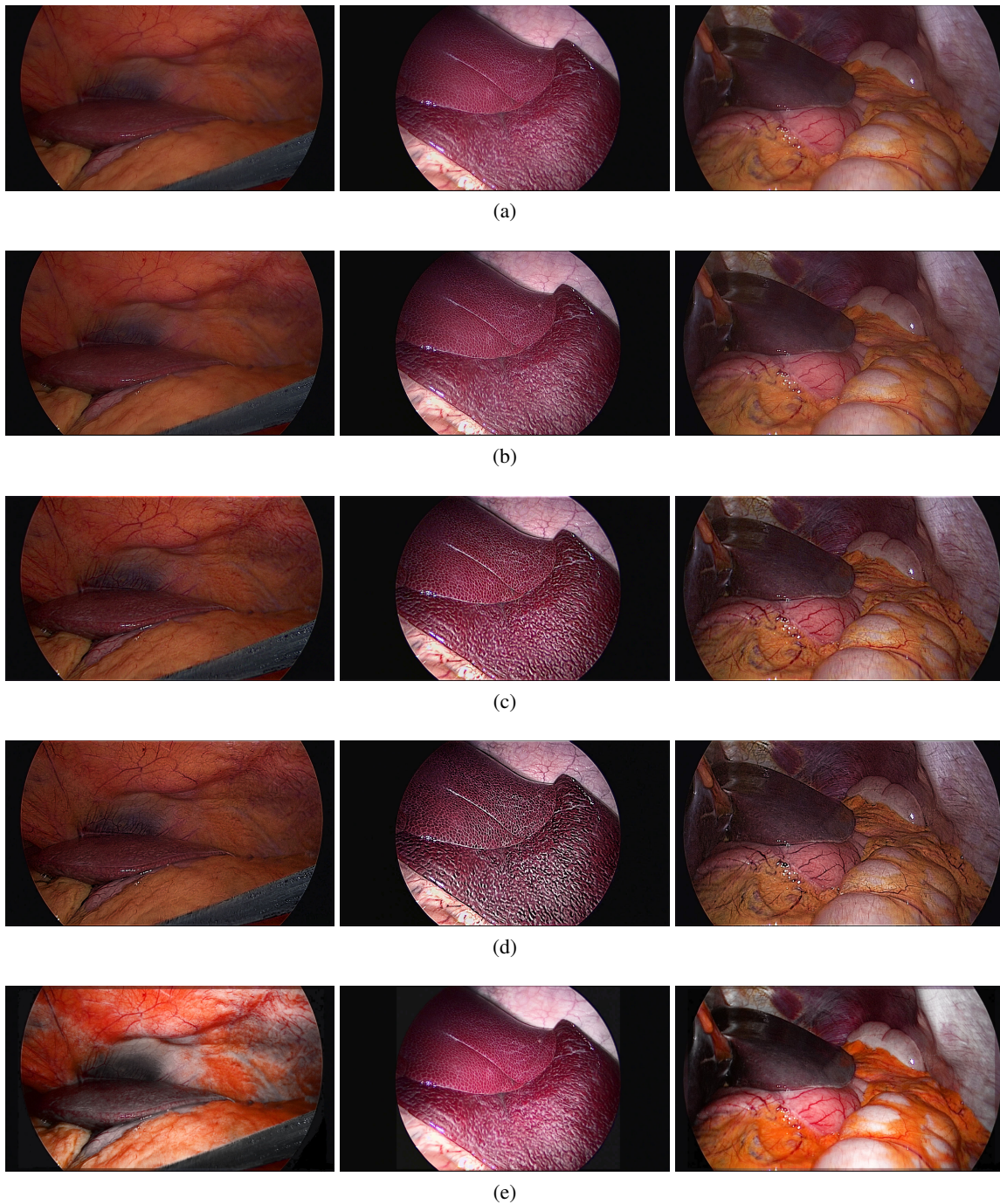


FIGURE 4.9 – Exemple des pré-traitements utilisés sur des séquences d’images en chirurgie laparoscopique. a) image originale, b) toggle, c) top-hat, d) toggle top-hat, e) égalisation lissée.

Suivi	Pas de pré-traitement		
	Detection	Top Hat	Égal lissée
GFTT	77.7%	83.3%	81.1%
SURF	81.6%	90.4%	91.3%
BRISK	81.8%	85.6%	88.3%
FAST	77.5%	82.2%	84.8%

TABLE 4.3 – Moyenne des pourcentages de points robustes avec et sans pré-traitement.

Les Fig.4.10e, 4.10f montrent l'histogramme des erreurs ϵ du suivi pour un seuil fixe qui permet de détecter 500 points avec les détecteurs GFTT, SURF, BRISK et FAST. Nous affichons le nombre de points dont la dérive est inférieure à 1 pixel. Il apparaît clairement que le pré-traitement augmente également la précision du suivi. Le tableau 4.3 montre les résultats du pourcentage de points robustes moyennés pour toutes les valeurs des seuils et pour les 6 séquences vidéos. Les résultats justifient sur différentes séquences vidéo et type de tissu, qu'appliquer un pré-traitement aux images pour le suivi, améliore les performances indépendamment des méthodes utilisées pour la détection de points.

4.3.5.2 Évaluation de l'influence du pré-traitement à la détection de points

Pour évaluer dans quelle mesure le processus de pré-traitement influe sur la détection de points en terme de robustesse et de précision, nous comparons les ensembles de points de la même taille détectés avec et sans pré-traitement. Par conséquent, pour un nombre donné de points à détecter, nous adaptons les seuils en conséquence pour chaque pré-traitement, grâce à la configuration des seuils pré-calculés. Nous suivons l'ensemble des points détectés dans le même mode que du pré-traitement utilisé pour la détection (pour l'ensemble des points détectés sans pré-traitement nous n'utilisons aucun pré-traitement pour le suivi). Le processus est répété pour un nombre différent de points (et donc pour des valeurs de seuil différentes). Le tableau 4.4 fournit le pourcentage de points robustes pour une séquence exemple sur un foie d'humain (C.1) et aussi la moyenne des résultats pour les 6 séquences. Nous remarquons qu'en moyenne, la précision du suivi est améliorée en utilisant le top-hat (resp. égalisation lissée) de 8,2 % (resp. 9,4 %).

L'évaluation de l'influence des pré-traitements sur les performances de suivi, a été étudiée exhaustivement grâce à notre stratégie de suivi décrite dans 3.2.1. Il est à noter que la tendance entre le top-hat et l'égalisation lissée pour améliorer les performances de suivi, varie en fonction des séquences utilisées. Ce phénomène s'explique par la différence des tissus et des scènes opératoires dans le cadre des images en chirurgie laparoscopique. C'est pourquoi, il serait intéressant de trouver le bon pré-traitement associé à chaque type de scène visualisé.

4.4 Dérivation d'une approche contexte spécifique pour le suivi de points

Nous avons vu précédemment, qu'un pré-traitement améliore les performances de suivi, et ceci pour différents détecteurs. Toutefois, les performances de suivi varient pour chaque détecteur et pré-traitement en fonction de la scène visualisée. Proposer une approche de suivi spécifique à la scène, dans le but d'améliorer la performance de suivi de points, semble donc être pertinente.

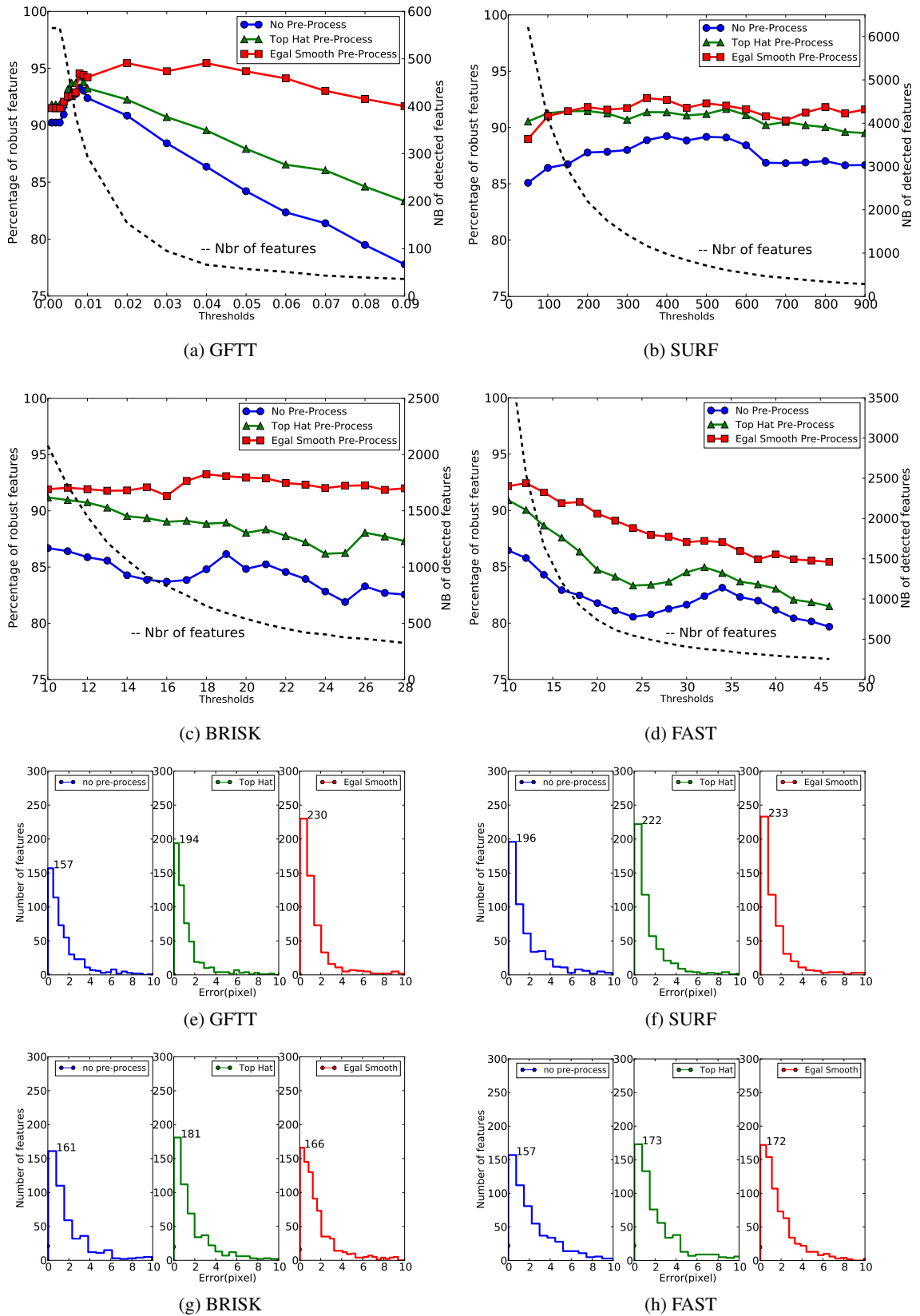


FIGURE 4.10 – (a,b,c,d) Pourcentage de points robustes avec et sans pré-traitement pour différents seuils de détection. (e,f,g,h) L’histogramme des erreurs du suivi pour 500 points détectés.

Pre-processes Sequence	Pas de pré-traitement		Top Hat		Égal lissée	
	Foie	Moyenne	Foie	Moyenne	Foie	Moyenne
GFTT	84.3%	78.7%	93.1%	82.1%	94.3%	83.8%
SURF	87.5%	76.4%	96.7%	84.3%	95.8%	84.5%
BRISK	84.1%	77.1%	91.3%	81.1%	93.5%	82.4%
FAST	83.1%	74.8%	88%	79.2%	91.3%	82%

TABLE 4.4 – Pourcentage de points robustes en fonction du pré-traitement appliquer.

Notre idée se base sur notre méthodologie de validation de la trajectoire du suivi de points, comme expliqué dans le chapitre 3 3.2.1. Nous proposons de sélectionner à partir d'un ensemble de pré-traitements, de détecteurs et d'algorithmes de suivi de points la combinaison qui garantit la meilleure performance de suivi de points en termes de précision, robustesse, et de distribution de points pour chaque séquence vidéo. En se basant sur quelques images de la séquence vidéo (environ 3 secondes). Le but est d'évaluer les différents algorithmes et de trouver la meilleure combinaison possible qui garantit la meilleure performance de suivi de point, adapté à la scène courante. La combinaison d'algorithmes trouvés sur les premières images de la séquence vidéo à l'initialisation, est supposée rester la meilleure durant le reste de la séquence. En effet, en chirurgie laparoscopique, les chirurgiens se concentrent généralement sur un organe d'intérêt ce qui se traduit par une même scène visionnée (couleur, texture du tissu) durant l'intervention. Cependant, dans le cas où le chirurgien oriente la caméra sur un nouveau organe ou qu'un changement significatif se produit dans la scène (sang, ablation des tissus, changement de l'intensité la lumière, ...), nous adaptons la combinaison en redémarrant l'évaluation.

4.4.1 Description des indicateurs de performances

Dans le but de fournir une approche spécifique au contexte du suivi et d'intégrer les processus du pré-traitement de l'image lors de la détection et du suivi, nous proposons de décomposer le suivi en 4 étapes tout en fournissant une analyse temps réel des différents pré-traitements ainsi que des algorithmes de détection et de suivi sur des données courantes. Les 4 étapes de suivi se décompose de la sorte :

1. un pré-traitement pour la détection
2. détection de points
3. un pré-traitement pour le suivi de points
4. suivi de points

Un pré-traitement efficace pour la détection ne l'est peu être pas pour le suivi, ce qui justifie notre choix d'intégrer les pré-traitements dans les deux phases (détection et suivi). Une combinaison est donc caractérisé par un quadruplet d'algorithmes. De ce fait, pour chaque quadruplet et pour un nombre de points N à détecter nous adaptons les seuils des différents détecteurs en incluant les pré-traitements pour la détection. Nous utilisons une approche dichotomique permettant d'adapter les seuils des détecteurs. Cette approche permet une évaluation plus cohérente. Cependant, comme vu dans le chapitre 3 adapter les seuils pour détecter le même nombre de points ne garantit pas une distribution identique des points détectés dans l'image. La distribution et la densité des points détectés dans la première image sont également des facteurs importants et doivent être évaluées, avant de mesurer

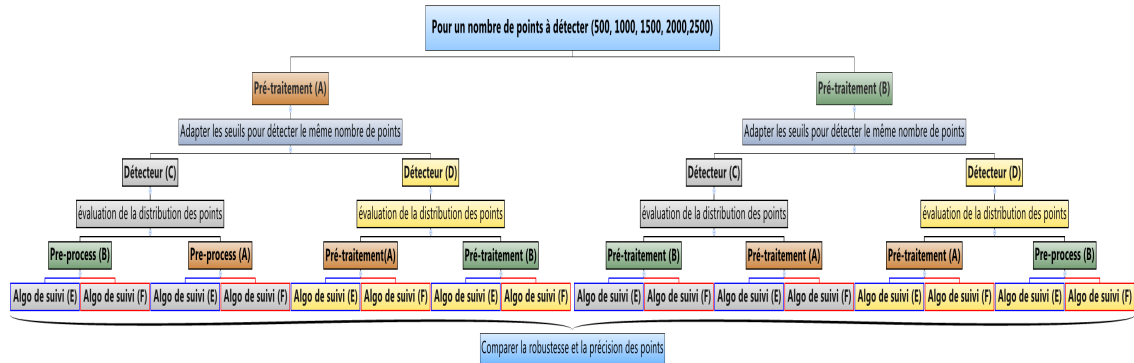


FIGURE 4.11 – Diagramme illustrant les différentes évaluations pour trouver la meilleure combinaison possible à partir d'un ensemble de : 2 pré-traitements, 2 détecteurs et 2 algorithmes de suivi.

la robustesse et la précision du suivi de points. Si les distributions sont trop différentes entre les détecteurs, la comparaison n'a pas de sens. Par conséquent, pour chaque quadruplet nous évaluons d'abord la distribution (comme expliqué dans le chapitre 3), et par la suite la précision et la robustesse des points suivis. La figure 4.11 résume notre stratégie qui permet d'identifier la meilleure combinaison possible à partir d'un ensemble d'algorithmes disponibles (détection, pré-traitement, suivi).

4.4.2 Évaluation du contexte-spécifique pour le suivi de points

Pour montrer l'intérêt de notre approche spécifique au contexte du suivi, nous avons utilisé 6 séquences vidéos dont 3 séquences réalisées sur des patients humains (C.1, C.2 et C.3), et 3 sur des cochons (C.5, C.6 et C.7).

Pour l'ensemble des algorithmes, nous proposons d'évaluer 4 pré-traitements et aussi le fait de n'appliquer aucun pré-traitement à l'image : sans pré-traitement (N), Égal lissée (E), top-hat (H), toggle (t), toggle basé top-hat (T). Nous proposons d'évaluer l'influence de 2 algorithmes de suivi : Lucas-Kanade (LK) [Bou01] et le robuste local optical flow (RLOF) [SES12] avec les paramètres suivants : pour LK une fenêtre de recherche de taille (25×25) et une pyramide de niveau 3. Pour RLOF, la petite et la grande fenêtre de recherche ont été fixées à 20×20 et 25×25 le nombre d'itération maximale à 20 et une pyramide de niveau 3. Nous proposons aussi 4 détecteurs : GFTT (G) [Tom94], SURF (S) [Bay08], FAST (F) [Ros06] et BRISK (B) [LCS11]. Les seuils des détecteurs ont été adaptés afin de détecter un nombre de points choisis par l'utilisateur : nous avons défini 5 nombres de points à détecter (500, 1000, 1500, 2000, 2500).

Après avoir lancé nos évaluations des performances de suivi sur un morceau d'une séquence courante (live) pour tous les quadruplets et les nombres de points possibles, nous pouvons faire une comparaison exhaustive de tous les quadruplets. Donc à partir d'un ensemble d'algorithmes, l'utilisateur peut choisir la combinaison la plus adaptée en fonction de la séquence vidéo utilisée.

4.4.2.1 Exemple d'utilisation

Grâce à notre méthodologie d'évaluation et de l'approche spécifique au contexte du suivi, plusieurs cas de figures de sélection d'algorithmes de suivi sont possibles. Par exemple, nous pouvons aussi bien

Nbr de points détectés	500	1000	1500	2000	2500
GFTT	141.7 ± 128.5	118.8±122.7	111.8±123.3	108.2±123.8	103.1±123.5
SURF	120.4 ± 102.8	96.2 ± 94.5	88.8 ± 95.7	84.3 ± 94.8	80.9 ± 95.1
FAST	130.9 ± 132.6	103.5 ± 106.6	89.5 ± 97.7	84.8 ± 96.2	80.9 ± 94.7
BRISK	106.28 ±100.2	98.1 ± 104.3	89.1 ± 99.7	84.1 ± 96.1	81.1 ± 96.1

TABLE 4.5 – La moyenne et l'écart-type de la transformée de distance pour les différents nombres de points détectés à l'aide de GFTT, SURF, FAST et BRISK.

fournir la meilleure combinaison détecteur/algorithme de suivi sans inclure de pré-traitement pour la séquence live.

À titre d'illustration nous montrons sur la figure 4.12 un exemple d'évaluation des détecteurs sur une séquence vidéo de foie humain (séquence C.1) en utilisant LK (figure 4.12a) et RLOF (figure 4.12b) pour le suivi de points et sans appliquer de pré-traitement à l'image. Les résultats montrent que pour la séquence considérée, l'algorithme de suivi RLOF augmente la robustesse du suivi quel que soit le détecteur utilisé. Ce qui justifie que l'algorithme RLOF sera sélectionné pour le suivi de points, au lieu de LK. De plus, en moyenne pour tous les nombres de points détectés, le détecteur SURF offre le pourcentage le plus élevé de points robustes pour presque tous les nombres de points détectés. De ce fait, pour la séquence considérée, une combinaison du détecteur SURF et de l'algorithme de suivi RLOF garantit la meilleure performance de suivi.

Toutefois, si l'utilisateur avait comme spécificité de détecter 1000 points et d'utiliser l'algorithme de LK pour le suivi, il est clair qu'à partir de la figure 4.12a le pourcentage des points robustes est identique pour GFTT et SURF. Néanmoins, à partir du tableau 4.5, qui montre les évaluations de la distribution, il apparaît que le meilleur détecteur est l'algorithme SURF. Ceci est confirmé visuellement entre la figure 4.12c et figure 4.12d qui montre la distribution des points détectés en utilisant GFTT et SURF respectivement. Par conséquent, à partir d'un ensemble de détecteurs et d'algorithmes de suivi, l'utilisateur peut choisir la meilleure combinaison possible en fonction de la séquence vidéo utilisée et du nombre des points détectés.

4.4.2.2 Évaluation de la persistance du contexte spécifique

Afin d'évaluer la pertinence de notre approche « contexte spécifique pour le suivi », pour chaque séquence vidéo nous redémarons notre évaluation après 100 images ce qui correspond approximativement à un déplacement de 300 pixels de la scène. En moyenne, nous fournissons trois évaluations pour une séquence vidéo. En particulier, la figure 4.13 montre deux évaluations (sur 2 sous-séquences) de l'influence des algorithmes utilisés dans chaque étape du suivi sur la séquence de foie humain (séquence C.3) de l'image 0 à 69 (a,c,e,g) et de l'image 170 à 330 (b,d,f,h). Un déplacement total de 345 pixels (approximativement 3 cm) a été observé entre les deux séquences différentes. Le graphique (a) montre le pourcentage de points robustes par rapport au pré-traitement utilisé avant l'étape de détection pour différent nombre de points détectés. Chaque barre représente le pourcentage moyen de points robustes de toutes les combinaisons de détecteurs, pré-traitements pour le suivi et des algorithmes de suivi, le pré-traitement pour la détection étant fixé. Sur cette vidéo, les résultats montrent que l'ordre des algorithmes qui fournissent le meilleur pourcentage de points robustes dans chaque étape reste le même.

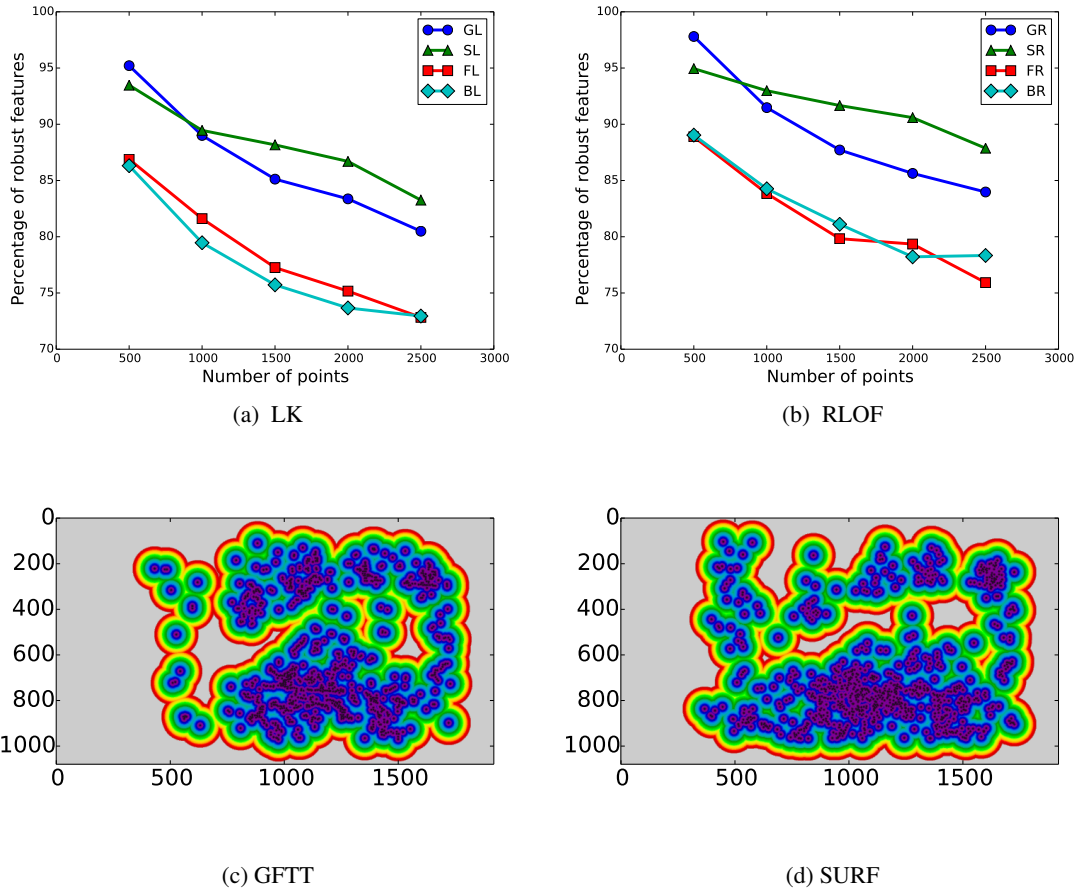


FIGURE 4.12 – Exemple de l'évaluation des détecteurs. (G) GFTT, (S) SURF, (F) FAST, (B) BRISK, (L) LK, (R) RLOF.

Considérant la sélection du meilleur quadruplet parmi toutes les combinaisons possibles, pour la première sous-séquence le quadruplet TGHR donne le meilleur résultat avec 91,3% de point robustes et une erreur moyenne de 1,84 pixels. Pour la deuxième sous-séquence, le résultat de l'évaluation montre à nouveau que le quadruplet TGHR offre le meilleur résultat jusqu'à 79,8% de points robustes et une erreur moyenne de 2,8 pixels. Nous remarquons une baisse des performances de suivi entre les deux sous-séquences, qui est expliquée par une différence de la durée des sous-séquences (approximativement de 3 secondes pour la première et de 7 secondes pour la deuxième), qui influe sur les algorithmes de suivi utilisant les approches récursives.

Comme nous l'attendions, sur toutes les séquences vidéos utilisées, un quadruplet qui offre la meilleure performance pour le suivi de points dans les premières images d'une séquence vidéo, reste à chaque fois le meilleur sur tout le reste de la séquence. Il est important de noter que le quadruplet qui donne la meilleure performance de suivi est spécifique à chaque séquence vidéo et qu'il varie entre les séquences vidéo sans corrélation significative.

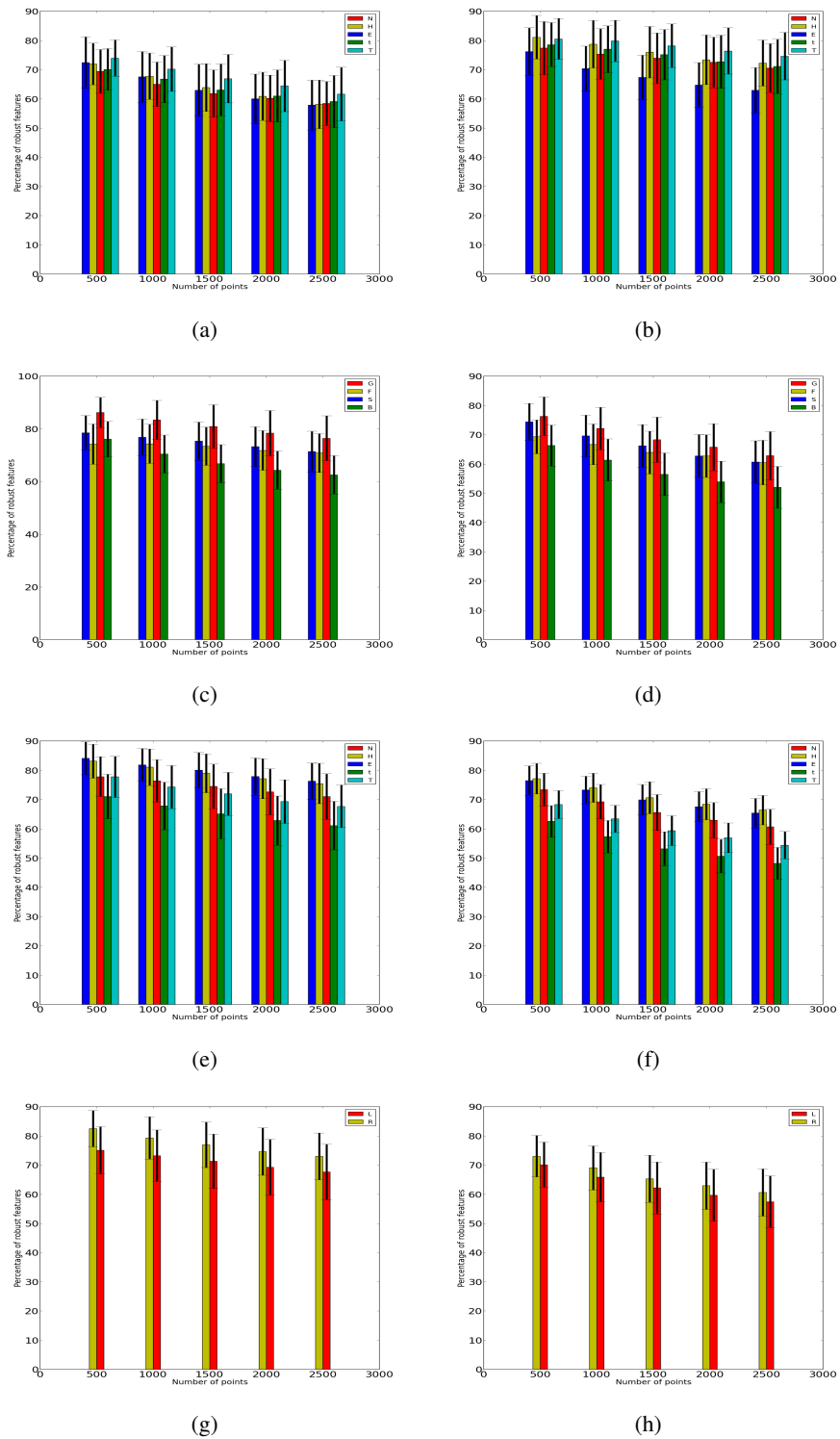


FIGURE 4.13 – Moyenne des points robustes et écart-type (barre noire) pour chaque étape du suivi. (a,b) pré-traitements pour le suivi. (c,d) détecteur utilisé. (e,f) pré-traitements pour le suivi (g,h) algorithmes de suivi. Dans les figures (a,b,e,f), (N) sans pré-traitement, (H) top-hat, (E) égalisation lissé, (t) toggle, (T) toggle basé top-hat. Dans les figures (c,d), (G) GFTT, (S) SURF, (F) FAST, (B) BRISK. Dans les figures (g,h), (L) LK, (R) RLOF.

Vidéos	Séquence.C.1	Séquence. C.6	Séquence. C.5	Séquence. C.3	Séquence. C.7	Séquence. C.2
Quadruplets	TGHR	TFNL	tGHL	TSHR	HGHR	TSHL
Robustesse	85.6%	95.4%	69.9%	87.7%	96.3%	99.8%
Précision	1.84	1.32	1.31	1.77	1.25	0.09

TABLE 4.6 – Pourcentage des points robustes ainsi que la précision du suivi pour le meilleur quadruplet sur les différentes séquences vidéos.

Vidéos	Séquence.C.1		Séquence. C.6		Séquence. C.5		Séquence. C.3		Séquence. C.7		Séquence. C.2	
	NGNL	NSNL	NGNL	NSNL	NGNL	NSNL	NGNL	NSNL	NGNL	NSNL	NGNL	NSNL
Robustesse	74.3%	67.5%	91.7%	94.2%	50.3%	61.8%	79.6%	80.6%	89.9%	88.2%	99.7%	97.8%
Précision	2.92	3.06	1.08	0.79	1.13	1.57	2.21	2.21	1.42	1.73	0.1	0.16

TABLE 4.7 – Pourcentage des points robustes ainsi que la précision du suivi des algorithmes standards sur les différentes séquences vidéos.

4.4.2.3 Évaluation de l'influence du contexte spécifique sur les performances de suivi

Pour évaluer l'influence de notre méthode sur les performances de suivi ainsi que son intérêt, nous proposons de comparer les meilleures combinaisons trouvées grâce à notre approche contexte-spécifique aux méthodes classiques et courante, utilisant l'algorithme de détection GFTT ou SURF et l'algorithme de LK pour le suivi de points, sans appliquer de pré-traitement à l'image. Le meilleur quadruplet qui garantit la meilleure performance pour le suivi de points est trouvé à partir d'un ensemble d'algorithmes de pré-traitement de détection et de suivi (décrit dans la section 4.4.2). Le tableau 4.6 fournit le pourcentages des points robustes ainsi que la précision des points calculés à partir des erreurs de suivi du meilleur quadruplet. Nous donnons aussi dans le tableau 4.7 les résultats en utilisant les algorithmes standards. A partir des deux tableaux nous remarquons clairement que la robustesse est améliorée jusqu'à 19% pour la séquence du foie de cochon et 18% pour le foie humain. En moyenne sur toutes les séquences, la robustesse et la précision sont améliorées respectivement de 8.5% et 13.4%. Le pourcentage des points robustes pour les séquences : foie humain avec cirrhose (séquence C.2) et rate du cochon (séquence C.5) est élevé, y compris avec les algorithmes standards. Ceci est dû à la nature des tissus qui sont plus texturées et qui ne nécessitent pas forcément un pré-traitement.

4.4.2.4 Évaluation du temps de calcul

Notre approche spécifique au contexte du suivi ainsi que les pré-traitements utilisés ont été implémentés avec un PC Intel Core I7-3.4 GHz processor, 8 GB of RAM et une carte graphique Nvidia GTS 450. Pour les pré-traitements, notre implémentation utilise l'accélération CPU-GPU de la librairie OpenCV sans une optimisation ou une parallélisation du code. Le tableau 4.8 fournit le temps de calcul nécessaire des pré-traitements sur des images FHD avec une résolution de (1920 × 1080) et des images WVGA avec une résolution (854 × 480). Dans le tableau 4.9 nous fournissons le temps de calcul nécessaire pour déterminer le meilleur quadruplet à partir d'un ensemble de 2 pré-traitements, 2 détecteurs et 2 algorithmes de suivi. Nous soulignons que le temps de calcul pour la recherche dichotomique qui permet d'adapter le seuil pour les détecteurs est inclus dans la valeur du temps de calcul. Comme nous pouvons le remarquer, la résolution de l'image et le nombre de points détectés affectent

Pré-traitements	E		H		t		T	
	FHD	WVGA	FHD	WVGA	FHD	WVGA	FHD	WVGA
Résolution vidéo								
Temps de calcul (<i>ms</i>)	0.01	0.008	0.06	0.02	0.05	0.01	0.11	0.03

TABLE 4.8 – Temps de calcul par images (FHD, WVGA). (E) égalisation lissée, (H) top-hat, (t) toggle, (T) toggle basé top-hat.

Résolution vidéo	FHD		WVGA	
Nombre d'images	70		98	
Nombre de points détectés	500	1500	200	700
Temps de calcul (<i>s</i>)	153	180	85	104

TABLE 4.9 – Temps de calcul pour différent nombre de points et résolution d'image.

largement le temps de calcul, ceci car le temps de calcul des algorithmes de suivi dépend largement de ces deux paramètres. Dans notre analyse, et pour toutes les combinaisons possibles (1000 évaluations générés), l'évaluation prend environ 3 à 4 heures pour une séquence moyenne de 150 images, ce qui est évidemment non acceptable pour fournir une application temps réel en salle opératoire. Cependant, nous sommes convaincus que ce temps de calcul peut être considérablement réduit à 3 ou 5 minutes à l'aide d'une optimisation de code (multi-threading, parallélisation, GPU) de tous les algorithmes.

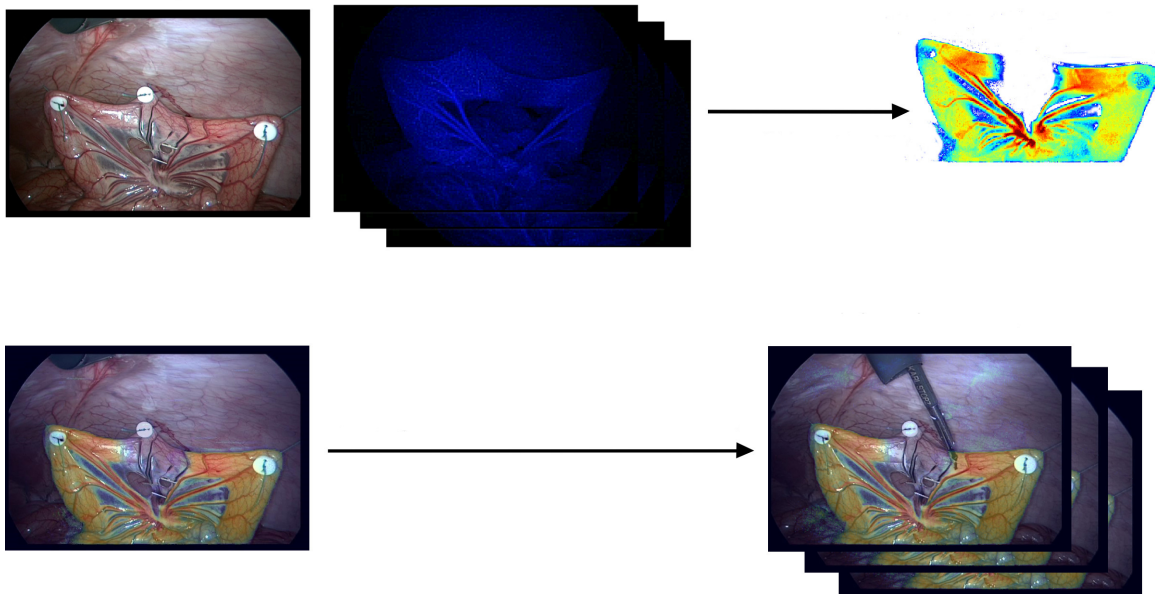
4.5 Synthèse

Dans ce chapitre, nous avons relevé trois points essentiels. Premièrement, l'analyse RGB a montré qu'il est pertinent et important de prendre en compte la totalité de l'information de l'image (image couleur) au lieu de l'image en niveau de gris. De plus, qu'un suivi est plus robuste sur une image bien contrasté (qui a été le canal vert). Deuxièmement, appliquer des pré-traitements à l'image durant la phase de détection et de suivi, afin d'améliorer le contraste de l'image, améliore les performances de suivi. Troisièmement, que chaque pré-traitement et/ou algorithmes de détection et de suivi, est spécifique pour chaque scène, d'où l'intérêt de fournir une approche spécifique au contexte du suivi. Nous avons validé d'une manière cohérente et pertinente l'aspect des trois propositions sur des données *in-vivo* prouvant ainsi qu'il est possible d'augmenter les performances de suivi grâce à des techniques simples et peu coûteuse en temps de calcul.

Il est à noter qu'une analyse exhaustive de plusieurs algorithmes pour notre approche spécifique au contexte, nécessite un ordinateur performant non pas pour appliquer le choix de la combinaison (quadruplet), mais pour faire l'évaluation au début de l'intervention.

APPLICATION EN RÉALITÉ AUGMENTÉE

L'intérêt de la vision par ordinateur en chirurgie mini-invasive, est de fournir des applications amenant à aider le chirurgien durant le geste opératoire. Fournir un outil précis et en temps réel, permet d'une part, de garantir au chirurgien un plus grand confort à la réalisation du geste chirurgical, et d'autre part, une plus grande sécurité pour la vie du patient.



5.1 Contexte clinique

Le cancer colorectal est le troisième cancer dans le monde avec plus de 1,3 million de nouveaux cas diagnostiqués chaque année [Fer13]. Le traitement chirurgical standard, est de pratiquer une anastomose viscérale. Elle consiste à rétablir la continuité du tube digestif après que l'on ait procédé à une résection du segment du côlon contenant les tumeurs.

Pour effectuer la résection, le chirurgien doit clamper les vaisseaux irriguant le segment contenant les tumeurs. Une fois clampée, la partie à résecter prend une couleur bleutée due à une ischémie (manque d'oxygène dans les cellules dû à un manque d'afflux sanguin), ce qui donne une idée approximative des sites où la résection doit être effectuée. Le principal risque de cette chirurgie est un défaut de cicatrisation de l'anastomose, provoqué par une mauvaise sélection des sites. Ceci provoque une ischémie sur la région du branchement qui remet en cause la continuité du circuit colique (fuite des matières fécales et risque de septicémie). Dans le cas d'une complication postopératoire (3% de cas) le taux de mortalité est de 32% [CLH06]. C'est pourquoi il est important que les extrémités digestives soient correctement vascularisées.

Ce type de procédure chirurgicale nécessite un temps supplémentaire qui est en général d'au moins une heure pour repérer les sites ischémiques (le temps que la zone devienne bleutée). De plus, cette méthode de diagnostic est très subjective et dépend de l'expérience du chirurgien. L'idée de fournir un système d'analyse quantitatif et rapide capable de repérer les sites ischémiques est nécessaire. Parmi les recherches récentes qui fournissent une évaluation quantitative de l'ischémie [UPdPV11], celui de Kudzus et al.[KRS10] permet de réduire de 60% la révision d'anastomose grâce aux propriétés de fluorescence du vert d'indocyanine (ICG) lorsqu'il est injecté par voie intraveineuse. En effet, l'utilisation de l'ICG comme outil de diagnostic, a augmenté ces dernières années. Lorsque l'ICG est injecté dans le corps humain, la molécule est liée rapidement aux protéines plasmatiques et reste dans le circuit vasculaire (du fait de la taille de la molécule)[Miw10]. ICG génère une fluorescence dans le proche infrarouge (845 nm longueur d'onde) sous excitation de lumière de longueur d'onde entre 750 à 800 nm. La longueur d'onde d'émission de l'ICG est loin des 2 pics d'absorption majeur dans le corps humain qui sont le sang et l'eau (cf. Fig. 5.1), ce qui fait que cette molécule a de très bonnes propriétés pour l'analyse des tissus [Miw10]. Il faut environ 40 secondes après une injection en bolus par voie intraveineuse pour que le produit se diffuse dans le côlon [Dia14a].

De nos jours, les caméras de fluorescence ont été miniaturisées et sont déjà disponibles dans le cadre de la chirurgie mini-invasive (MIS). Par exemple la caméra D-Light P de Storz qui permet de basculer entre la vue normale (lumière blanche) et la vue fluoroscopique par l'utilisation de filtres spécifiques (avec un retard de commutation de 1 seconde). La vue en fluorescence ne permet de distinguer les détails des tissus (pas de lumière) ce qui empêche les chirurgiens de définir sur les tissus les zones ischémiques. Nous verrons à travers ce chapitre les étapes nécessaires pour le développement d'une application en réalité augmentée qui rentre dans le cadre de la chirurgie MIS, dont le but est de fournir un outil qui permet d'aider le chirurgien durant la phase de définition des sites ischémiques. Premièrement, nous rappelons comment, à partir de l'évolution du signal de fluorescence, il est possible de créer une carte de perfusion (débit sanguin) qui permet de repérer les zones ischémiques [Dia14a]. Deuxièmement, nous décrivons notre stratégie employée pour déformer et recalibrer la carte de perfusion sur la séquence vidéo courante observée par le chirurgien. Troisièmement, nous évaluons la précision de notre approche sur des données animales.

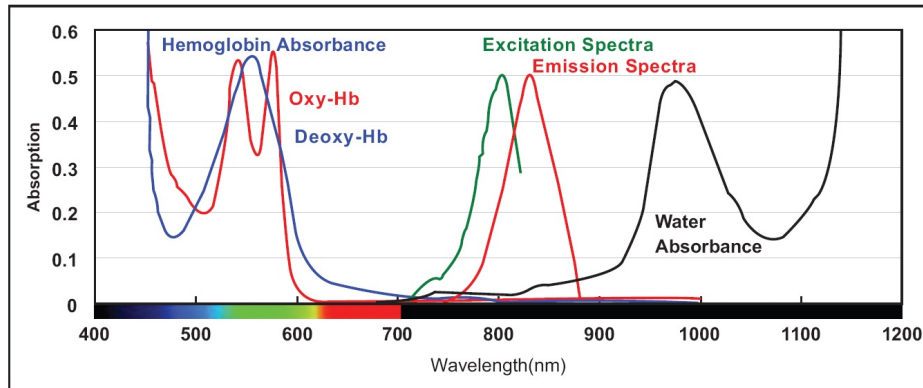


FIGURE 5.1 – Spectres d’excitation (ligne verte) et d’émission (ligne rouge) de l’ICG. Spectres d’absorption de l’hémoglobine (ligne bleu), et de l’eau (ligne noire). La lumière visible au-dessous de 650 nm de longueur d’onde est absorbée fortement par l’hémoglobine. Au-dessus de 900 nm la lumière infrarouge sera également fortement absorbée par l’eau.

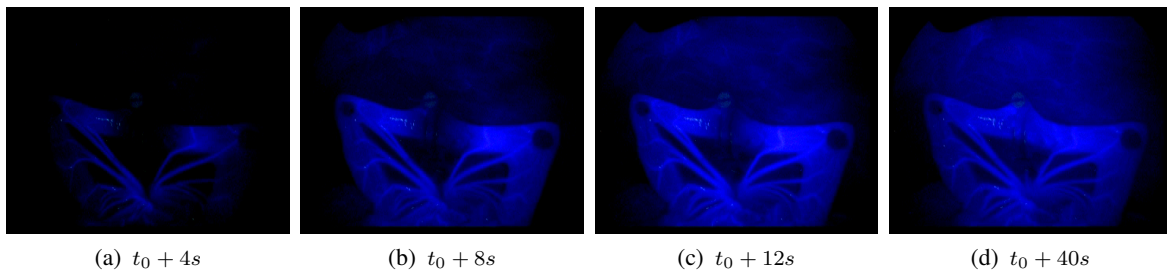
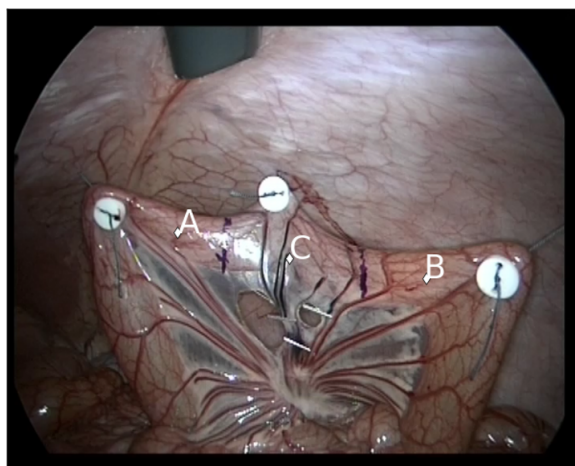


FIGURE 5.2 – Séquence représentant l’évolution du signal de la fluorescence en fonction du temps. t_0 début d’apparition de la fluorescence.

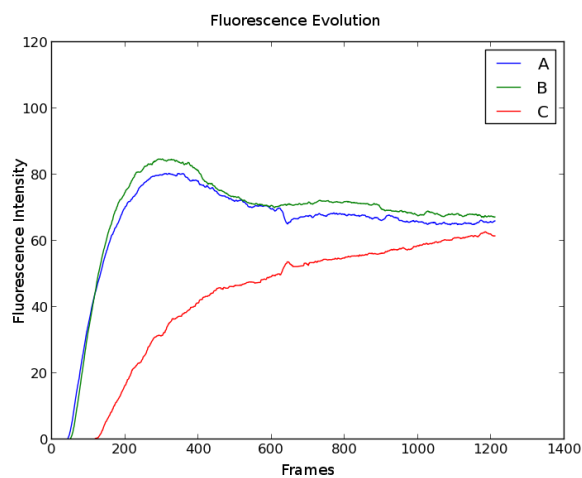
5.1.1 Génération de la carte de perfusion

Afin d’illustrer notre stratégie de l’analyse du signal de fluorescence de l’ICG pour générer une carte de perfusion, nous définissons artificiellement trois zones sur le côlon d’un cochon (cf. Fig. 5.3). A et B sont des zones bien irriguées (non ischémique). La zone C est artificiellement rendue ischémique par clampage des vaisseaux mésentériques (non visibles chez l’homme, mais visible chez le cochon). Après injection de l’ICG, nous basculons en vue fluoroscopique (cf. Fig. 5.2) afin d’analyser le signal de fluorescence de l’ICG (cf. Fig. 5.3a). La figure.5.3b montre l’évolution du signal de la fluorescence sur les trois zones. Nous remarquons clairement que le signal croît lentement pour la zone C par rapport aux zones bien irriguées (A et B). Au bout de 40 secondes le signal est constant partout, par conséquent, après ce délai, la commutation en mode de fluorescence endoscopique pour évaluer visuellement les emplacements d’ischémie est inutile. (cf. Fig. 5.2.d).

Il est donc nécessaire de réaliser une analyse dynamique du signal : l’information la plus pertinente se trouve dans la vitesse de croissance de la fluorescence c’est à dire sa pente. Nous proposons de créer une image où chaque pixel, affiche un code couleur correspondant au taux de croissance de la fluores-



(a)



(b) Évolution du signal de la fluorescence en fonction des zones A, B et C

FIGURE 5.3 – (a) Côlons du cochon avec A et B zones non ischémiques et C zone ischémique. (b) évolution du signal de fluorescence en fonction des trois zones (A, B et C).

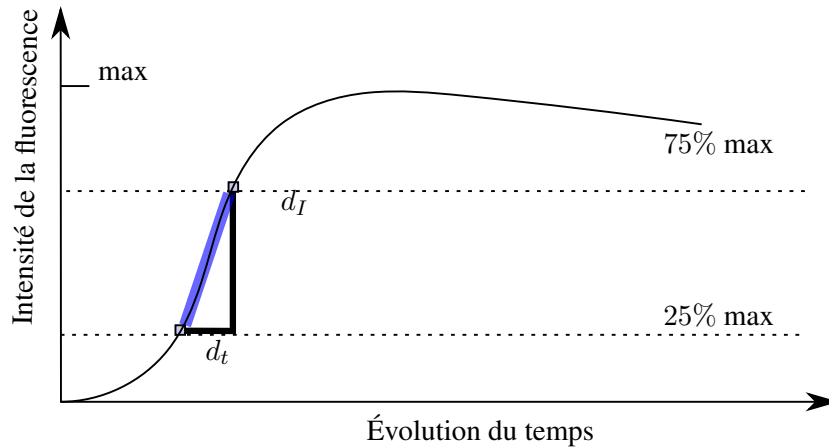


FIGURE 5.4 – Schéma illustrant l'évolution de l'intensité de la fluorescence en fonction du temps. Nous proposons d'étudier la pente (en bleu) entre le quartile 25% et 75% du signal maximal

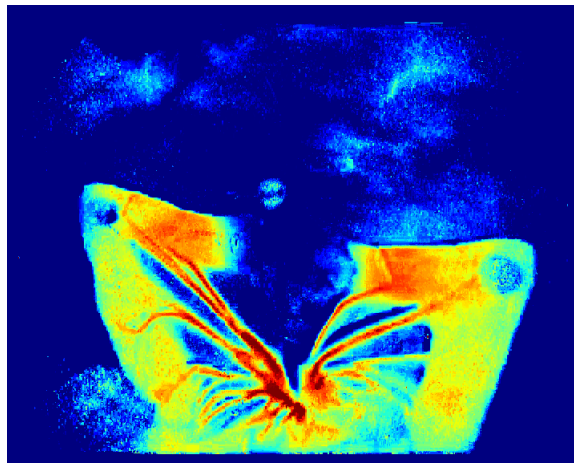


FIGURE 5.5 – Exemple de la carte de perfusion calculée à partir du signal de fluorescence de l'ICG.

cence. Pour chaque pixel de l'image la pente de croissance de la fluorescence est calculée en utilisant le rapport de progression de la fluorescence divisé par le laps de temps pris pour que la fluorescence passe de 25% du signal maximal à 75% du signal maximal (cf. Fig. 5.4). L'usage du quartile 25% et 75% permet d'être plus robuste au bruit et permet d'être indépendant de l'estimation du début de l'injection de l'ICG. Nous supposons que la scène reste statique (caméra fixée et respiration bloquée pendant 40 seconde) pendant le calcul de la pente de croissance de la fluorescence.

Finalement, la figure 5.5 illustre une image de carte de perfusion, montrant la vitesse de croissance du signal de fluorescence en chaque pixel. Nous pouvons facilement différencier sur cette image les zones ischémiques (couleurs froides) des zones bien perfusées (couleurs chaudes).

5.2 Recalage de la carte de perfusion

Dans cette section nous décrivons notre stratégie afin de pouvoir superposer directement la carte de perfusion sur la vue endoscopique courante et fournir une vue en réalité augmentée, dynamique et en temps réel. Cela évitera au chirurgien d'effectuer un recalage mental de la carte d'ischémique (figure.5.6.c) calculée à partir de la vue de référence (figure.5.6.c). En effet, la respiration, le mouvement péristaltique du côlon et les interactions des instruments avec les tissus, modifient continuellement la scène chirurgicale.

La méthode que nous proposons pour recalibrer la carte de perfusion est divisée en deux parties. Premièrement, nous estimons le mouvement des tissus à partir du suivi de points d'intérêt. Deuxièmement, nous calculons une carte de déformation dense à partir du suivi de points ce qui permet de recalibrer la carte de perfusion de façon non-rigide. Le principe général de notre recalage est résumé dans la figure 5.6. Nous avons vu que la première partie de l'application, consiste à générer la carte de perfusion (figure. 5.6.c). Ici est exploité le fait que la caméra a été fixée et que la respiration a été bloquée pendant l'acquisition de la fluorescence. De ce fait, l'image de référence (capturée juste avant la fluorescence) (figure.5.6.a) sera utilisée pour extraire des points d'intérêts qui seront appariés aux images de la séquence vidéo courante (live). Nous nous basons sur ces appariements pour calculer un champ de déformation dense, qu'il sera appliqué à la carte de perfusion. Le recalage de la carte de perfusion avec l'image de référence correspond à une vue statique en réalité augmentée (figure.5.6.d).

5.3 Estimation du mouvement des tissus

Afin de garantir un suivi robuste aux occlusions avec aucune hypothèse sur la continuité spatio-temporelle des points suivis, Le suivi de points basé sur les techniques descripteurs (appariement de points) semble le plus approprié. En effet, le suivi doit se faire d'une image de référence à une image en cours. L'intervalle de temps entre ces des 2 images est de minimum de 40 secondes et le côlon peut subir de grande déformation (mouvement péristaltique).

En général, les approches descripteurs, trouvent un ensemble d'appariement de deux ensembles de points d'intérêt extrait de deux images en exploitant la similarité entre leur apparence (couleurs, angle, échelle, gradient). Cette information doit être distincte pour chaque point d'intérêt afin de garantir un appariement robuste. Nous rappelons qu'une signature identique entre deux descripteurs est représentée par une distance des descripteurs proche de zéro (voir chapitre 2 section 2.3.3.1).

Nous avons choisi d'utiliser l'algorithme SURF [BTV06] pour la détection de points d'intérêt. Ce choix est motivé par notre approche spécifique au contexte du suivi (chapitre 4 section 4.4) sur les tissus du côlon de cochon, qui montre que le détecteur SURF offre la meilleure performance de suivi. Pour ce qui est du descripteur de points, nous avons choisi l'algorithme BRISK [LCS11] qui est un descripteur binaire, ce qui permet d'effectuer rapidement le calcul des appariements nécessaire pour notre application.

Toutefois, les appariements initiaux basés uniquement sur l'apparence peuvent contenir un grand nombre d'appariements incorrects, qui sont dus à l'apparence similaire des points d'intérêts dans les images en chirurgie laparoscopique (similarité de la texture et des tissus vaisseau sanguin du côlon). Les mauvais appariements sont évidemment problématiques pour calculer une déformation cohérente de l'image (recalage incohérent et présence d'artéfacts : cf. Fig.5.7). Un certain nombre de filtres

(proximité physique, similitude d'échelle, orientation, rapport de la distance des descripteurs, ...) ont été proposés pour améliorer la robustesse de l'appariement des points [Yip12, PSM13a].

Dans notre contexte, nous ne pouvons pas nous appuyer sur la pertinence temporelle entre deux images, parce que nous comparons une image qui représente la vue courante transmise au chirurgien, à une image référence acquise plusieurs minutes avant. De ce fait, la position de la caméra et la forme du côlon sont différentes en raison du mouvement péristaltique et des interactions des instruments. Par conséquent, nous écartons les filtres de proximité physique, d'échelle et d'orientation pour améliorer la robustesse des appariements et nous proposons la stratégie suivante.

Soient S_R (resp. S_L) l'ensemble des points d'intérêts détectés dans l'image de référence (resp. l'image courante). Pour chaque point de l'ensemble S_R , nous calculons les deux meilleurs appariements dans l'image courante en utilisant la distance de Hamming entre leur descripteur, et pour chaque point de l'ensemble S_L , les deux meilleurs appariements dans l'image de référence. Nous obtenons ainsi deux ensembles d'appariements $M_{R \rightarrow L}, M_{L \rightarrow R}$, où chaque point d'intérêt est apparié à deux candidats. La prochaine étape, basée sur un critère de symétrie, consiste à filtrer ces appariements et se décompose en deux étapes.

Premièrement, pour chacun des ensembles ($M_{R \rightarrow L}, M_{L \rightarrow R}$), nous rejetons les deux candidats si la distance de Hamming des deux candidats est similaire. Par ailleurs, la condition qu'un appariement soit pertinent n'est satisfaite que si la distance de Hamming du premier candidat est très faible par rapport au second, le premier candidat est alors accepté comme appariement correct. Nous avons déterminé à partir de nos résultats expérimentaux qu'un ratio de 0,65 entre les distances des deux candidats donne de bons résultats. Deuxièmement, à partir des appariements corrects des deux ensembles ($M_{R \rightarrow L}, M_{L \rightarrow R}$) nous ne gardons que les appariements qui sont corrects dans les deux ensembles. Cela signifie que pour qu'un appariement soit accepté, il faut que les deux points soient le meilleur candidat de l'autre dans les deux ensembles.

Cette approche impose un filtre très restrictif qui peut éliminer des appariements corrects. Cependant, la présence de mauvais appariements impliquerait des erreurs importantes lors de l'étape de déformation de l'image (cf. Sec. 5.3.1) et il est donc crucial de garantir la robustesse des appariements de points d'intérêt. Nous verrons dans la section 5.4 que suffisamment d'appariements corrects sont calculés malgré ce choix.

5.3.1 Calcul de la déformation

Nous nous basons sur un recalage 2D de la carte de perfusion afin de fournir une vue dynamique en réalité en augmentée, qui est suffisant dans le contexte de notre application. En effet, au cours de l'intervention chirurgicale (à partir du début de la fluorescence à la définition des sites de résection) nous nous attendons à peu de mouvement de perspective dans la scène, puisque le mouvement de la caméra reste limité par le trocart. Nous utilisons la méthode décrite par Schaefer et al. [SMW06] afin de déformer l'image à partir d'un ensemble de points d'intérêt appariés et ainsi recalculer la carte de perfusion. La méthode proposée par Schaefer est basée sur la technique Moving Least Squares (MLS) qui utilise diverses classes de fonctions linéaires dont des transformations affines, de similitude et rigides. Ces fonctions génèrent un champ de déplacement dense à partir du déplacement connu de quelques points. Les méthodes MLS ont été choisies car elles sont faciles à mettre en œuvre et fournissent un calcul rapide des déformations.

Il est expliqué ci-dessous le principe de la méthode MLS. Soit P un ensemble de points d'intérêt dans l'image de référence (Fig.5.6.a) et Q l'ensemble des points d'intérêt correspondant dans l'image

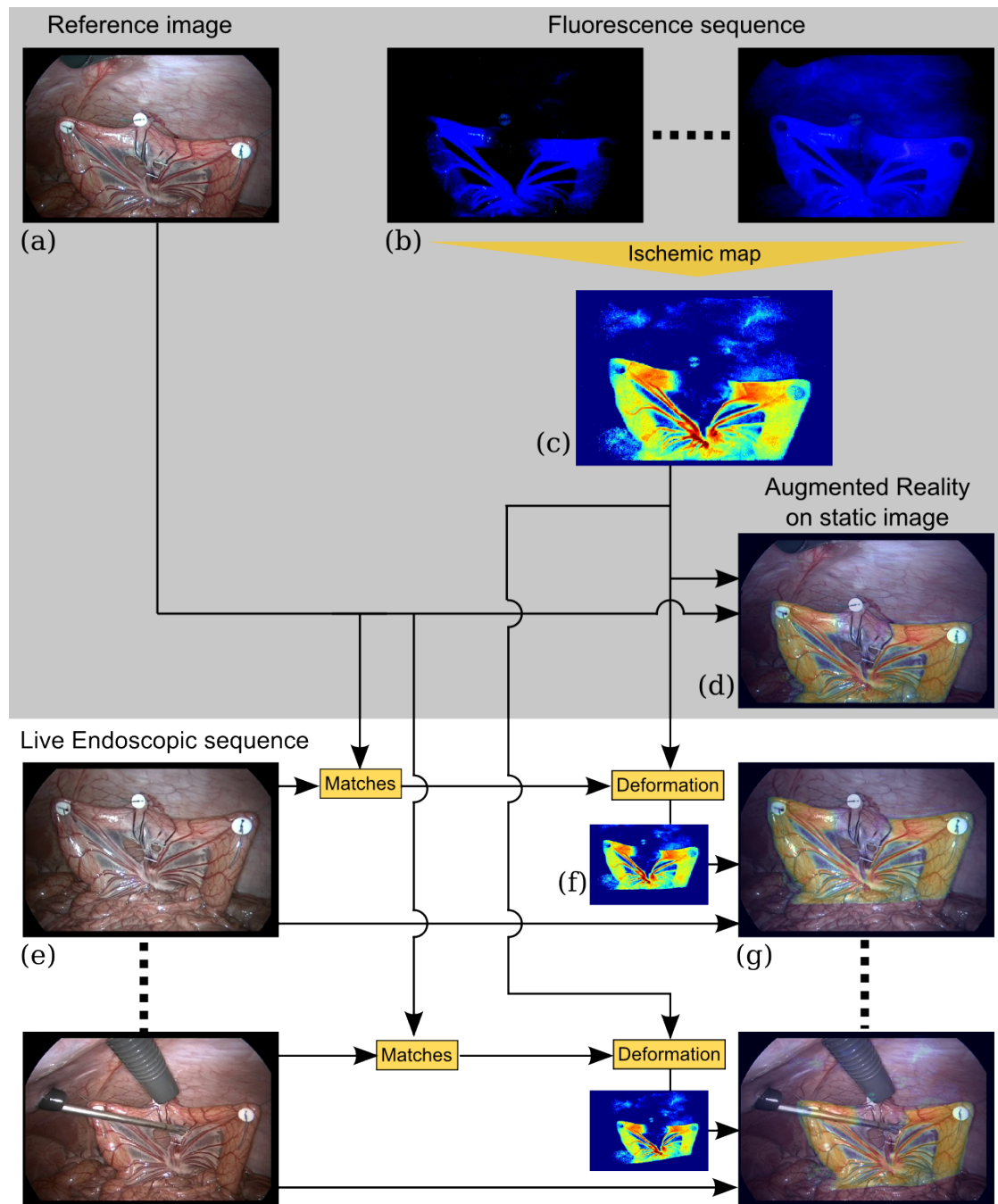


FIGURE 5.6 – Schéma illustrant notre application en réalité augmentée à base de fluorescence, avec l’image de référence (a), la séquence de fluorescence (b) et la carte de perfusion (c) calculées à partir de l’évolution temporelle du signal de la fluorescence. Vue statique en réalité augmenté (carré gris) : la carte de perfusion (c) est superposée à l’image de référence (a) pour fournir une vue statique de réalité augmentée statique (d). Vue dynamique et temps réel en réalité augmentée : un appariement entre les points d’intérêt est effectué entre l’image de référence (a) et l’image courante (e). A partir de ces points d’intérêt détectés, la carte de perfusion (c) est déformée pour obtenir une mise à jour de la carte (f). Nous superposons ensuite cette carte de perfusion déformée (f) sur la séquence en direct (e) offrant une vue de réalité augmentée dynamique (g).

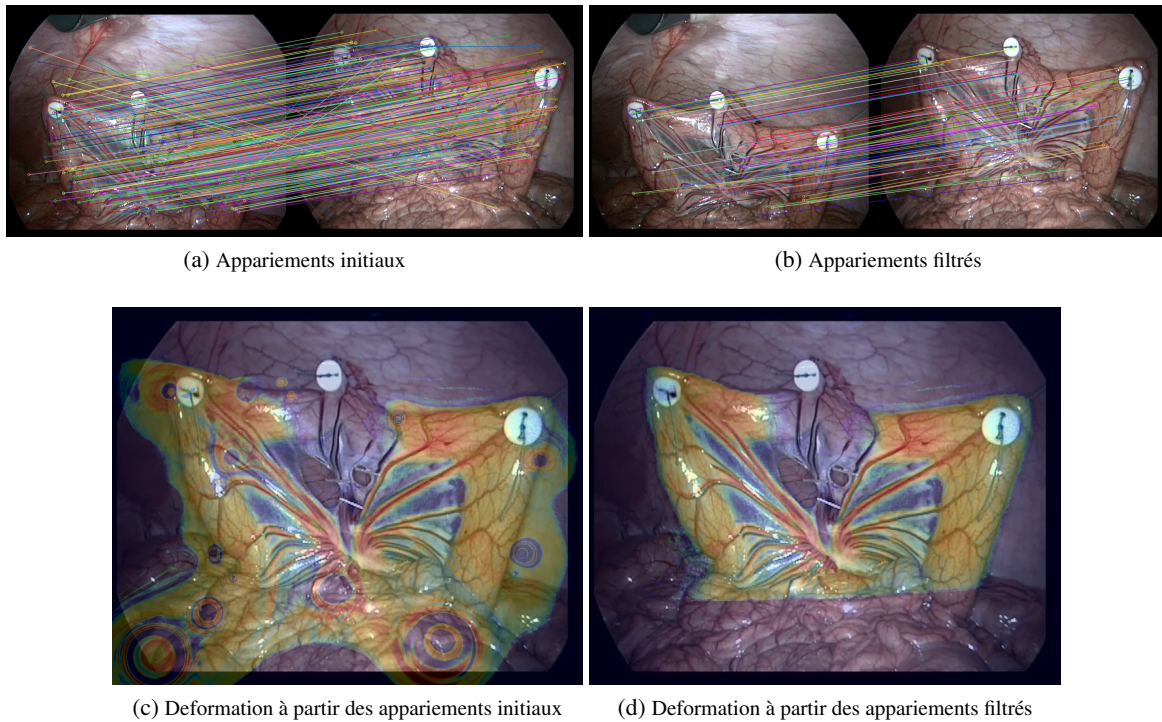


FIGURE 5.7 – Influence de notre étape de filtrage des appariements sur la qualité de la déformation de l’image de perfusion.

courante (Fig.5.6.e). La transformation de chaque classe (affine ou similitude ou rigide) $I_v(x) : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ est calculée pour chaque point v selon le critère de minimisation suivant :

$$\sum_{p_i \in P, q_i \in Q} w_i |I_v(p_i) - q_i|^2 \quad (5.1)$$

où w_i est un poids égal à : $w_i = \frac{1}{|p_i - v|^{2\alpha}}$ avec $\alpha \simeq 0.5$ et p_i, q_i les points appariés.

Enfin, un champ de déplacement dense est obtenu à partir des points appariés que nous utilisons pour déformer et recalibrer la carte de perfusion (Fig.5.6.f). Ce recalage est maintenu durant toute la définition des sites de résection par le chirurgien (cf. Fig. 5.11).

5.4 Évaluation du recalage de la carte de perfusion

La performance de notre méthode a été évaluée sur un ensemble de données *in vivo* de côlon de cochon. Nous avons effectué nos expérimentations sur 5 cochons différents. L’ensemble des vidéos fournissent des conditions variables telles que des translations, rotations de la caméra, déformations du tissu (mouvements péristaltiques, interaction des instruments chirurgicaux), changements des conditions d’illuminations et des occlusions. La caméra utilisée est une D-Light P de Storz avec une résolution de 768×576 à $25fps$ qui permet de basculer entre la vue normale et la vue fluoroscopique par l’utilisation de filtres spécifiques (avec un retard de commutation d’une seconde). La méthode de

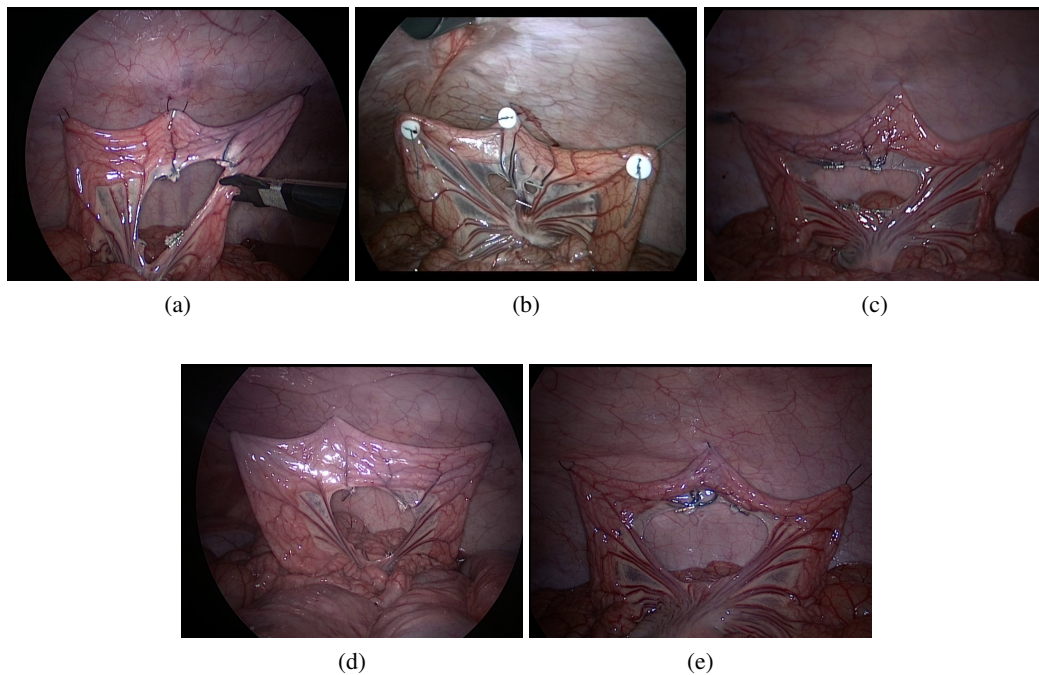


FIGURE 5.8 – Exemples d’images des différentes séquences utilisées dans nos expériences.

déformation MLS a été implémentée sans aucune parallélisation ni optimisation du code. Nous avons utilisé des grilles composées de (144×115) sommets sur lesquels la déformation est appliquée au lieu d’appliquer la déformation sur chaque pixel afin de réduire le temps de calcul. Nous utilisons ensuite une interpolation bilinéaire pour combler les quads résultant comme dans [SMW06]. Calculer le champ de déformation sur chaque pixel de l’image, augmente la précision en moyenne de 2%, mais ralentit considérablement le temps de calcul et donc ne permet pas de satisfaire les conditions cliniques.

Nous décrivons d’abord dans la section 5.4.1 comment nous obtenons les données vérité terrain pour évaluer notre méthode de recalage et présentons par la suite l’évaluation de la précision du recalage. En second lieu, nous présentons dans la section 5.4.2 les temps de calculs de chaque classe de fonctions de la méthode MLS.

5.4.1 Évaluation de la précision

Les données vérité terrain pour notre évaluation sont obtenues manuellement en sélectionnant dix points sur le côlon (tissu d’intérêt) qui sont suivis sur dix images acquises longtemps après l’image de référence. Ces dix images sont espacées en moyenne d’un délai de 5 secondes ce qui implique d’importants mouvements de la caméra et du tissu. Nous veillons à ce que les points cliqués soient relativement éloignés (distance minimale de 20 pixels) des points d’intérêts appariés lors de notre recalage afin de ne pas biaiser l’évaluation. Il est à noter que ces points sélectionnés ne sont pas forcément des points d’intérêts. Cela nous permet d’évaluer quantitativement le recalage non-rigide fourni par notre méthode et de montrer son intérêt sur le côlon. La figure 5.9 montre un exemple des points vérité

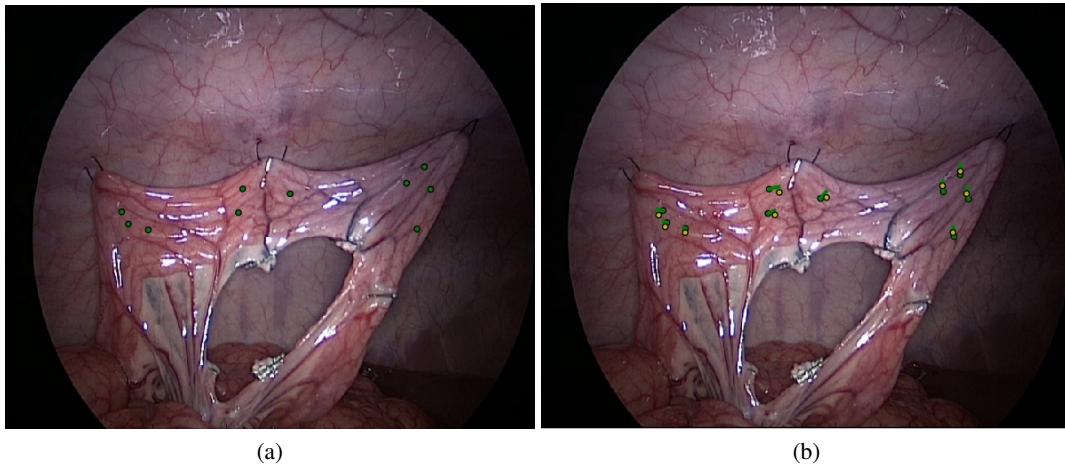


FIGURE 5.9 – Évaluation du recalage. (a) avant déformation, (b) après déformation. Points verts : position vérité terrain, points jaunes : les points estimés à partir d'un MLS rigide, ligne verte : le déplacement vérité terrain.

terrain (en vert) suivis sur deux images différentes et les points estimés à l'aide de champ de déformation calculé à partir d'un MLS rigide. L'erreur de recalage correspond à la distance euclidienne entre le déplacement des points vérité terrain et le champ de déformation calculé au coordonnées des points sélectionnés à l'image correspondante. Nous soulignons que six pixels correspondent approximativement à 1 mm, ce qui correspond à l'épaisseur des vaisseaux du côlon mesurée avec une règle au cours de la procédure chirurgicale. Dans le tableau 5.1 nous indiquons les erreurs de recalage obtenues en utilisant trois méthodes MLS différentes (affine, rigide et de similarité) sur une séquence vidéo (Fig. 5.8b). Pour chaque point annoté, nous donnons l'erreur moyenne et le déplacement moyen des points sur dix images. Nous remarquons que pour cette séquence la méthode rigide MLS donne le résultat de recalage le plus précis, qui reste inférieur à 1 mm.

Dans le tableau 5.2 nous fournissons l'erreur moyenne de recalage des dix points sur les dix images et selon les trois méthodes pour les différentes séquences vidéos. La tolérance d'erreur pour le chirurgien pour ce type d'intervention est d'environ 5 mm [Dia14b] : notre première évaluation *in vivo* montre clairement que notre approche a le potentiel de satisfaire les exigences cliniques. De plus, nous notons que la méthode MLS rigide fournit en moyenne des résultats légèrement meilleurs, du fait qu'il n'est pas observé un changement de zoom significatif dans les séquences.

5.4.2 Évaluation du temps de calcul

Le système proposé (cf. Fig. 5.11) a été mis en œuvre sur un PC avec un processeur Intel Core I7 - 3.4 GHz et 8 Go de RAM. Le code implémenté en c++ sous environnement Windows. Dans la figure 5.10 nous fournissons le temps de calcul pour chaque fonction de la méthode MLS et ceci pour un nombre différent de points d'intérêt appariés utilisés pour recalculer l'image. Nous pouvons voir que le nombre de points appariés utilisés pour calculer la déformation influe sur le temps de calcul. Nous remarquons aussi que la transformation MLS affine donne le meilleur résultat en termes de temps de calcul (environ 0.18s pour environ 200 points). Cependant, le temps de calcul peut être

TABLE 5.1 – Erreur moyenne du recalage et son écart-type des méthodes MLS affine, rigide et de similitude (annotée Simil) pour une séquence (Fig.5.8b).

Points	pt 0	pt 1	pt 2	pt 3	pt 4	pt 5	pt 6	pt 7	pt 8	pt 9
Déplacement (pixels)	47.3 ±40.3	46.5 ±40.4	47.2 ±41.6	46.9 ±41.1	51.1 ±43.2	52.8 ±44.7	47.8 ±40.7	48.5 ±41.6	47.4 ±40.5	48.2 ±41.1
Méthode	MLS affine									
Erreur (pixels)	2.6 ± 2.2	4.5 ± 3.7	2.9 ± 1	4.5 ± 3.9	2.8 ± 2.2	6.1 ± 6.5	4.1 ± 2.2	3.1 ± 2.4	2.4 ± 1.6	4.5 ± 3.7
Méthode	MLS rigide									
Erreur (pixels)	1.6 ± 1.2	1.2 ± 1.7	1.1 ± 0.4	1.4 ± 1.8	2.6 ± 0.6	4.7 ± 1.9	1.6 ± 1.2	1.1 ± 0.5	2.7 ± 2.7	3.6 ± 4.8
Méthode	MLS Simil									
Erreur (pixels)	1.5 ± 1.6	4.3 ± 3.4	2.4 ± 0.7	4.3 ± 4.1	2.5 ± 2.1	6.1 ± 6.6	3.2 ± 2.3	3.1 ± 2.6	2.4 ± 1.6	4.1 ± 4

TABLE 5.2 – Erreur moyenne du recalage et son écart-type des méthodes MLS affine, rigide et de similitude pour chaque séquence .

Videos	Fig.5.8a	Fig.5.8b	Fig.5.8c	Fig.5.8d	Fig.5.8e	Moyenne
Affine	3.7 ± 0.9	3.7 ± 1.4	3.4 ± 1.1	6.2 ± 1.3	1.7 ± 0.6	3.7 ± 0.6
Rigide	3.9 ± 0.4	2.1 ± 1.6	3.4 ± 0.7	6.1 ± 1.1	1.9 ± 0.6	3.4 ± 0.4
Simil	3.6 ± 0.7	3.3 ± 1.5	3.1 ± 1.1	6.3 ± 1.2	1.6 ± 0.5	3.5 ± 0.5

réduit par une meilleure optimisation du code. Nous soulignons qu'en moyenne le nombre de points d'intérêt appariés utilisés pour notre expérience est d'environ 240 points où l'étape de suivi proposé prend 0.07s, ce qui inclut la détection des points d'intérêt, le calcul des descripteurs et le calcul des appariements. Le temps de calcul est donc principalement dû au calcul de la déformation. Le logiciel fournit un recalage de cinq images par seconde, ce qui satisfait les besoins des chirurgiens. Toutefois, il serait préférable de réduire ce temps de calcul en optimisant le code, en le parallélisant par exemple.

5.5 Synthèse

Dans ce chapitre, nous avons proposé une application visant à fournir un système de réalité augmentée qui aide les chirurgiens à localiser le risque d'ischémie durant les procédures d'anastomose, en superposant une carte de perfusion directement sur la séquence vidéo courante. Notre méthode repose sur le suivi de points d'intérêt pour recalibrer de manière non-rigide la carte de probabilité en utilisant la transformation Moving Least Squares sur une vidéo endoscopique. Nous avons montré que la précision et le temps de calcul des méthodes proposées satisfont grandement les contraintes cliniques. Nos futures recherches se concentrent sur l'évaluation de notre application sur des données d'humains et étendre notre application en utilisant le guidage visuel basé sur la fluorescence à d'autres pathologies comme l'infarctus intestinal.

L'arrivée sur le marché d'autres caméras comme le SPIES de Storz offrant les deux vues (lumière blanche et fluorescence) en même temps, permet de faire évoluer l'application. En effet, durant le calcul de la carte de perfusion nous bloquons la respiration et nous supposons que l'environnement

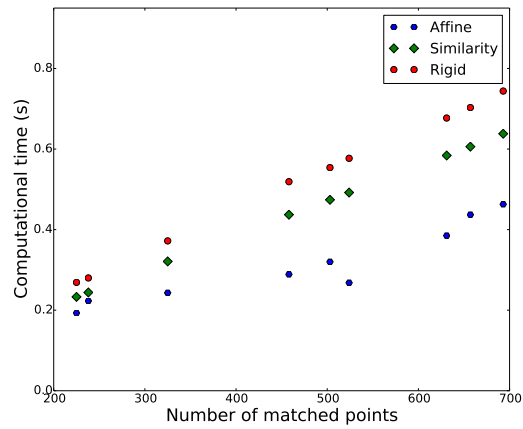


FIGURE 5.10 – Temps de calcul en fonction du nombre de points appariés utilisé pour calculer le recalage basé sur la méthode MLS (affine, similitude et rigide).

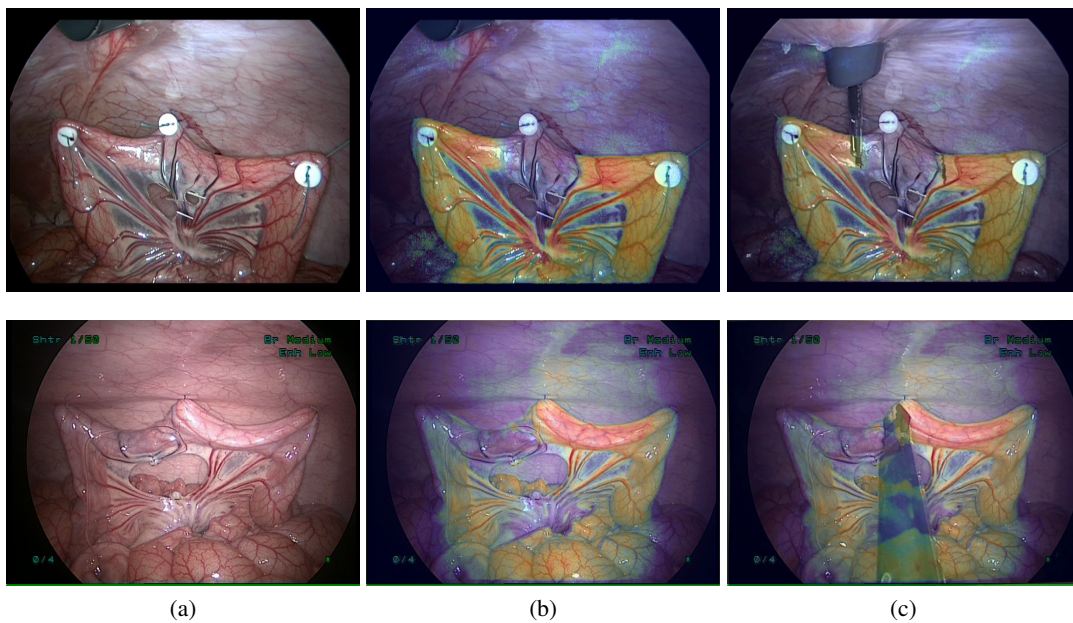


FIGURE 5.11 – Illustration d’une vue en réalité augmentée par notre système sur de nombreuses images qui subissent d’importantes déformations et des occlusions. a. Image de référence, b. Superposition de la carte de perfusion sur l’image de référence, c. Recalage non-rigide de la carte de perfusion

est statique. Avec les deux vues, il est possible d'effectuer un suivi de la déformation des tissus sur la vue normale et de compenser les mouvements dans la vue en fluorescence afin de calculer la carte de perfusion.

CONCLUSION

Cette thèse fut le fruit de trois années de travail au sein de l'institut de recherche contre le cancer de l'appareil digestif à Strasbourg. Mes travaux de recherche avaient pour but de développer un outil d'assistance en chirurgie laparoscopique à partir des images de la caméra endoscopique. En particulier, un outil basé sur la réalité augmentée permettant le guidage visuel du geste chirurgical. Les systèmes de guidage reposent principalement sur le suivi des organes d'intérêt qui doit être robuste et précis. Ce suivi est généralement mis en œuvre par une première étape de détection de points d'intérêt dans l'image, accompagné par une étape de suivi de points ("feature tracking"). Ces points d'intérêts correspondent à des caractéristiques visuelles facilement identifiables, caractérisés par un changement ou une discontinuité du gradient.

Face au choix d'une méthode de suivi parmi celles existantes, nous avons été limités par le manque d'information concernant leur performance. En particulier, il était difficile d'évaluer le suivi, du fait de l'absence de vérité terrain. La littérature mettait en évidence le manque de méthodologie permettant de choisir une technique de suivi adaptée à notre problème. Nous avons donc choisi dans un premier temps d'axer notre travail de recherche sur une méthode d'évaluation rigoureuse permettant de comparer différentes approches de suivi et donc de déterminer celle qui pouvait convenir à notre contexte.

Par ailleurs, nous avons rappelé qu'il est difficile de garantir un suivi de points robuste dans le cas des images en chirurgie laparoscopique. Les performances des techniques de suivi classiques n'étant pas toujours satisfaisantes, en particulier au niveau de la robustesse et de la répartition des points d'intérêt dans l'image, nous nous sommes intéressés dans un second temps à améliorer les performances de suivi.

Finalement, nous montrons sur une application médicale comment notre méthode peut être utilisée afin d'améliorer les performances de suivi, et fournir une information de guidage en réalité augmentée au chirurgien.

6.1 Contributions

Plusieurs contributions ont été fournies durant ce travail, elles se décomposent en trois sections.

6.1.1 Contributions méthodologiques

Tout d'abord, nous avons mis en place une méthodologie d'évaluation des performances de suivi, qui permet de comparer d'une manière exhaustive et pertinente, différents algorithmes de détection et de suivi. Cette méthodologie repose sur une nouvelle stratégie de suivi basée sur un suivi avant-arrière en utilisant les images paires-impaires, qui permet de fournir une vérité terrain automatique pour chaque scène.

Par la suite, nous avons proposé une approche spécifique au contexte du suivi qui repose essentiellement sur notre méthodologie de validation du suivi. À partir des premières images de la séquence vidéo, nous trouvons la meilleure méthode de suivi qui garantit la meilleure performance de suivi pour la séquence visualisée. La persistance et la pertinence de notre approche a été démontré sur différentes séquences vidéos.

6.1.2 Contributions expérimentales

Nous avons tout d'abord mis en évidence la pertinence de l'utilisation de toute l'information R,G,B au lieu de l'image en niveau de gris, car elle permet d'exploiter la couleur des tissus qui influe grandement sur le gradient de l'image. En effet, nous avons pu observer que le suivi de points sur le canal vert, offre une meilleure qualité du suivi par rapport aux autres canaux et de l'image en niveau de gris, du fait que le gradient est plus fort dans ce canal-là. De ce fait, nous avons ensuite dirigé nos travaux vers l'étude des pré-traitements basés sur l'amélioration du gradient de l'image, permettant ainsi d'augmenter les performances de suivi. Ces pré-traitements sont appliqués en tenant compte des spécificités des images en chirurgie laparoscopique (bifurcation de vaisseau sanguin, texture et couleur des tissus, condition d'illumination,...) pour augmenter la qualité du suivi. L'évaluation quantitative de l'influence des pré-traitements sur le suivi a mis évidence qu'il est nécessaire d'adapter chaque pré-traitement pour la scène visualisée. Grâce à notre approche spécifique au contexte du suivi, nous sommes capable de trouver le meilleur pré-traitement en fonction de la scène visualisée.

6.1.3 Contributions médicales

Finalement, nous avons développé une application en réalité augmentée qui aide les chirurgiens à localiser le risque d'ischémie durant les procédures d'anastomose, en superposant une carte de perfusion directement sur la séquence vidéo courante. Cette carte de perfusion est calculée en fonction des propriétés de fluorescence du vert d'indocyanine (ICG). Grâce à notre système, le chirurgien peut facilement définir les zones ischémiques lors de la résection des parties cancéreuses du côlon, ce qui permet de réduire le risque de morbidité post-opératoire chez le patient. Nous avons montré à travers nos évaluations, que les contraintes cliniques sont grandement satisfaites en termes de précision et de temps de calcul.

6.2 Perspectives

Nous discutons dans cette partie de plusieurs perspectives complémentaires à nos travaux, qui sont décomposées en trois sections.

6.2.1 Perspectives théoriques

Nous avons proposé une approche spécifique au contexte du suivi, qui permet de trouver la meilleure combinaison d'algorithmes pour la scène visualisée. Toutefois, l'ajout d'un mécanisme qui adapte des zones de recherches dans l'image, s'il y a présence de plusieurs types de tissus dans la séquence vidéo, garantit une identification du meilleur algorithme de suivi en fonction de chaque zone dans l'image. Ces zones de recherches peuvent être définies par une segmentation des tissus en analysant la couleur et la saturation des tissus. Cette analyse peut se faire en utilisant les deux composantes H (hue) et S (saturation) de l'espace de couleur HSV ou bien par des combinaisons des canaux afin de séparer les différents tissus en fonction de l'absorption de la lumière.

6.2.2 Perspectives pratiques

À partir de nos évaluations des pré-traitements sur les performances de suivi, nous avons remarqué que chaque pré-traitement influe différemment en fonction du tissu et/ou de la scène. Néanmoins, trouver une corrélation du choix des pré-traitements pour chaque type de tissus nécessite en effet un plus grand nombre de données. De plus, une classification des types de tissus semble pertinente pour fournir une analyse sur la corrélation du choix des pré-traitements en fonction des tissus. Par exemple, l'approche proposée par Quelhas et al. [QO06] considère la construction d'un dictionnaire qui décrit l'image à partir de ses caractéristiques visuelles locales. De cette manière nous pourrions associer chaque pré-traitement à un type de tissu.

Concernant notre application en réalité augmentée, notre évaluation a été limitée à l'animal, car actuellement il n'est pas possible d'utiliser l'ICG en France dans un tel cadre opératoire chez l'humain. Une procédure d'autorisation est en cours, et de ce fait il faudra certainement l'adapter aux données de l'homme.

En général, un travail complémentaire qui est l'optimisation des méthodes proposées, reposant sur une optimisation du code avec les techniques de parallélisation et d'accélération GPU/CPU, mettra à profit les travaux réalisés.

6.2.3 Perspectives médicales

Les propriétés de l'ICG pour l'étude de la vascularisation, permettent d'aboutir à plusieurs applications pour le guidage basé sur la fluorescence à d'autres pathologies comme l'infarctus intestinal. L'idée repose sur le même principe qui est le calcul d'une carte de perfusion, et d'un recalage dynamique sur le tissu d'intérêt en se basant sur le suivi de points.

La version actuelle de l'application repose sur la caméra D-Light P de Storz qui permet de basculer entre la vue normale et la vue fluoroscopique. Durant le calcul de la carte de perfusion nous bloquons la respiration et nous supposons que l'environnement est statique. Avec l'arrivée sur le marché de

nouvelles caméras et dispositifs comme le SPIES de Storz pour l'étude la fluorescence, le calcul de la carte de perfusion peut être effectué dynamiquement.

ANNEXE - A**A.1 Espace de couleur RGB**

L'espace RGB est l'espace engendré par les 3 composantes primaires (Rouge, Vert et Bleu). Cet espace de couleur correspond à la façon dont les couleurs sont généralement codées informatiquement à partir des capteurs CCD. Étant donné que le codage RGB repose sur trois composantes proposant la même gamme de valeur (0 à 255), l'ensemble des couleurs produites se représente comme l'intérieur d'un cube dont chacun des axes correspond à une couleur primaire. La diagonale du blanc au noir représente les niveaux de clarté (cf. Fig. A.1)

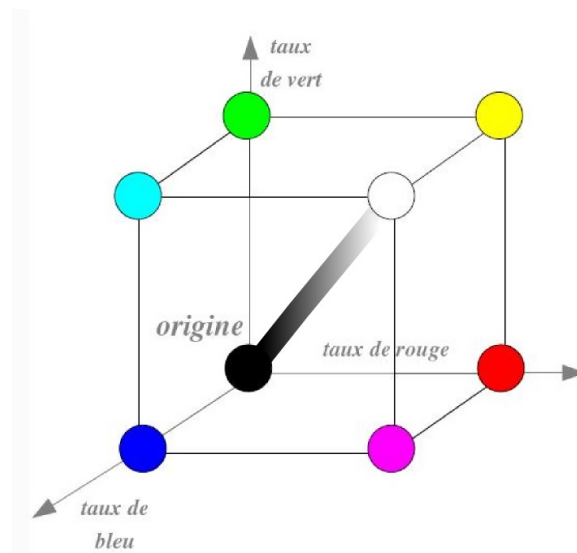


FIGURE A.1 – Espace de couleur RGB.

A.2 Espace de couleur HSV

Espace de couleur HSV, est un espace colorimétrique, proche du modèle de la perception physiologique de la couleur par l'œil humain. L'espace HSV est issue d'une transformation non linéaire de l'espace de couleur RGB, défini en fonction de ses trois composantes (cf. Fig. A.2) :

1. Hue (teinte), correspondant à la perception de la couleur,
2. Saturation (saturation), décrivant la pureté de la couleur, c'est-à-dire son caractère vif ou pâle,
3. Value (valeur), indiquant la quantité de lumière de la couleur, c'est-à-dire son aspect clair ou sombre.

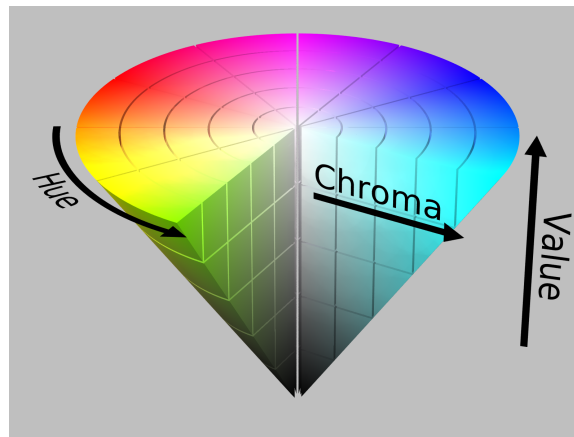


FIGURE A.2 – Espace de couleur HSV.

L'espace HSV contient toutes les couleurs, ce qui permet à chaque couleur de l'espace RGB d'avoir une correspondance HSV. En revanche, la conversion n'est pas réversible, c'est-à-dire que les teintes HSV ne sont pas toutes représentées dans l'espace RGB.

Soit $h \in [0, 360]$ et $s, v, r, g, b \in [0, 1]$

r, g, b désignent respectivement les coordonnées RGB.

h, s, v désignent respectivement les coordonnées HSV.

M , désigne respectivement le max (r,g,b).

m , désigne respectivement le min (r,g,b).

La conversion de l'espace RGB vers le HSV et vice-versa est décrite ci-dessous.

A.2.1 conversion RGB HSV

$$h = \begin{cases} 0, & \text{si } M = m \\ (60^\circ \times \frac{g-b}{M-m} + 360^\circ) \bmod 360^\circ, & \text{si } M = r \\ 60^\circ \times \frac{b-r}{M-m} + 120^\circ, & \text{si } M = g \\ 60^\circ \times \frac{r-g}{M-m} + 240^\circ, & \text{si } M = b \end{cases}$$

$$s = \begin{cases} 0, & \text{si } M = 0 \\ 1 - \frac{m}{M}, & \text{sinon} \end{cases}$$

$$v = M$$

A.2.2 conversion HSV RGB

$$\begin{aligned}
 h_i &= \left\lfloor \frac{t}{60} \right\rfloor \bmod 6 \\
 f &= \frac{t}{60} - h_i \\
 l &= v \times (1 - s) \\
 m &= v \times (1 - f \times s) \\
 n &= v \times (1 - (1 - f) \times s) \\
 (r, g, b) &= \begin{cases} (v, n, l), & \text{si } h_i = 0 \\ (m, v, l), & \text{si } h_i = 1 \\ (l, v, n), & \text{si } h_i = 2 \\ (l, m, v), & \text{si } h_i = 3 \\ (n, l, v), & \text{si } h_i = 4 \\ (v, l, m), & \text{si } h_i = 5 \end{cases}
 \end{aligned}$$

A.3 Espace de couleur HSL

Espace de couleur HSL, est un espace calorimétrique, proche de celui du HSV, il est basé sur les trois grandeurs de Hue (Teinte), Saturation (Saturation) et Lightness (Luminance). La Figure A.3 donne une représentation de cet espace. Alors que dans l'espace HSV la valeur est définie par la plus grande valeur des composantes de couleur RGB. Dans l'espace HSL la luminance est la moyenne de la plus grande et de la plus petite valeur des composantes de couleur RGB.

$$L = 0.5 * \max(R, G, B) + 0.5 * \min(R, G, B)$$

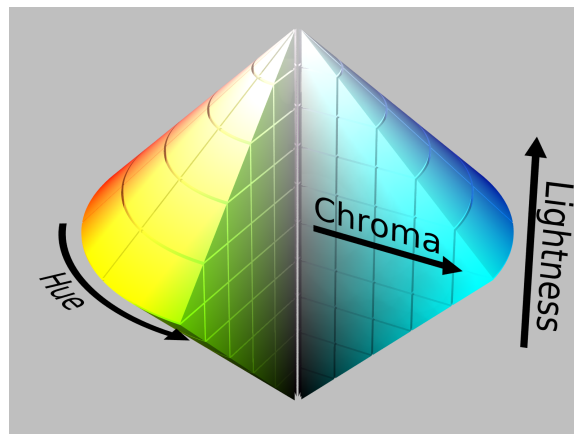


FIGURE A.3 – Espace de couleur HSV.

ANNEXE - B

B.1 Les opérateurs morphologiques en traitement d'image

La morphologie mathématique est un outil particulièrement utile, qui permet l'analyse des formes spatiales dans l'image. La morphologie mathématique fournit plusieurs applications dans le domaine du traitement d'image notamment, le filtrage, la segmentation et l'analyse des formes. Initialement, elle est destinée au traitement des images binaires et s'appuie sur différents opérateurs. La dilatation et l'érosion sont les opérateurs de base de la morphologie mathématique. Pratiquement tous les autres (ouverture, fermeture, chapeau haut de forme,...) peuvent être définis à l'aide de ceux-ci, en utilisant des compositions de fonctions et des opérations ensemblistes.

B.1.1 Élément structurant

Un élément structurant est un masque binaire (constitué de pixels blancs et noirs) muni d'un point d'ancrage. Dans la figure, nous montrons, à titre d'exemple, quatre éléments structurants, le point d'ancrage étant marqué par un point.

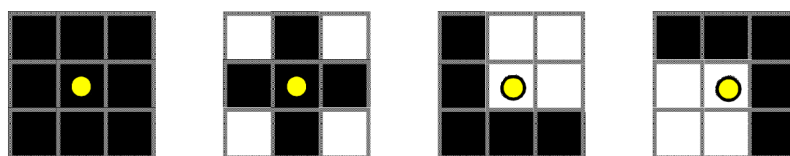


FIGURE B.1 – Exemples des éléments structurants, ou le point jaune représente le point d'ancrage.

B.1.2 Dilatation

Une dilatation morphologique consiste à déplacer l'élément structurant sur chaque pixel de l'image, et à regarder si l'élément structurant intersecte avec la structure d'intérêt. Le résultat est une structure qui plus grosse que la structure d'origine (Figure.B.2). En fonction de la taille de l'élément structurant, certaines particules peuvent se trouver connectées, et certains trous disparaître. Soit une image binaire I et un élément structurant B . La dilatation de I par B est définie par :

$$\delta_B(I) = \{x \in R^2, B_x \cap I \neq \emptyset\} \quad (\text{B.1})$$

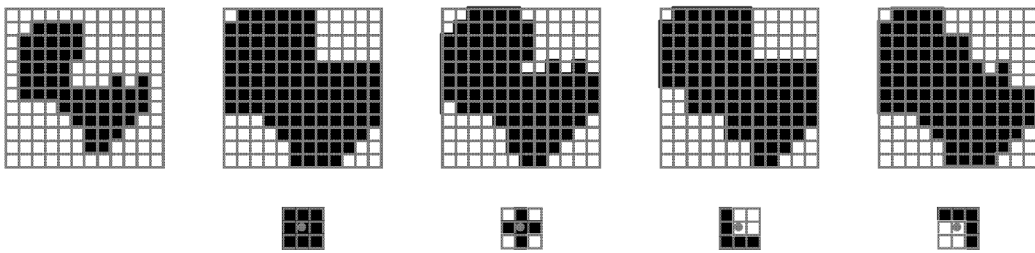


FIGURE B.2 – Illustration d'une dilatation selon 4 types d'éléments structurants.

B.1.3 Érosion

L'érosion est l'opération inverse d'une dilatation du complémentaire de la structure. Elle consiste à chercher tous les pixels pour lesquels l'élément structurant centré sur ce pixel touche l'extérieur de la structure d'intérêt. Le résultat est une structure rognée (Figure. B.3). En fonction de l'élément structurant utilisé, il est observé la disparition des structures de taille plus petites, et la séparation éventuelle des structures de taille supérieure. Soit une image binaire I et un élément structurant B . L'érosion de I par B est définie par :

$$\varepsilon_B(I) = \{x \in R^2, B_x \subseteq I\} \quad (\text{B.2})$$

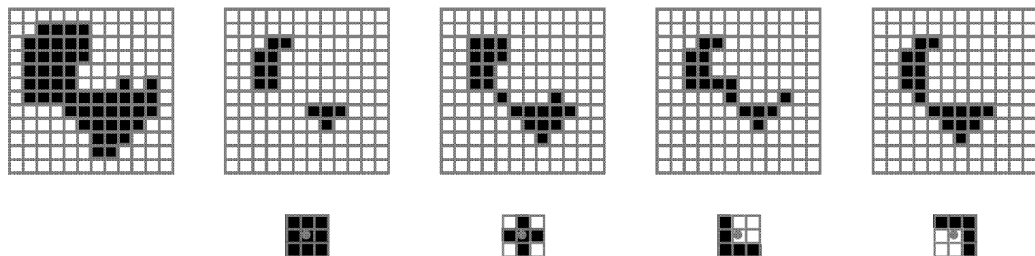


FIGURE B.3 – Illustration d'une érosion selon types d'éléments structurants.

B.1.4 Ouverture

L'ouverture morphologique est obtenue par la succession d'une érosion et d'une dilatation. Elle a pour effet d'éliminer les objets de taille inférieure à celle de l'élément structurant. La structure résultante est plus régulière et moins riche en détails que l'ensemble initiale. Une ouverture (γ) d'une image I avec un élément structurant B est définie par l'érosion de I par B suivi par une dilatation avec l'élément structurant \check{B} symétrique à B .

$$\gamma_B(I) = \delta_{\check{B}}[\varepsilon_B(I)] \quad (\text{B.3})$$

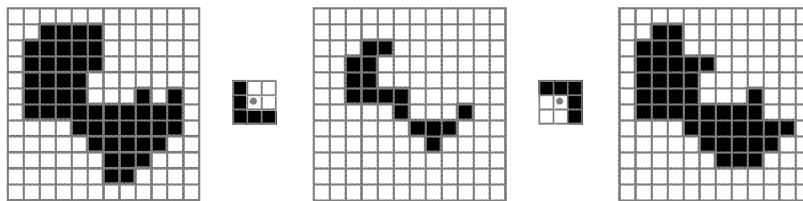


FIGURE B.4 – Illustration d'une ouverture par un élément structurant en «L».

B.1.5 Fermeture

La fermeture est obtenue par la succession d'une dilatation et d'une érosion. Elle a pour effet de boucher les trous des objets qui sont plus petits que l'élément structurant. Une fermeture (ϕ) d'une image I avec un élément structurant B est définie par la dilatation de I par B suivi par l'érosion avec l'élément structurant \check{B} symétrique à B .

$$\phi_B(I) = \varepsilon_{\check{B}}[\delta_B(I)] \quad (\text{B.4})$$

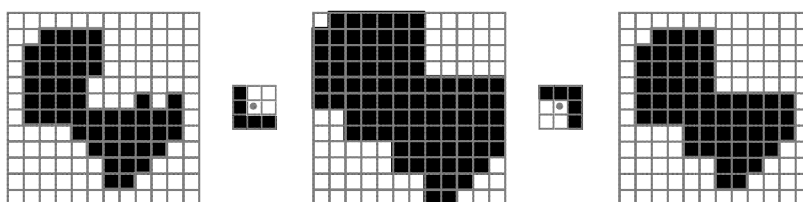


FIGURE B.5 – Illustration d'une fermeture par un élément structurant en «L».

B.1.6 Chapeau haut de forme

La transformation chapeau haut de forme (top-hat) peut se diviser en : chapeau haut de forme noir (black top-hat) et en chapeau haut de forme blanc (white top-hat). Le chapeau haut de forme noir

(BTH) permet la détection des zones foncées des images. Il correspond à la différence d'une fermeture d'une image $\phi(I)$ à l'image originale I .

$$BTH = \phi(I) - I \quad (\text{B.5})$$

Le chapeau haut de forme blanc (WTH) permet de détecter ce que l'ouverture a fait disparaître, c'est-à-dire les pics ou parties claires de l'image originale. Il correspond à la différence d'une image originale I à son ouverture $\gamma(I)$.

$$WTH = \phi(I) - I \quad (\text{B.6})$$

ANNEXE - C**C.1 Séquences vidéos utilisées**

Nous donnons les séquences *in vivo* sur lesquelles nous avons effectué nos évaluations. Toutes les séquences vidéos sont acquises avec une caméra Storz Image1 Hub Tri-CCD FHD (1920 × 1080) à 25 *fps* avec une lentille endoscopique à 0° degré, sauf susmentionné. À partir des séquences décrites ci-dessous, nous les avons découpés en sous-séquences en moyenne de 180 images afin de réaliser nos évaluations. Ceci correspond approximativement à 7 secondes.

C.1.1 Foie d'humain**C.1.1.1 Patient - A**

La séquence C.1 correspond à une exploration abdominale, visualisant un foie humain. Le chirurgien effectue un mouvement de balayage panoramique aller-retour avec une vitesse constante. Aucune occlusion ou interaction des instruments avec les tissus n'est observée. Elle contient 1025 images.



FIGURE C.1 – Extraits de séquence foie humain Patient - A.

C.1.1.2 Patient - B

La séquence C.2 avec une résolution de (854 × 480), elle correspond à une exploration abdominale visualisant un foie humain accompagné d'une dégénérescence des tissus hépatiques due à la cirrhose. La caméra bouge autour du foie avec des mouvements de translations et de rotations. Il est observé de la fumée et des saignements dû à la cautérisation des tissus et des occlusions provoquées par les instruments. Elle contient 850 images.

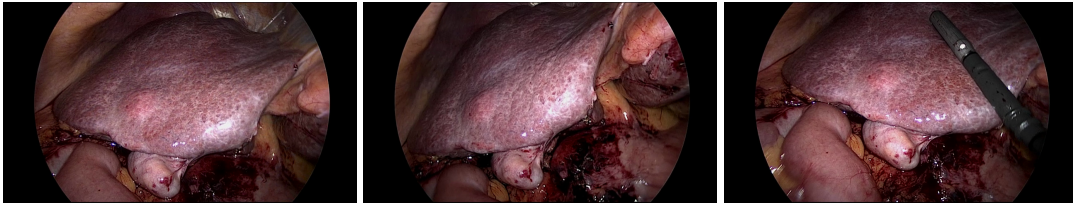


FIGURE C.2 – Extraits séquence foie humain Patient - B.

C.1.1.3 Patient - C

La séquence C.3 correspond à une exploration abdominale, visualisant un foie humain. Le chirurgien effectue des mouvements de rotations et de translations autour du foie. Il est observé des occlusions provoquées par les instruments. Elle contient 675 images.

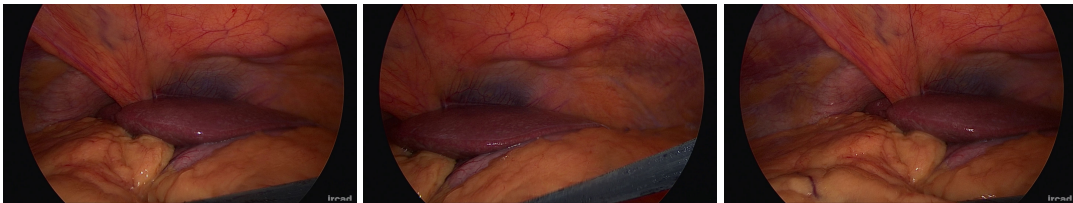


FIGURE C.3 – Extraits séquence foie humain Patient - C.

C.1.1.4 Patient - D

La séquence C.4 correspond à une cholécystectomie chez un patient humain. La caméra effectue des mouvements de rotations et de translations autour du foie. Il est observé des occlusions provoquées par les instruments ainsi que de la fumée et des saignements dû à la cautérisation des tissus. Elle contient 3725 images.

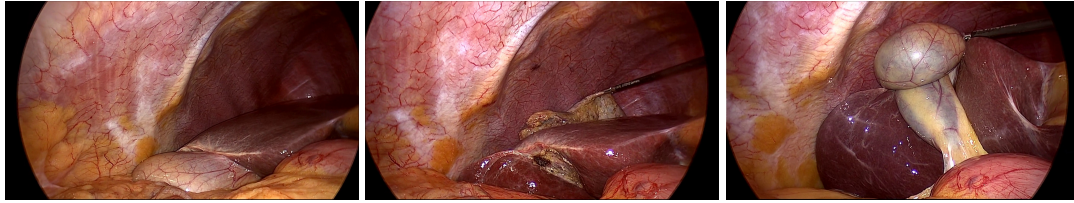


FIGURE C.4 – Extraits séquence foie humain Patient - D.

C.1.2 Foie et rate de cochon

C.1.2.1 Foie

La séquence C.5 correspond à une exploration abdominale chez un cochon, visualisant son foie. Le chirurgien effectue des mouvements de rotations et de translations autour du foie. Une occlusion partielle est observée. Aucune interaction des instruments avec les tissus n'est réalisée. Elle contient 6000 images.

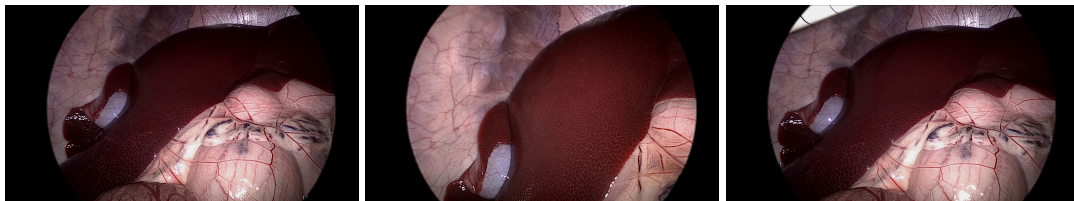


FIGURE C.5 – Extraits séquence foie de cochon.

C.1.2.2 Rate

La séquence C.5 correspond à une exploration abdominale chez un cochon, visualisant la rate posée sur foie. Le chirurgien effectue des mouvements de rotations et de translations autour de la rate. Un grand mouvement de glissement du foie sur la rate est observé. Aucune occlusion ou interaction des instruments avec les tissus n'est observée. Elle contient 675 images.

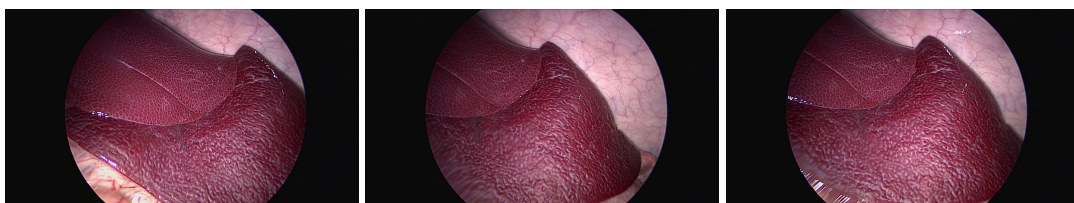


FIGURE C.6 – Extraits séquence rate cochon.

C.1.3 Intestins de cochon

Les deux séquences C.7 (625 images) et C.8 (875 images) correspondent à des explorations abdominale chez deux cochons différents, visualisant les intestins avec un mouvement péristaltique important. Le chirurgien effectue des mouvements de balayage dans la cavité abdominale. Aucune occlusion ou interaction des instruments avec les tissus n'est observée.

C.1.3.1 cochon - A



FIGURE C.7 – Extraits séquence des intestins du cochon - A.

C.1.3.2 cochon - B

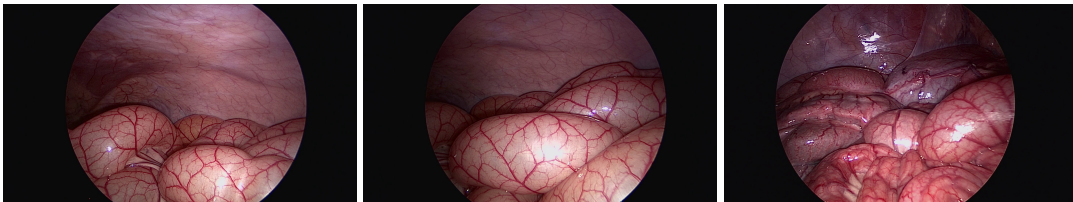


FIGURE C.8 – Extraits séquence des intestins du cochon - B.

BIBLIOGRAPHIE

- [AB12] PF Alcantarilla and Adrien Bartoli. Deformable 3D Reconstruction with an Object Database. *BMVC*, pages 133.1–133.12, 2012.
- [Adi85] Gilad Adiv. Determining three-dimensional motion and structure from optical flow generated by several moving objects. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (4) :384–401, 1985.
- [AGD07] IC Albitar, Pierre Graebbling, and Christophe Doignon. Robust structured light coding for 3d reconstruction, 2007.
- [AGG⁺09] Selen Atasoy, Ben Glocker, Stamatia Giannarou, Diana Mateus, Alexander Meining, Guang-Zhong Yang, and Nassir Navab. Probabilistic region matching in narrow-band endoscopy for targeted optical biopsy. *Medical Image Computing and Computer-Assisted Intervention*, 12 :499–506, January 2009.
- [AK08] Motilal Agrawal and Kurt Konolige. Censure : Center surround extremas for realtime feature detection and matching. pages 102–115, 2008.
- [AOV12] Alexandre Alahi, Raphael Ortiz, and P Vandergheynst. FREAK : Fast Retina Keypoint. *Computer Vision and . . .*, 2012.
- [Azu97] RT Azuma. A survey of augmented reality. *Presence-Teleoperators and Virtual Environments*, 4(August) :355–385, 1997.
- [Bak04] Simon Baker. Lucas-kanade 20 years on : A unifying framework. *International Journal of Computer Vision*, 56(3) :221–255, 2004.
- [Bay08] H Bay et. al. Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3) :346–359, June 2008.
- [BBPW04] Thomas Brox, Andrés Bruhn, Nils Papenberg, and Joachim Weickert. High accuracy optical flow estimation based on a theory for warping. In *Computer Vision-ECCV 2004*, pages 25–36. Springer, 2004.
- [BBS⁺10] Alexander Behrens, Michael Bommers, Thomas Stehle, Sebastian Gross, Steffen Leonhardt, and Til Aach. Real-time image composition of bladder mosaics in fluorescence endoscopy. *Computer Science - Research and Development*, 26(1-2) :51–64, September 2010.
- [BCBC12] Adrien Bartoli, Toby Collins, Nicolas Bourdel, and Michel Canis. Computer assisted Minimally Invasive Surgery : Is medical Computer Vision the answer to improving laparosurgery ? *Medical hypotheses*, 79(6) :858–63, December 2012.

- [BFB94] John L Barron, David J Fleet, and Steven S Beauchemin. Performance of optical flow techniques. *International journal of computer vision*, 12(1) :43–77, 1994.
- [BHN⁺12] J Bano, Alexandre Hostettler, SA Nicolau, Stephane Cotin, Christophe Doignon, HS Wu, MH Huang, Luc Soler, and Jacques Marescaux. Simulation of pneumoperitoneum for laparoscopic surgery planning. pages 91–98, 2012.
- [BM04] S. Baker and I. Matthews. Lucas-kanade 20 years on : A unifying framework. *International Journal of Computer Vision*, 56(3) :221–255, 2004.
- [Bou01] JY Bouguet. Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the algorithm. *In Practice*, 1(2) :1–9, 2001.
- [BTV06] Herbert Bay, Tinne Tuytelaars, and L. Van Gool. Surf : Speeded up robust features. *Computer Vision–ECCV 2006*, pages 404–417, 2006.
- [BZX12] Xiangzhi Bai, Fugen Zhou, and Bindang Xue. Toggle and top-hat based morphological contrast operators. *Computers & Electrical Engineering*, 38(5) :1196–1204, September 2012.
- [CB12] Toby Collins and Adrien Bartoli. 3D reconstruction in laparoscopy with close-range photometric stereo. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, 15(Pt 2) :634–42, January 2012.
- [CC11] Toby Collins and B Compte. Deformable Shape-From-Motion in Laparoscopy using a Rigid Sliding Window. *and Analysis (MIUA 11)*. *To appear*, pages 1–5, 2011.
- [CLH06] Hok-Kwok Choi, Wai-Lun Law, and Judy W C Ho. Leakage after resection and intra-peritoneal anastomosis for colorectal malignancy : analysis of risk factors. *Diseases of the colon and rectum*, 49(11) :1719–25, November 2006.
- [CLO⁺11] Michael Calonder, Vincent Lepetit, Mustafa Ozuysal, Tomasz Trzcinski, Christoph Strecha, and Pascal Fua. BRIEF : Computing a Local Binary Descriptor very Fast. *IEEE transactions on pattern analysis and machine intelligence*, 34(7) :1281–1298, November 2011.
- [CLSF10] Michael Calonder, Vincent Lepetit, Christoph Strecha, and Pascal Fua. BRIEF : Binary robust independent elementary features. In *European Conference on Computer Vision*, pages 778–792, 2010.
- [CPB⁺13] Toby Collins, Daniel Pizarro, Adrien Bartoli, Michel Canis, and Nicolas Bourdel. Real-time wide-baseline registration of the uterus in laparoscopic videos using multiple texture maps. In *Augmented Reality Environments for Medical Imaging and Computer-Assisted Interventions*, pages 162–171. Springer, 2013.
- [DATa] http://ilab.cs.ucsb.edu/tracking_dataset_ijcv/.
- [DATb] <http://www.robots.ox.ac.uk/vgg/research/affine/detectors.html>.
- [Dia14a] Michele Diana et. al. Enhanced-reality video fluorescence : A real-time assessment of intestinal viability. *Annals of surgery*, 2014.

- [Dia14b] Michele Diana et. al. Probe-based confocal laser endomicroscopy (cellvizio) and enhanced-reality for real-time assessment of intestinal microcirculation at the future anastomotic site. Accepted to SAGES, 2014.
- [EP11] Haytham Elhawary and A Popovic. Robust feature tracking on the beating heart for a robotic guided endoscope. *The international journal of medical robotics and computer assisted surgery*, (July) :459–468, 2011.
- [ESR07] Hashem B El-Serag and K Lenhard Rudolph. Hepatocellular carcinoma : epidemiology and molecular carcinogenesis. *Gastroenterology*, 132(7) :2557–2576, 2007.
- [Fer13] Jacques Ferlay et. al. GLOBOCAN 2012 v1.0 Cancer Incidence and Mortality Worldwide : IARC CancerBase No. 11, 2013.
- [GCM11] O.G. Grasa, Javier Civera, and JMM Montiel. EKF monocular SLAM with relocalization for laparoscopic sequences. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 4816–4821. IEEE, 2011.
- [GHT11] Steffen Gauglitz, Tobias Höllerer, and Matthew Turk. Evaluation of interest point detectors and feature descriptors for visual tracking. *International Journal of Computer Vision*, 94(3) :335–360, 2011.
- [Gou99] François Goulette. *Modélisation 3D automatique : outils de géométrie différentielle*. Presses des MINES, 1999.
- [GVSY12] Stamatia Giannarou, Marco Visentini-Scarzanella, and Guang-Zhong Yang. Probabilistic Tracking of Affine-Invariant Anisotropic Regions. *IEEE transactions on pattern analysis and machine intelligence*, pages 1–14, March 2012.
- [Han04] Allan Hanbury. The morphological top-hat operator generalised to multi-channel images. *International Conference on Pattern Recognition.*, pages 672–675, 2004.
- [Han11] JH Han et. al. A Novel 3-D Color Histogram Equalization Method With Uniform 1-D Gray Scale Histogram. *Image Processing, IEEE*, pages 506–12, 2011.
- [HDP13] Nazim Haouchine, Jeremie Dequidt, and Igor Peterlik. Image-guided simulation of heterogeneous tissue deformation for augmented reality during hepatic surgery. 2013.
- [HFB⁺08] Johann Hummel, Michael Figl, Michael Bax, Helmar Bergmann, and Wolfgang Birkfellner. 2d/3d registration of endoscopic ultrasound to ct volume data. *Physics in medicine and biology*, 53(16) :4303, 2008.
- [HMG⁺09] Xishi Huang, John Moore, Gerard Guiraudon, Douglas L Jones, Daniel Bainbridge, Jing Ren, and Terry M Peters. Dynamic 2d ultrasound and 3d ct image registration of the beating heart. *Medical Imaging, IEEE Transactions on*, 28(8) :1179–1189, 2009.
- [HNR⁺10] A Hostettler, SA Nicolau, Y Rémond, J Marescaux, and L Soler. A real-time predictive simulation of abdominal viscera positions during quiet free breathing. *Progress in biophysics and molecular biology*, 103(2) :169–184, 2010.

- [HPE⁺07] Mingxing Hu, Graeme Penney, Philip Edwards, Michael Figl, and David Hawkes. 3D reconstruction of internal organ surfaces for minimal invasive surgery. *Medical image computing and computer-assisted intervention : MICCAI*, 10(Pt 1) :68–77, January 2007.
- [HS81] Berthold K Horn and Brian G Schunck. Determining optical flow. In *1981 Technical Symposium East*, pages 319–331. International Society for Optics and Photonics, 1981.
- [HS88] Chris Harris and Mike Stephens. A combined corner and edge detector. In *In Proceedings of The Fourth Alvey Vision Conference (1988)*, pages 147–152, 1988.
- [HSN06] Mitsuhiro Hayashibe, Naoki Suzuki, and Yoshihiko Nakamura. Laser-scan endoscope system for intraoperative geometry acquisition and surgical robot safety management. *Medical Image Analysis*, 10(4) :509–519, 2006.
- [Hu 10] Mingxing. Hu et. al. Reconstruction of a 3D surface from video that is robust to missing data and outliers : Application to minimally invasive surgery using stereo and mono endoscopes. *Medical image analysis*, December 2010.
- [HZ03] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.
- [Iei11] S. Ieiri et. al. Augmented reality navigation system for laparoscopic splenectomy in children based on preoperative CT image using optical tracking device. *Pediatric surgery international*, December 2011.
- [JG09] Luo Juan and Oubong Gwun. A comparison of sift, pca-sift and surf. *International Journal of Image Processing (IJIP)*, 3(4) :143–152, 2009.
- [Kal10] Zdenek Kalal et. al. Forward-Backward Error : Automatic Detection of Tracking Failures. *International Conference on Pattern Recognition*, pages 2756–2759, August 2010.
- [KF10] S. Kodashima and M. Fujishiro. Novel image-enhanced endoscopy with i-scan technology. *World Journal of Gastroenterology : WJG*, 16(9) :1043, 2010.
- [KRSH10] Stefanie Kudszus, Christian Roesel, Alexander Schachtrupp, and Jörg J Höer. Intraoperative laser fluorescence angiography in colorectal surgery : a noninvasive analysis to reduce the rate of anastomotic leakage. *Langenbeck's archives of surgery / Deutsche Gesellschaft für Chirurgie*, 395(8) :1025–30, November 2010.
- [LCS11] Stefan Leutenegger, Margarita Chli, and Roland Y. Siegwart. BRISK : Binary Robust invariant scalable keypoints. *2011 International Conference on Computer Vision*, pages 2548–2555, November 2011.
- [Lin99] Tony Lindeberg. Principles for automatic scale selection. 1999.
- [LK81] Bruce D Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. *IJCAI*, 81 :674–679, 1981.
- [Low04] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2) :91–110, November 2004.

- [Mah12] Nader. Mahmoud et. al. Fast 3d structure from motion with missing points from registration of partial reconstructions. *Articulated Motion and Deformable Objects*, 7378 :173–183, 2012.
- [Mal12] Abed Malti et. al. Template-Based Conformal Shape-from-Motion-and-Shading for Laparoscopy. *Information Processing in Computer-Assisted*, pages 1–10, 2012.
- [MCUP04] J Matas, O Chum, M Urban, and T Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and vision computing*, 22(10) :761–767, September 2004.
- [MHTS⁺08] Lena Maier-Hein, Aysun Tekbas, Alexander Seitel, Frank Pianka, Sascha A Müller, Stefanie Satzl, Simone Schawo, Boris Radeleff, Ralf Tetzlaff, Alfred M Franz, et al. In vivo accuracy assessment of a needle-based navigation system for ct-guided radiofrequency ablation of the liver. *Medical physics*, 35(12) :5385–5396, 2008.
- [Miw10] Mitsuharu Miwa. The Principle of ICG Fluorescence Method. *Open Surgical Oncology Journal*, pages 26–28, 2010.
- [MJQR03] Calvin R Maurer Jr, Rensheng Qi, and Vijay Raghavan. A linear time algorithm for computing exact euclidean distance transforms of binary images in arbitrary dimensions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(2) :265–270, 2003.
- [Mou07] Peter. Mountney et. al. A probabilistic framework for tracking deformable soft tissue in minimally invasive surgery. *MICCAI*, 10(Pt 2) :34–41, January 2007.
- [MS05] Krystian Mikolajczyk and Cordelia Schmid. A performance evaluation of local descriptors. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(10) :1615–1630, 2005.
- [MSDY06] Peter Mountney, Danail Stoyanov, Andrew Davison, and Guang-Zhong Yang. Simultaneous stereoscope localization and soft-tissue mapping for minimal invasive surgery. *Medical image computing and computer-assisted intervention : MICCAI*, 9(Pt 1) :347–54, January 2006.
- [MTS⁺05] Krystian Mikolajczyk, Tinne Tuytelaars, Cordelia Schmid, Andrew Zisserman, Jiri Matas, Frederik Schaffalitzky, Timor Kadir, and Luc Van Gool. A comparison of affine region detectors. *International journal of computer vision*, 65(1-2) :43–72, 2005.
- [MY08] Peter Mountney and Guang-Zhong Yang. Soft tissue tracking for minimally invasive surgery : learning local deformation online. *Medical image computing and computer-assisted intervention : MICCAI*, 11(Pt 2) :364–72, January 2008.
- [MY09] Peter Mountney and Guang-Zhong Yang. Dynamic view expansion for minimally invasive surgery using simultaneous localization and mapping. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society.*, 2009 :1184–7, January 2009.
- [MY11] Peter Mountney and Guang-Zhong Yang. Context specific descriptors for tracking deforming tissue. *Medical image analysis*, 16(3) :550–561, May 2011.

- [Nic11] Stéphane Nicolau et. al. Augmented reality in laparoscopic surgical oncology. *Surgical oncology*, 20(3) :189–201, September 2011.
- [Par09] P Parisot. Suivi d’objets dans des séquences d’images de scènes déformables. 2009.
- [Pen09] Jochen Penne. Time-of-flight 3-d endoscopy. *Medical image computing and computer-assisted intervention*, pages 467–74, 2009.
- [PSM13a] G Puerto-Souza and G Mariottini. A Fast and Accurate Feature-Matching Algorithm for Minimally-Invasive Endoscopic Images. *IEEE transactions on medical imaging*, (c), January 2013.
- [PSM13b] Gustavo A Puerto-Souza and Gian Luca Mariottini. Toward long-term and accurate augmented-reality display for minimally-invasive surgery. pages 5384–5389, 2013.
- [QO06] Pedro Quelhas and JM Odobez. Natural scene image modeling using color and texture visterms. *Image and Video Retrieval*, pages 411–421, 2006.
- [RBS⁺11] Sebastian Roehl, Sebastian Bodenstedt, Stephan Suwelack, Hannes Kenngott, Beat P. Muller-Stich, Rudiger Dillmann, and Stephanie Speidel. Real-time surface reconstruction from stereo endoscopic images for intraoperative registration. In *SPIE Medical Imaging*, volume 7964, pages 796414(1–13), 2011.
- [Ros06] Edward Rosten. Machine learning for high-speed corner detection. *European Conference on Computer Vision, ECCV 2006*, pages 430–443, 2006.
- [RPD10] Edward Rosten, Reid Porter, and Tom Drummond. Faster and better : a machine learning approach to corner detection. *IEEE transactions on pattern analysis and machine intelligence*, 32(1) :105–19, January 2010.
- [RRKB11] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. ORB : An efficient alternative to SIFT or SURF. *2011 International Conference on Computer Vision*, pages 2564–2571, November 2011.
- [Sch] Cameron Schaeffer. A Comparison of Keypoint Descriptors in the Context of Pedestrian Detection : FREAK vs. SURF vs. BRISK. *cs229.stanford.edu*.
- [Sel] F. Selka et. al. Performance evaluation of simultaneous RGB analysis for feature detection and tracking in endoscopic images. In *Medical Image Understanding and Analysis. 2012*, pages 249–254.
- [SES12] Tobias Senst, Volker Eiselein, and T Sikora. Robust Local Optical Flow for Feature Tracking. *IEEE Transactions On Circuits And Systems For Video Technology*, 22(9) :1377–1387, 2012.
- [SEY09] Danail Stoyanov, Dan Elson, and Guang-Zhong Yang. Illumination position estimation for 3D soft-tissue reconstruction in robotic minimally invasive surgery. *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2628–2633, October 2009.
- [SMB00] Cordelia Schmid, Roger Mohr, and Christian Bauckhage. Evaluation of Interest Point Detectors. *International Journal of Computer Vision*, 37(2) :151–172, 2000.

- [SMD⁺05] Danail Stoyanov, George P. Mylonas, Fani Deligianni, Ara Darzi, and Guang-Zhong Zhong Yang. Soft-tissue motion tracking and structure estimation for robotic assisted MIS procedures. *Medical image computing and computer-assisted intervention : MICCAI*, 8(Pt 2) :139–46, January 2005.
- [SMW06] Scott Schaefer, T McPhail, and Joe Warren. Image deformation using moving least squares. *ACM Transactions on Graphics (TOG)*, 2006.
- [SVA⁺09] Li-Ming Su, Balazs P. Vagvolgyi, Rahul Agarwal, Carol E. Reiley, Russell H. Taylor, and Gregory D. Hager. Augmented reality during robot-assisted laparoscopic partial nephrectomy : toward real-time 3D-CT to stereoscopic video registration. *Journal of Urology*, 73(4) :896–900, April 2009.
- [SVSPY10] Danail Stoyanov, Marco Visentini-Scarzanella, Philip Pratt, and Guang-Zhong Yang. Real-time stereo reconstruction in robotically assisted minimally invasive surgery. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, volume 13, pages 275–282, January 2010.
- [TM07] Tinne Tuytelaars and Krystian Mikolajczyk. Local Invariant Feature Detectors : A Survey. *Foundations and Trends® in Computer Graphics and Vision*, 3(3) :177–280, 2007.
- [Tom91] Carlo Tomasi. Detection and Tracking of Point Features Technical Report CMU-CS-91-132. *Image (Rochester, N.Y.)*, (April), 1991.
- [Tom94] Carlo Tomasi. Good features to track. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition CVPR-94*, pages 593–600, 1994.
- [Tot11] J. Totz et. al. Enhanced visualisation for minimally invasive surgery. *International journal of computer assisted radiology and surgery*, June 2011.
- [UPdPV11] Linas Urbanavičius, Piet Pattyn, Dirk Van de Putte, and Donatas Venskutonis. How to assess intestinal viability during surgery : A review of techniques. *World journal of gastrointestinal surgery*, 3(5) :59–69, May 2011.
- [Vem] AS. Vemuri et. al. Endoscopic Video Mosaicing : Application to Surgery and Diagnostics. *Living Imaging Workshop, December 2011*.
- [VJ04] Paul Viola and Michael J Jones. Robust real-time face detection. *International journal of computer vision*, 57(2) :137–154, 2004.
- [Wan10] X. Wang et. al. Endoscopic video texture mapping on pre-built 3-D anatomical objects without camera tracking. *IEEE transactions on medical imaging*, 29(6) :1213–23, June 2010.
- [Wu04] Chia-Hsiang Wu. Automatic extraction and visualization of human inner structures from endoscopic image sequences. *Proceedings of SPIE*, 5369 :464–473, 2004.
- [Yip12] M. Yip et. al. Tissue Tracking and Registration for Image-Guided Surgery. *IEEE transactions on medical imaging*, 31(11) :2169–2182, August 2012.