PolyPublie
Polytechnique Montréal

POLYTECHNIQUE MONTRÉAL
UNIVERSITÉ D'INGÉNIERIE

Titre: Title:	Bayesian differentiation of multi-scale line-structures for model-free instrument segmentation in thoracoscopic images
Auteur: Author:	Luke Windisch
Date:	2005
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Windisch, L. (2005). Bayesian differentiation of multi-scale line-structures for model-free instrument segmentation in thoracoscopic images [Master's thesis, École Polytechnique de Montréal]. PolyPublie. https://publications.polymtl.ca/7693/

Document en libre accès dans PolyPublie
Open Access document in PolyPublie

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/7693/
Directeurs de recherche: Advisors:	Farida Cheriet, & Guy Grimard
Programme: Program:	Unspecified

Ce fichier a été téléchargé à partir de PolyPublie, le dépôt institutionnel de Polytechnique Montréal
This file has been downloaded from PolyPublie, the institutional repository of Polytechnique Montréal

https://publications.polymtl.ca

UNIVERSITÉ DE MONTRÉAL

BAYESIAN DIFFERENTIATION OF MULTI-SCALE LINE-STRUCTURES FOR MODEL-FREE INSTRUMENT SEGMENTATION IN THORACOSCOPIC IMAGES

LUKE WINDISCH
INSTITUT DE GÉNIE BIOMÉDICAL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION DU
DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES
(GÉNIE BIOMÉDICAL)
AOÛT 2005

© Luke Windisch, 2005.

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-16866-0
Our file Notre référence
ISBN: 978-0-494-16866-0

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé:

BAYESIAN DIFFERENTIATION OF MULTI-SCALE LINE-STRUCTURES FOR MODEL-FREE INSTRUMENT SEGMENTATION IN THORACOSCOPIC IMAGES

présenté par: WINDISCH Luke
en vue de l'obtention du diplôme de: Maîtrise ès sciences appliquées
a été dûment accepté par le jury d'examen constitué de:

M. MATHIEU Pierre, D.Sc.A., président
Mme. CHERIET Farida, Ph.D., membre et directrice de recherche
M. GRIMARD Guy, M.D., membre et codirecteur de recherche
M. BILODEAU Guillaume-Alexandre, Ph.D., membre

To my wife Andrea,
whose support was invaluable and indelible.

AKNOWLEDGEMENTS

While the accolades for a completed Master's thesis fall principally to its author, he or she could not possibly endure the rigors of research without a deep network of support. This network has been essential to the success of this work, and I would like to thank all those who contributed in their own way.

In the first place, thank you to my supervisor Farida. Since the November day I first showed up in her lab and expressed an interest in a Master's, her enthusiasm, patience and technical guidance have been wholly appreciated. My co-director, Dr. Guy Grimard, at the Sainte-Justine hospital , also deserves thanks. His unfailing willingness to let me observer surgeries and discuss the project from a user's perspective have added a valuable dimension to this work while broadening my appreciation of the demands of medical-based research.

To the countless others who have contributed to my growth as a researcher – Dr. Ian Stokes at the University of Vermont, the other students and researches associated with the LIV4D lab – thank-you for always being open to discussions, whether technical or not, and providing so many worthy lessons that cannot be found in textbooks.

Underlying all these contributions is the necessity of funding. For financial support, I am deeply appreciative of the MENTOR program for research into spinal deformities. Not only did the funding sustain me as a student, but it helped ensure I could pursue important opportunities that have allowed me to launch a career in the domain of orthopedics. Thank-you also to the Natural Sciences and Engineering Research Council of Canada, whose funding supports the project that this thesis contributes toward.

Finally, thanks to my family: my parents and sister for their general encouragement, and to Andrea, who not only supported me through many long hours, but continues to teach me the value of true dedication to your goals. It could not have happened without you.

RÉSUMÉ

Les chirurgies minimalement invasives du rachis peuvent réduire le risque et la période de convalescence pour le patient, comparées aux chirurgies conventionnelles qui exposent entièrement les structures internes du thorax. Dans une chirurgie par approche antérieure comme la discoidectomie le chirurgien insère les instruments à travers un certain nombre de petites incisions effectuées sur le torse du patient et visualise le champ opératoire par l'intermédiaire de la vidéo d'un endoscope, doté d'une caméra monoculaire et une source lumineuse intégrée. La visualisation du champ opératoire à l'aide d'une séquence monoculaire d'images pose de nombreuses difficultés, y compris une perte de perception de profondeur avec peu de possibilité de compensation vu que la proximité de l'endoscope aux objets d'intérêt crée un champ de vision étroit. Finalement, la longue période d'apprentissage du chirurgien pour développer une certaine habilité pour ce genre de technique a limité la diffusion de cette approche minimalement invasive ainsi que les avantages associés.

Un système de réalité augmentée permettant d'intégrer les modèles 3D préopératoire du patient avec la séquence vidéo fournie par l'endoscope aiderait à résoudre les problèmes de la perte de profondeur et de contexte. Un élément central au succès de ce système est le re-calibrage automatique des paramètres intrinsèques et extrinsèques de l'endoscope, qui peuvent changer durant la chirurgie suite au mouvement de la caméra et aux opérations de `focus' manuelles effectuées par le chirurgien. Les paramètres extrinsèques seront mis à jour par un système optique de suivi de mouvement. Par contre, la mise à jour des paramètres intrinsèques exige le suivi des instruments à travers la séquence vidéo de l'endoscope. Ces derniers doivent être alors détectés automatiquement, de sorte que leurs mouvements apparents sur les images puissent être corrélés avec les mouvements retournés par le système de suivi optique.

L'identification automatique des instruments chirurgicaux sur les images est rendue difficile par la nature complexe des images de l'endoscope. Finalement, un schéma viable de segmentation de régions correspondant aux instruments est exigé afin d'isoler les régions d'image qui doivent être suivies tout au long de la séquence. Les algorithmes

classiques de segmentation sont limités par les défis des images d'endoscope, et une méthode alternative basée sur la stabilité des objets de forme rectiligne et une formulation rigoureuse caractérisant les méthodes probabilistes Bayésiennes a été développée.

Une structure rectiligne a une longueur plus grande comparée à sa largeur (un rapport longueur à largeur élevé), et les instruments utilisés en chirurgie minimalement invasive possèdent cette caractéristique: ils doivent être assez étroits pour s'insérer à travers de petites incisions, mais assez longs pour traverser le thorax et atteindre la partie antérieure du rachis. Les opérateurs de détection de lignes ont été développés pour exploiter la structure 'longue et mince' de certaines structures anatomiques dans d'autres domaines d'imagerie médicale -- détection du réseau vasculaire sur des images IRM , par exemple -- mais aucun n'a été appliqué aux images endoscopiques, où le contraste faible des objets et la superposition d'un arrière plan complexe paralysent les détecteurs standards de structures rectilignes.

Ce mémoire porte sur cette limitation, en remaniant la théorie de structures rectilignes dans un schéma de régularisation Bayésienne qui peut distinguer efficacement les réponses générées par des structures rectilignes correspondant à des instruments chirurgicaux, et celles générées par des structures rectilignes correspondant à des objets de l'arrière plan comme les vaisseaux sanguins. De plus, il applique les principes de détection de contours pour améliorer la précision de la détection des structures rectilignes dans les images endoscopiques. Enfin, l'algorithme de segmentation proposé rebranche naturellement avec les méthodes de détection de structures rectilignes en intégrant un schéma de détection basé sur les valeurs propres comme une contrainte *a priori* dans le modèle Bayésien.

L'approche de segmentation Bayésienne de structures rectilignes offre également un point de départ intuitif pour le suivi des instruments chirurgicaux à travers la séquence d'images endoscopiques. Un récent développement dans la théorie du flux optique, qui est basé sur un formalisme Bayésien, est modifié en utilisant la réponse en valeur propre de structures rectilignes générée dans l'étape de segmentation. Comparée à la méthode de suivi Bayésienne originale, qui dépend fortement de l'information d'intensité, la formulation basée sur la réponse rectiligne offre un indicateur plus stable de forme et de

localisation de l'instrument.

Les résultats actuels de segmentation et de suivi des instruments chirurgicaux sont extrêmement encourageants. L'utilisation de la théorie de structures rectilignes code la forme à un niveau perceptuel élevé, il en résulte que les méthodes sont robustes même en présence de faibles intensités et de flou inhérent aux déplacements de la caméra. La flexibilité mathématique et algorithmique des concepts Bayésiens permet une grande latitude dans la formulation de fonctions appropriées pour la description des spécificités du contenu des images endoscopiques, y compris les changements apparents de la taille des instruments dus au mouvement de la caméra. Finalement, ces éléments combinés ont permis de produire un algorithme de segmentation qui détecte de façon fiable et automatique les régions correspondant aux instruments chirurgicaux.

Les résultats de suivi sont également prometteurs. En raison de la représentation plus stable de la structure obtenue grâce à l'utilisation de la réponse en valeur propre de structures rectilignes, les résultats de la méthode Bayésienne de flux optique sont améliorés ; cependant, les résultats ne sont pas aussi avancés que ceux de la segmentation et beaucoup de travail reste à effectuer. Toujours, pour le moins, les résultats actuels et la théorie sous-jacente devraient fournir un bon point de départ pour les travaux de recherche futurs.

En termes de travaux de recherche futurs, la faiblesse la plus notable de la méthode courante de segmentation est la sous-segmentation. Ce problème est dans la grande partie une conséquence de choix délibérés de conception, mais plusieurs avenues sont disponibles pour améliorer les résultats. Deux des solutions les plus notables incluent l'amélioration des résultats en utilisant la position et la forme des instruments retournées par l'étape de segmentation comme contraintes rigides, et la correction des réflexions spéculaires, qui compliquent la sélection Bayésienne de structures rectilignes.

Les travaux futurs sur le problème de suivi des instruments peuvent se concentrer sur des détails tels qu'une meilleure intégration de la réponse en valeur propre de structures rectilignes. Cependant, la résolution du problème de suivi devrait commencer par identifier la nature des primitives géométriques qui peuvent être extraites de l'instrument tout au long de la séquence d'images afin de résoudre le problème de calibrage.

ABSTRACT

Minimally invasive techniques for spinal surgery can reduce patient risk and in-hospital recovery periods compared to surgeries that fully expose the surgical site. In an anterior approach minimally invasive discectomy, the surgeon inserts instruments through a number of small incisions on the patient's torso and views the surgical site via video from an endoscope, a monocular camera with integrated light source. The monocular view presents numerous difficulties, including a loss of depth perception with few compensating depth cues since the proximity of the endoscope to the objects being imaged creates a narrow field-of-view. Ultimately, the long surgeon training to develop comfort with the view has limited the spread of this minimally invasive approach and its associated benefits.

An enhanced reality surgical system to integrate 3D preoperative patient models with the endoscope video stream would help solve the problems of depth and context loss. Central to the success of this system is automatic recalibration of the extrinsic and intrinsic endoscope parameters, which may vary throughout the surgery due to camera motion and manual focusing. Extrinsic parameters will be updated by tracking instruments using a global optical positioning system. Updating the intrinsic parameters requires tracking instrument objects in the endoscope video that can be easily and automatically detected, so that their image-based movements can be correlated with movements returned by the optical tracking system.

Identifying instrument objects in the images is made difficult by the complex characteristics of endoscope imagery. Ultimately, a viable scheme to segment instrument regions is required in order to isolate which image objects are appropriate for tracking. Typical segmentation algorithms are limited by the challenges of endoscope images, and an alternative method based on the stability of line-objects and rooted in the rigour of Bayesian probabilistic methods has been developed.

A line-structure has an extended length compared to its width (a high 'length-to-width ratio'), and the instruments used in minimally invasive surgery exhibit this general trait: they must be narrow enough to fit through small incisions, yet long enough to traverse the

thorax and reach the anterior spine. Line-detector operators have been developed to exploit the 'long-but-thin' structure of certain anatomical structures in other medical imaging domains—detection of vascular trees in MRI imaging, for example—but none have been extended to sub-optimal images, where poor object contrast and complicated backgrounds paralyze standard line-object detectors.

This thesis addresses this limitation, recasting the theory of line-objects into a Bayesian regularization scheme that is able to effectively distinguish between line-object responses generated by target instruments, and those generated by both real and apparent line-like background objects. Moreover, it applies edge-detection principles to improve the precision of line-object detection for the endoscope task. Finally, the proposed segmentation algorithm reconnects naturally with the motivating line-object methods by integrating a previous eigenvalue-based detection scheme into an *a priori* constraint in the Bayesian model.

The line-object, Bayesian-based approach to segmentation also offers an intuitive starting point for an investigation into instrument tracking in endoscope images. A recent advancement in optical-flow theory that uses a Bayesian framework is modified with the eigenvalue line-object response produced in the segmentation step. Compared to the original Bayesian-based tracking method, which depended heavily on intensity information, the line-response based formulation offers a more stable indicator of instrument shape and location.

The current results of the segmentation and tracking methods are extremely encouraging. The use of line-object theory encodes shape at a high perceptual level, making the methods robust to low intensity as well as the strong blurring often required for noise removal. The mathematical and algorithmic flexibility of Bayesian concepts afford large latitude in the formulation of functions that are well suited to the specifics of endoscopy, including changes in apparent instrument size due to unpredictable camera motion. Ultimately, these elements have combined to produce a segmentation algorithm that reliably and automatically returns instrument regions.

The tracking results are also promising. Because of the more stable representation of structure gained through the use of the eigenvalue line-object response, results compared

to the original intensity formulation of the Bayesian optical flow method are improved; however, the results are not as advanced as those of the segmentation algorithms and much work remains to be done. Still, at the very least, the current results and underlying theory should provide a good starting point for future research.

In terms of future research, the most notable shortcoming of the current segmentation method is under-segmentation. This problem is in large part a consequence of deliberate design choices, but several avenues are available to improve the results. Two of the most notable areas for development include refinement of the results using instrument positions and shapes returned by the current segmentation as powerful constraints, and removal of specular reflections, which complicate the Bayesian selection of line-objects.

Future work on the tracking problem may focus on details such as better integration of the eigenvalue response. However, the tracking investigation should also investigate in detail the image objects that will be tracked to achieve the calibration objective. This investigation is still pending, and its results will help reveal the most important issues that need to be addressed by subsequent tracking research.

CONDENSÉ

Introduction

Les risques associés à l'exposition de l'intérieur du corps d'un patient pendant la chirurgie du rachis peuvent être atténués par les techniques de chirurgie minimalement invasives, durant lesquelles un chirurgien insère des instruments à travers un certain nombre de petites incisions et visualise le champ opératoire par l'intermédiaire de la vidéo d'un endoscope, une caméra monoculaire avec une source lumineuse intégrée. Comparées aux chirurgies conventionnelles qui exposent entièrement les structures internes du thorax, le risque pour le patient, la période de convalescence et les coûts d'hospitalisation sont réduits.

Malheureusement, l'adoption de ces procédures a été ralenti par les difficultés considérables auxquelles font face les chirurgiens, qui perdent la perception de profondeur puisque le champ opératoire est visualisé indirectement par l'intermédiaire de l'endoscope monoculaire. De plus, les compensations contextuelles de profondeur sont faibles vu que la proximité de l'endoscope aux objets d'intérêt crée a champ de vision étroit (FOV), de l'ordre de 5-centimètres ou moins. Finalement, la longue période d'apprentissage requise pour que le chirurgien acquiert le confort et l'expertise nécessaires pour fonctionner dans un environnement aussi restreint a limité la diffusion de cette approche.

Un système de réalité augmentée intégrant les modèles 3D préopératoires de patients avec la séquence vidéo fournie par l'endoscope aiderait à résoudre les problèmes de la perte de profondeur et de contexte. Un élément central au succès de ce système est le re-calibrage automatique des paramètres extrinsèques et intrinsèques de l'endoscope, qui peuvent changer durant la chirurgie suite aux mouvements de la caméra et aux opérations de focalisation manuelles effectuées par le chirurgien. Les paramètres extrinsèques de l'endoscope sont mis à jour en temps réel en utilisant un système optique de suivi de mouvement. Cependant, la mise à jour des paramètres intrinsèques exige le suivi et la détection automatique des instruments chirurgicaux à travers la séquence d'images. Les mouvements relatifs réels entre la caméra et les instruments chirurgicaux peuvent être

ainsi reliés aux mouvements apparents des instruments identifiés sur la séquence d'images afin de mettre à jour les paramètres intrinsèques de la caméra. Etant donné le besoin du système en précision, les objets appropriés pour le calibrage sont ceux pour lesquels le mouvement réel 3D peut être inféré à partir du mouvement apparent sur les images avec une bonne précision : des points singuliers tels que des coins sont probablement les plus utiles, mais d'autres primitives telles que les contours d'instruments ou leur forme peuvent être également appropriées. À cette étape de développement du système, la recherche sur le type de primitives les plus appropriées est encore en suspens.

Sans se soucier du type de primitives finalement utilisé, leur identification est compliquée pour plusieurs raisons, entre autres les caractéristiques du contenu des images endoscopiques elles-mêmes. Les difficultés dues à la nature du contenu sont énumérées ci-dessous:

- La présence de flou provenant du sang et de la vapeur ;
- Occlusions dues aux tissus coupés ;
- Intensité et contraste fortement variables ;
- Réflexions spéculaires fortes dues aux tissus moites et aux instruments métalliques ; et
- Changement brusque du rapport d'aspect dans les images provoqué par les mouvements relatifs entre les instruments et l'endoscope.

Pour des primitives de bas niveau comme des points, d'autres complications proviennent du fait qu'il est difficile de distinguer l'ensemble éparse de primitives identifiées sur des surfaces lisses d'instruments de celles appartenant à l'arrière plan-- un critère rigoureux de détection qui peut supprimer le fond risque également de tailler trop fortement les instruments. Par conséquent, pour identifier les primitives d'intérêt un schéma viable de segmentation des instruments à partir des images endoscopiques est exigé. Ce schéma doit être robuste aux caractéristiques d'images décrites ci-dessus, fournissant au moins un des deux résultats : (i) un ensemble approprié de primitives identifiables à travers une séquence d'images, tels que des formes caractérisant des instruments, ou (ii) des régions d'image à partir desquelles des points singuliers comme des coins peuvent être extraits et suivis le long de la séquence.

Une fois que des primitives appropriées sont obtenues, celles-ci doivent être suivies de façon robuste à travers toute la séquence d'images de sorte que leur mouvement apparent soit utilisé pour le calibrage. La résolution de ce problème est actuellement limitée par le fait que, comme mentionné précédemment, les primitives qui pourraient être suivies de façon robuste à travers la séquence d'images n'ont pas encore été définies. Néanmoins, une recherche préliminaire sur les algorithmes de suivi a été explorée. Comme les algorithmes de segmentation, certaines exigences fondamentales du système doivent être respectées pendant le développement si les algorithmes doivent être utilisés dans les étapes postérieures de la conception du système de calibrage.

En général, les solutions des problèmes de segmentation et de suivi doivent contribuer au développement d'un système bénéfique à l'utilisateur final : l'équipe chirurgicale. Pour être utile, le système doit être correct, rapide, et peut être intégré naturellement dans le protocole standard des opérations chirurgicales. Plus spécifiquement, la profondeur et l'information de contexte doivent être précises, le système ne peut pas exiger à l'équipe chirurgicale d'attendre après les résultats de certaines procédures (ou augmenter le risque pour le patient et les frais d'opération), et il doit limiter les interactions avec l'équipe chirurgicale. Pour aider à assurer le respect de ces contraintes, il est important d'étudier la situation actuelle des technologies dans les domaines de segmentation et de suivi.

État de l'art

La plupart des algorithmes de segmentation peuvent être répartis dans deux catégories : les algorithmes basés sur les techniques de classification (clustering) et les algorithmes basés sur la détection de contours. Dans la première catégorie, des régions de pixels sont itérativement formées et séparées jusqu'à ce que la division de l'image optimise une certaine mesure d'homogénéité aux niveaux inter régions et intra région. La segmentation s'arrête une fois que tous les pixels ont été assignés à leur groupe optimal ; cependant, une interprétation des régions résultantes est rarement fournie. En revanche, les algorithmes basés sur la détection de contours ont habituellement pour but de segmenter l'image en des régions "cibles" et des régions de "fond", et une classification plus raffinée de ce résultat dichotomique est rarement désirée. Dans ces approches un contour représentant la

frontière de la région cible est itérativement déplacé (contour actif) jusqu'à ce qu'une certaine métrique exprimant l'accord entre la position actuelle du contour et la position prévue de la frontière, étant donné des mesures réelles extraites de l'image, soit optimisée.

Pour le problème de calibrage, l'objectif est incontestablement d'isoler des instruments cibles du fond composé de tissus. Une grande panoplie de métriques pour évaluer la similarité entre les contours actifs et ceux prévus existent et sont basées soit sur l'intensité du gradient (information contenue sur les arêtes) soit sur l'intensité des régions (information de texture ou des descripteurs de couleur) soit sur d'autres informations contenues dans l'image. Ces métriques seules sont rarement suffisantes pour contraindre adéquatement le contour; la notion du problème bien-posé est fréquemment violée, et des solutions au problème de segmentation -- comme avec la plupart des problèmes de vision assistée par ordinateur -- doivent être obtenues par l'intermédiaire du concept de la régularisation.

Régularisation

L'idée fondamentale de la régularisation est d'introduire une connaissance a priori du problème pour contraindre la solution quand la métrique de similarité entre la solution courante et la solution prévue est insuffisante. Deux des cadres théoriques les plus communs sont la régularisation de Tikhonov et la régularisation Bayésienne. L'esprit des deux schémas est identique : soumettre de nombreuses solutions candidates dans une fonctionnelle combinant la connaissance a priori et la métrique de similarité, et prendre la solution qui optimise le résultat. Dans la régularisation de Tikhonov, la solution est le minimum de la fonctionnelle, alors que dans l'approche Bayésienne, la fonctionnelle prend la forme d'une probabilité conditionnelle et la solution maximise cette probabilité. En général, l'approche Bayésienne est beaucoup plus flexible, et en fait, l'approche de Tikhonov peut s'avérer comme un cas particulier de l'approche Bayésienne.

Les contours actifs pour la segmentation

Généralement le schéma de segmentation, basé sur la détection de contours, le plus utilisé découle du cadre de régularisation de Tikhonov. L'approche des contours actifs de base utilise l'information de courbure pour compléter la mesure ambiguë du gradient

d'intensité de l'image, en pénalisant le développement de courbures élevées dans le contour actif. Une telle courbure pourrait, par exemple, se développer autour du bruit local, qui crée des gradients forts mais ne correspond pas aux frontières d'un objet cible. Malheureusement cette approche est légèrement limitée parce que le contour actif doit être paramétrisé par un ensemble de particules, qui doivent être supprimées, rajoutées, et redistribuées pour maintenir la stabilité numérique pendant l'évolution du contour. Cette stratégie da calcul est rarement intuitive ni évidente à développer.

Les contours actifs géodésiques (GAC) sont une méthode alternative populaire d'évolution des contours également basée sur des gradients d'intensité d'images et de courbure, mais mis en application par l'intermédiaire des lignes de niveau au lieu des contours paramétrisés. Les schémas de lignes de niveau offrent l'avantage considérable d'une paramétrisation implicite de la courbure, et sont donc beaucoup plus stables. De plus, ils manipulent naturellement les changements de topologies changeantes du contour de segmentation, qui est particulièrement avantageux pour des problèmes où il peut y avoir plusieurs objets cibles – un contour initialisé pour entourer toutes les cibles se subdivisera facilement en nombre approprié des frontières séparées.

En général, la méthode de GAC fonctionne bien, mais elle peut tirer profit de contraintes additionnelles. À cet effet, la régularisation par l'intermédiaire de connaissances a priori sur la forme d'objets cibles a été prouvée efficace. Des contraintes appropriées de forme sont souvent basées sur des modèles obtenus à partir d'un ensemble d'entraînement de grande taille par l'intermédiaire de l'analyse par composantes principales. Cependant, pour des images d'endoscope cette philosophie est trop restrictive : des mouvements relatifs entre les instruments et l'endoscope font qu'il est impossible de considérer tous les modèles PCA pour représenter toutes les vues possibles de tous les instruments possibles. Par conséquent, un modèle plus général de forme adaptable au changement d'aspect des différents instruments serait utile.

Un autre problème avec les contraintes basées sur la forme, ainsi qu'avec les schémas de segmentation basés sur les contours en général, est qu'ils exigent souvent l'initialisation de la frontière de segmentation dans la région cible. Pour réduire au minimum le besoin d'interaction de l'équipe des chirurgiens avec le système de calibrage,

une méthode de segmentation qui peut identifier automatiquement les objets d'intérêt est extrêmement souhaitable.

Segmentation d'objets rectilignes

Un schéma efficace de segmentation basé sur la forme qui est à la fois assez général pour des images endoscopiques et qui peut identifier automatiquement des régions d'image cibles a été développé en exploitant la caractéristique long-et-mince des instruments cible. Cette propriété provient du fait que les instruments utilisés en chirurgie minimalement invasive doivent être assez minces pour pouvoir être insérés à travers de petites incisions effectuées dans le thorax d'un patient et assez longs pour atteindre les organes du site chirurgical.

Pour exploiter cette caractéristique de forme, les concepts d'objets rectilignes sont utiles. Comme leur nom l'indique, les objets rectilignes ont une longueur beaucoup plus importante que leur largeur, et l'intensité de l'image d'un objet rectiligne devrait montrer une variation beaucoup plus forte à travers son profil transversal que son profil longitudinal. De nombreux schémas de segmentation basés sur cette propriété ont été développés pour l'extraction du réseau vasculaire à partir d'images et de volumes obtenus par résonance magnétique (MRI). Bien que les détails changent, le concept sous-jacent à ces schémas est de rechercher une image de pixels où la variation d'intensité dans la direction du gradient d'intensité est considérablement plus faible que dans la direction transversale du gradient. L'extraction de l'ensemble de ces pixels produira non seulement la segmentation finale de l'image, mais indiquera immédiatement les régions de l'image qui représentent les objets cibles.

Suivi des instruments à travers la séquence d'images

La théorie d'objets rectilignes motive également une étude préliminaire sur le suivi des instruments à travers la séquence d'images qui, bien que limité par le fait le type d'objets à suivre doit encore être défini, ça vaut la peine de la considérer puisqu'elle découle naturellement de la théorie précédente de segmentation et servira assurément comme une bonne base pour les travaux futurs. Comme la segmentation, la théorie du suivi d'objets à travers une séquence est vaste, et il y a trois classes d'approches : approches basées sur

des transformations, approches basées sur la mise en correspondance de primitives, et les approches basées sur les contours. Les approches basées sur la mise en correspondance telles que le filtre de Kalman sont probablement les plus applicables dans le contexte de calibrage, mais celles-ci exigent une mise en correspondance temporelle d'un ensemble de primitives qui doivent être suivies à travers la séquence. Cependant, les schémas basés sur le suivi de contours offrent un potentiel intéressant, puisque les méthodes sont souvent des extensions immédiates des algorithmes de segmentation de contours : au lieu d'une initialisation arbitraire, un contour est initialisé sur l'objet cible dans l'image précédente et évolue vers la position cible dans l'image courante.

Dans la classe d'approches de suivi de contours, le flot optique est une approche bien connue formulée dans cadre de régularisation de Tikhonov. Dans la méthode de flot optique, une métrique basée sur l'intensité de l'image est utilisée pour suivre une région cible, et une contrainte de lissage sur les champs de vitesse pour les pixels appartenant à la région cible est considérée pour contraindre la solution de suivi.

Les travaux récents ont remanié le problème de flot optique basé sur la régularisation de Tikhonov dans une méthode de suivi basée sur une régularisation Bayésienne et ont éliminé la nécessité de calculer le champ de vitesse dans la région cible. De plus, le suivi a été mis en application à l'aide d'une approche basée sur les lignes de niveau, ce qui offre les avantages associés à l'indépendance topologique et à la stabilité numérique. Cette propriété peut être d'un intérêt particulier pour les images endoscopiques, puisque l'apparence des instruments sur les images peut nettement changer, et plusieurs régions d'instruments peuvent être présentes en même temps. Cependant, schéma de suivi probabiliste est encore dépendant de l'information d'intensité, qui est de fiabilité douteuse dans des images endoscopiques quand elle est utilisée en l'absence d'autres mesures.

Méthodologie

Le travail principal dans ce projet est à d'adapter la théorie d'objets rectilignes aux défis uniques de la segmentation et du suivi dans les séquences d'images endoscopiques. Pour le problème de segmentation, la solution commence par détection de contour à l'aide du filtre de Canny pour localiser les contours qui peuvent correspondre aux images

d'instruments. Les contours résultants sont des indicateurs plus précis de frontières d'instrument que les résultats retournés par d'autres détecteurs d'objets rectilignes; cependant, comme d'autres détecteurs, la détection de contours n'est pas robuste aux arrières plans texturés. Pour commencer à traiter ce problème, la propriété générale de structure curviligne des frontières d'instruments est exploitée en appliquant un seuil de courbure aux contours détectés. Les contours dont la courbure excède ce seuil sont rejetés, et l'ensemble des contours restant est trié en enlevant ceux qui sont si courts qu'ils peuvent être classés avec confiance comme des structures appartenant à l'arrière plan. Les Chaînes de points équidistants entre les paires de contours restants sont alors extraits. Ces chaînes, ou squelettes, extraits le long du centre des objets curvilignes dans l'image contiennent chacun les informations de forme et de rapport d'aspect des objets curvilignes sous-jacents. Les squelettes représenteraient la segmentation finale des instruments, sauf que la procédure de tri basée sur le seuil de courbure ne peut pas garantir d'enlever tous les squelettes surgissant des objets curvilignes et des textures appartenant à l'arrière plan. D'où une différence fondamentale entre le schéma de segmentation d'objets curvilignes proposé et les méthodes précédentes : les autres méthodes ne considèrent pas la possibilité de détection d'objets curvilignes non désirables, et prendraient ces squelettes comme segmentation finale. Par conséquent, ces méthodes ne peuvent pas être appliquées telles qu'elles aux images avec un arrière plan complexe comme c'est le cas des images endoscopiques.

Schéma de régularisation Bayesienne pour la segmentation

Pour extraire uniquement les squelettes correspondant aux instruments, un schéma de régularisation Bayésien est développé, ce qui est concentré sur l'exploitations de certaines caractéristiques générales des objets rectilignes comme les instruments. En premier lieu, comme indiqué plus tôt, les objets rectilignes devraient être beaucoup plus longs que larges et un terme de probabilité, pour accorder la priorité aux squelettes qui présentent cette propriété, est formulé. La définition de ce terme de probabilité est compliquée vu les mouvements relatifs des instruments par rapport à l'endoscope, qui peuvent causer l'apparition d'un long-et-mince instrument avec ces proportions inversées dans l'image

s'il est près de l'endoscope. Le terme de probabilité est alors adapté à cette possibilité en tenant compte de plus court squelettes à de plus hautes niveaux d'échelle. En général, les instruments sont les objets de plus grande échelle dans une image, et ce biais dans ce terme de probabilité est adapté à cette réalité. Néanmoins, la possibilité de détecter des objets à grande échelle dans l'arrière plan doit être considérée. De plus, ce n'est pas tous les instruments qui vont être identifiés comme des objets à grande échelle, alors d'autres termes sont nécessaires pour différencier entre les instruments et les objets d'arrière plans à de basses échelles.

Termes Bayésiens a priori pour la segmentation

À cette fin, deux termes *a priori* ont été développés pour coder la connaissance (i) de la structure de la surface et des échelles relatives des instruments comparées aux objets rectilignes de l'arrière plan, aussi bien que (ii) l'importance d'une mesure basée sur la valeur propre de bons objets rectilignes. Le premier terme reflète le fait que la plupart des instruments sont réfléchissants et/ou facettés, et devrait donc produire une série d'objets rectilignes à différentes échelles à travers leur largeur. Puisque les instruments existent généralement à de grandes échelles, les objets rectilignes à de plus hautes échelles dans une image qui présentent aussi des réponses à de plus petites échelles ont une probabilité plus élevée d'être des instruments.

Le terme a priori de valeur propre provient du détecteur précédent d'objets rectilignes qui utilise la matrice Héssienne pour évaluer le "la rectilinéarité de l'objet" dans une région d'image. En bref, la première la valeur propre de la matrice Héssienne d'intensité à un pixel de l'image devrait être beaucoup plus grande que la deuxième valeur propre si le pixel fait partie d'un bon objet rectiligne. Une conséquence de l'utilisation de la théorie d'objets rectilignes dans un premier temps, les instruments devraient avoir une bonne réponse de valeur propre. Par conséquent, les objets rectilignes qui répondent non seulement bien au terme de probabilité et au premier terme a priori, mais ont également une bonne réponse de valeur propre sont bien plus candidats pour être des instruments. L'utilisation de cette réponse en valeur propre apparente bien le schéma de segmentation proposé aux méthodes originales d'objets rectilignes qui ont initialement motivé le travail.

Le suivi basé sur la réponse en valeur propre

L'analyse de la valeur propre appliquée dans la tâche de segmentation fournit également une alternative naturelle à la formulation d'une approche Bayésienne de flot optique. Même si elle utilise toujours l'information d'intensité, la réponse en valeur propre des objets rectilignes est signée selon le contraste des objets : une réponse négative indique un objet lumineux, et une réponse positive indique un objet foncé. Tandis que les valeurs spécifiques d'intensité peuvent fluctuer, les caractéristiques générales de contraste d'un instrument sont supposées être plus stables. En conséquence, la réponse signée de la valeur propre est substituée directement à l'information d'intensité dans le schéma Bayésien de flot optique, fournissant un indicateur de forme et de région d'instrument qui devrait être plus robuste aux fluctuations dans l'intensité locale.

Résultats et discussion

Les résultats du schéma Bayésien de segmentation des objets rectilignes dépendent fortement des valeurs des divers paramètres de la procédure d'extraction squelettique basée sur une détection de contours et la procédure Bayésienne de sélection de squelettes. L'analyse rigoureuse de ces paramètres a indiqué des valeurs stables pour plusieurs, bien qu'un certain nombre ait besoin d'être raffiné -- des paramètres pour l'extraction finale de squelettes à partir du seuil de courbure appliqué sur les contours, par exemple. Aussi, un ensemble fixe de paramètres a été déterminé, une fois appliqué à une grande série d'images, a produit des résultats généralement tout à fait bons de segmentation. Spécifiquement, les régions d'instrument ont été uniquement, sûrement, et automatiquement extraites.

En dépit de ces résultats prometteurs, beaucoup d'amélioration est encore possible. Pour commencer, les limites du schéma d'extraction basé sur la détection de contours ont démontré le potentiel une sur-segmentation de l'instrument dans un cas isolé. Ce problème peut probablement être résolu par une meilleure implémentation numérique ou une formulation améliorée du terme *a priori* Bayésien. Un autre problème plus significatif que cas isolé de sur-segmentation est le problème persistant de sous-segmentation des instruments. Tandis que cette situation provient principalement d'un

choix délibéré de conception -- pour un suivi non ambigu et le calibrage, il est préférable d'avoir un plus petit ensemble d'objets représentant de façon fiable des instruments qu'un grand ensemble d'objets dont certains sont des structures de l'arrière plan-- les résultats peuvent être améliorés par le raffinement de la segmentation initiale d'objets rectilignes en utilisant des schémas d'évolution de contours sous des contraintes de forme. La correction des réflexions spéculaires de l'image peut également améliorer les résultats. Ces réflexions sont la cause de la détection de plusieurs objets rectilignes de l'arrière plan, et leur correction devrait tenir compte de paramètres de segmentation plus flexibles afin de produire plus de véritables régions d'instruments.

Comparé aux résultats de la segmentation, ceux des algorithmes de suivi sont beaucoup plus préliminaires. Jusqu'ici, l'implémentation algorithmique du schéma Bayésien basé sur le flot optique n'a pas été modifié par rapport à la méthode originale basée sur l'information d'intensité. Le seul changement était de remplacer les mesures de gradient d'intensité avec les réponse en valeur propre des objets rectilignes comme décrit précédemment. Cependant, en dépit de la portée limitée de la recherche jusqu'ici, les résultats sont tout à fait prometteurs. D'abord, la formulation en valeur propre suit les régions d'intérêt de façon plus fiable et plus robuste dans les zones de basse intensité. De plus, des résultats améliorés ont été obtenus sans aucun effort d'optimisation des paramètres, ainsi il reste beaucoup de potentiel inexploité qui pourrait améliorer les résultats.

Conclusion

En dépit des limites mentionnées, les résultats de segmentation sont extrêmement prometteurs. Un avantage particulier est l'intégration des concepts d'objets rectilignes avec la régularisation Bayésienne. Les caractéristiques d'objets rectilignes encodent la forme de l'objet à un niveau perceptuel élevé, faisant ce type de représentation d'objets moins sensible à la basse intensité aussi bien qu'au flou fort souvent filtré pour atténuer le bruit dans l'image. Les concepts Bayésiens sont mathématiquement flexibles, permettant une grande latitude dans la formulation de fonctions qui sont bien adaptées aux détails spécifiques d'un problème -- changements de la taille apparente d'instrument, par

exemple.

Pour le problème de suivi, la recherche est extrêmement préliminaire. Les résultats sont encourageant, mais beaucoup de travail doit être fait. Pour le moins, les résultats actuels et la théorie sous-jacente devrait fournir un bon point de départ pour les travaux futurs.

En termes de travaux futurs, certaines directives importantes basées sur ce travail devraient être mentionnées. D'abord, l'amélioration des résultats de segmentation devrait se concentrer sur une meilleure implémentation numérique, évolution secondaire des contours en utilisant la segmentation initiale comme contrainte de forme, et correction des réflexions spéculaires. En second lieu, la recherche sur le suivi devrait se tourner d'abord vers le type de primitives approprié à extraire et suive à travers la séquence d'images. Après identification des primitives appropriées, des méthodes de suivi basées sur la mise en correspondance temporelle de primitives pourraient être évaluées. Leurs résultats aideront à déterminer si l'amélioration du travail courant basé sur la détection de contours est garantie.

TABLE OF CONTENTS

DEDICATION iv
AKNOWLEDGEMENTS v
RÉSUMÉ vi
ABSTRACT ix
CONDENSÉ xii
TABLE OF CONTENTS xxiv
LIST OF TABLES xxvi
LIST OF FIGURES xxvii
TABLE OF APPENDICES xxix
LIST OF ABBREVIATIONS xxx
INTRODUCTION 1
CHAPTER 1 - LITERATURE REVIEW 12
1.1 Advances in surgical technique 12
1.1.1 Surgical navigation concepts 13
1.1.2 Surgical navigation applied to minimally invasive surgery 14
1.2 Ill-posed problems & regularization 15
1.2.1 Ill-posed problems 15
1.2.2 Regularization 16
1.3 Segmentation 19
1.3.1 Contour-based segmentation paradigms 19
1.3.2 Intensity and colour segmentation constraints 27
1.3.3 Shape model constraints for contour-based segmentation 29
1.3.4 Line structures 32
1.3.5 Segmentation: Next steps 35
1.4 Object tracking 36
1.4.1 Transformation-based trackers 36
1.4.2 Correspondence-based trackers 37
1.4.3 Contour-based trackers 39
CHAPTER 2 - METHODOLOGY 47
2.1 Segmentation: Edge-based line-object detection 49
2.1.1 Limitations of current line-object detectors 49

2.1.2Modified line-object detection with Canny edge extraction....55
2.1.3Final comments on proposed detection method....61
2.2Segmentation: Probabilistic line-object selection....62
2.2.1Errors in current line-object segmentation methods....62
2.2.2Bayesian differentiation of target and non-target line-objects....64
2.2.3Final remarks on proposed line-object differentiation....71
2.2.4Validation procedure (complete segmentation scheme)....72
2.3Object Tracking....74
2.3.1Benefits of Bayesian tracking....74
2.3.2Eigenvalue-based Bayesian tracking....76
2.3.3Evaluation procedure....77
CHAPTER 3 - RESULTS & DISCUSSION....78
3.1Segmentation: Line-object extraction....78
3.1.1Canny detection....78
3.1.2Curvature-based edge removal....83
3.2Segmentation: Line-object selection....92
3.2.1Review of probabilistic terms and parameters....92
3.2.2Parameter investigation....93
3.3Segmentation: Final results and discussion....101
3.3.1Final parameters and results....101
3.3.2Discussion....103
3.4Tracking: Preliminary results and discussion....109
3.4.1Comparison with original intensity-based Bayesian scheme....109
3.4.2Discussion of eigenvalue-based tracking....112
3.4.3Next steps....114
CHAPTER 4 - CONCLUSION....116
REFERENCES....121
APPENDICES....125

LIST OF TABLES

Table 2.1: Aspects of segmentation algorithm validated by different test images..... 73
Table 3.1: Canny edge detection results for different parameter combinations.....80
Table 3.2: Best Canny detection parameters averaged across all endoscope test images..80
Table 3.3: Final results for Canny edge detection parameters..... 82
Table 3.4: Curvature thresholds yielding largest changes in number of edge pixels.....86
Table 3.5:Quantitative differences in edge content for changes in curvature threshold....87
Table 3.6: Quantitative results for varying skeleton extraction parameters.....96
Table 3.7: Quantitative results for varying probabilistic thresholds..... 101
Table 3.8: Final parameter values for the Bayesian line-object segmentation.....102
Table 3.9: Quantitative results for Bayesian-based tracking scheme.....112

LIST OF FIGURES

Figure 0.1: Elements of video-assisted minimally invasive surgery. 2
Figure 0.2: Operation of an enhanced reality system for MIS of the spine. 5
Figure 0.3: Problematic characteristics of endoscope images 9
Figure 0.4: Outline of each chapter of the thesis 11
Figure 1.1: Parameterized representation of a boundary 21
Figure 1.2: Initialization of the level set function to the signed distance function 24
Figure 1.3: Topological independence of level sets. 25
Figure 1.4: Unreliable edge/intensity gradient content of endoscope images 28
Figure 1.5: Problem of poor instrument contrast for segmentation 28
Figure 1.6: Line-object convolution response to first-derivative-of-Gaussian kernel 33
Figure 1.7: Probabilistic optical flow scheme 43
Figure 2.1: Structure of proposed segmentation and tracking methods 48
Figure 2.2: Response of the Hessian-based detector to an ideal line-object 50
Figure 2.3: Response of the Hessian-based detector to a non-ideal line-object 51
Figure 2.4: Poor line-object segmentation due to coincidental texture and blurring 53
Figure 2.5: Relationship between Gaussian blurring and response resolution 55
Figure 2.6: Gaussian line profile and Canny-detected edges. 56
Figure 2.7: Examples of the proposed skeleton extraction scheme 61
Figure 2.8: Problem of background line-objects in current segmentation methods 64
Figure 2.9: Likelihood term of the Bayesian line-object selection model 66
Figure 2.10: Endoscope images used to test the proposed segmentation scheme 72
Figure 2.11: Problems with intensity-based contour-tracking 75
Figure 3.1: Examples of Canny edge detection results on endoscope images. 83
Figure 3.2: Edge pixel gain for different curvature thresholds 86
Figure 3.3: Qualitative differences in edge content for changes in curvature threshold. ..88
Figure 3.4: Final results of initial edge detection and removal 91
Figure 3.5: Qualitative results for varying skeleton extraction parameters 95
Figure 3.6: Impact of a too-strong low-scale likelihood parameter. 99

Figure 3.7: Qualitative results for varying probabilistic thresholds..... 100
Figure 3.8: Final segmented images for the Bayesian line-object scheme..... 103
Figure 3.9: Limitations of the edge-based skeleton extraction scheme..... 107
Figure 3.10: Eigenvalue responses underlying endoscope images. 109
Figure 3.11: Preliminary tracking results..... 110
Figure 3.12: Advantages of modified tracking scheme compared to original. 113

TABLE OF APPENDICES

APPENDIX A: CONFERENCE PAPER..... 126
APPENDIX B: SEGMENTATION CODE..... 138
APPENDIX C: TRACKING CODE.....165

LIST OF ABBREVIATIONS

CT	Computed Tomography
ERS	Enhanced Reality Surgery
FOV	Field of View
GAC	Geodesic Active Contours
MAP	Maximum a Posteriori
MIS	Minimally Invasive Surgery
MRI	Magnetic Resonance Imaging
OR	Operating Room
PCA	Principal Components Analysis
3D	Three Dimensional

INTRODUCTION

Surgical intervention is often seen as a treatment of last resort for health complications if any alternatives are available. Not only does the surgical exposure itself put the patient at risk of infection, but the invasiveness of surgical procedures inevitably cause trauma to anatomical structures surrounding the operative zone. Furthermore, the recovery time can be long and painful for the patient, as well as costly for the hospital. Any technologies that can reduce the risks and trauma of surgery and improve patient recovery are of high interest in the medical field. For this reason, the concepts of minimally invasive surgery (MIS) have seen rapid and wide-spread development in the past 10-20 years.

Minimally invasive surgery

Compared to full (or "open") approaches, minimally invasive surgeries expose the interior anatomy of a patient only as much as absolutely demanded in order to perform the procedure. Not surprisingly definitions for the "as much as absolutely demanded" standard vary dramatically according to the operative objectives and the surgical equipment used. The use of video-assistance to facilitate MIS has seen increasing attention since the early 1980s.

In typical video-assisted MIS a small camera called an endoscope is inserted into the operative zone via a small incision on the surface of the patient's body. The surgeon will exclusively use the endoscope video displayed on a monitor to view the surgical site and control the position of his instruments, which are also inserted through small incisions (figure 0.1).

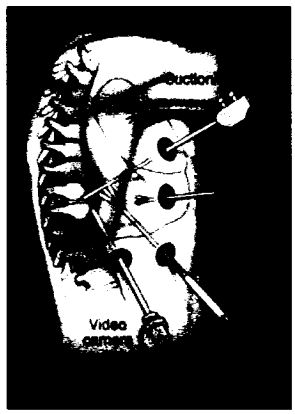


Figure 0.1: Elements of video-assisted minimally invasive surgery. Instruments for the surgery are inserted through a number of small incisions, and the operative site is viewed indirectly by video returned from the endoscope, which is also inserted through a small incision. (Source: www.spineuniverse.com/1p/ejouirnal/)

Video-assisted MIS was first popularized by Semm who used the technique for appendectomies and Mouret who applied it to cholecystectomies (gall bladder removal). Since then, numerous studies have demonstrated the considerable benefits of these two approaches ([Cox et al. 1996], [Guller et al. 2004], [Long et al. 2001]). In turn, these results have motivated application of video-assisted principles to many other surgical domains.

Despite the considerable patient and economic benefits derived from the much smaller incisions, reduced exposure, and improved recovery time, adoption of video-assisted MIS procedures has been slowed by the considerable difficulties faced by surgeons. In particular, a surgeon loses his/her depth perception since the surgical site is viewed indirectly via the monocular endoscope. Moreover, contextual depth cues are few, as the lens of the endoscope is typically very close to the objects being imaged so that the field-of-view (FOV) is extremely narrow, on order of 5-cm or less. Ultimately, the long training needed to gain comfort and expertise working with such a difficult view of the surgical site has limited the spread of this approach.

In particular, video-based minimally invasive spinal surgeries, including discectomies, corpectomies, and anterior correction of idiopathic scoliosis, are limited by these hurdles due to the proximity of numerous sensitive thoracic anatomic structures. To help make the MIS techniques easier to apply and learn, and by extension expedite the spread of

their benefits, an enhanced reality surgical system to restore depth and context information would be useful.

An enhanced-reality system for minimally invasive spinal thoracoscopy is currently under development by researchers at the LIV4D lab at the Ecole Polytechnique de Montreal in collaboration with the research center at Sainte-Justine hospital in Montreal. Critical to the success of this system are certain requirements that should be highlighted.

Requirements of an enhanced reality system (ERS)

The important design elements are linked together by the fundamental system objective discussed above, which can be formally expressed as:

> ***The ultimate goal of an enhanced reality surgery system for minimally invasive surgeries of the spine is to increase and accelerate the spread of minimally invasive techniques and their associated benefits.***

In no uncertain terms, achieving this goal will depend on the system's ability to:

- Reduce the training needed by surgeons by providing a more intuitive view of the surgical field, including depth and context information.
- Help surgeons perform minimally invasive procedures accurately and timely.

The design implications of these two points for all system components are summarized in the following points.

Accuracy

> ***Any information reported by the ERS must be precise and accurate.***

Surgeons will expect to be able to perform an enhanced reality procedure with equivalent or better precision and accuracy compared to current minimally invasive techniques. If depth (e.g. the relative depth of two anatomical landmarks) or context information—or any other data—provided by the system is not reliable, or gives the surgeon reason to question its validity, the system will be ineffective.

Economy

> ***The ERS must have an appropriate and manageable cost.***

Hospital budgets are under pressure from ever increasing demands on different fronts, including maintenance of current care standards and investing in myriad new

technologies that all promise improved treatment for patients. As the cost of a new treatment technology increases, especially if the technology is relatively unproven, the likelihood that the technology will be considered and purchased is correspondingly less.

Simplicity

System set-up and maintenance should be as straight-forward and simple as possible.

Anyone who has observed a surgical procedure has an appreciation for the complexities of the operating environment. Complexity arising from a need to ensure the well-being of a patient is tolerable, while any other sources simply exacerbate the difficulties of the OR for no good reason. Accordingly, the spread of the enhanced reality solution will be hampered by complexities in its set-up and maintenance.

Speed

The system should not add length to a procedure.

Operating room time is extremely valuable, and the risk posed to a patient increases as his/her surgical exposure and time under anesthetics grows (e.g. increased blood loss). Both these realities demand that the length of a surgery be as short as possible. As such, the system should not increase the time a procedure takes.

Transparency

The system needs to integrate naturally into the normal flow of a surgical procedure.

Additional to the complex OR environment, surgical procedures themselves are complex, and are also constrained by issues relating to the sterile operating field. The best enhanced reality systems will minimize complexity and sterility concerns by being transparent within the normal flow of a surgical procedure. Transparency in this context implies systems that are intuitive to use (for instance, provide data that is immediately and naturally understood) and require a minimum of intra-operative interactions.

During early development it may difficult to attain all of the above requirements. However, as long as they are at least respected in the design of each system component it will be easier in later development stages to build a final product that is effective and

embraced by surgeons. Thus, with this unifying set of design needs in place, the discussion can now shift focus to details of the actual system.

Design of an enhanced reality system

From a high level, the system architecture is described below and shown in figure 0.2:

(a) The system will integrate into the work flow of current minimally invasive procedures.

(b) As in current MIS systems, the endoscope of the system collects video sequences of the operative field, which is otherwise invisible to the surgical team.

(c) Prior to surgery, a set of 3-D models of the patients internal anatomy are developed using apppropriate medical imaging technology.

(d) Throughout the surgery, the endoscope video stream is integrated with the preoperative models in real-time, creating an augmented representation of the surgical site that restores depth and context information.

(e) The augmented representation is displayed on the system monitor, completing the enhanced reality system.

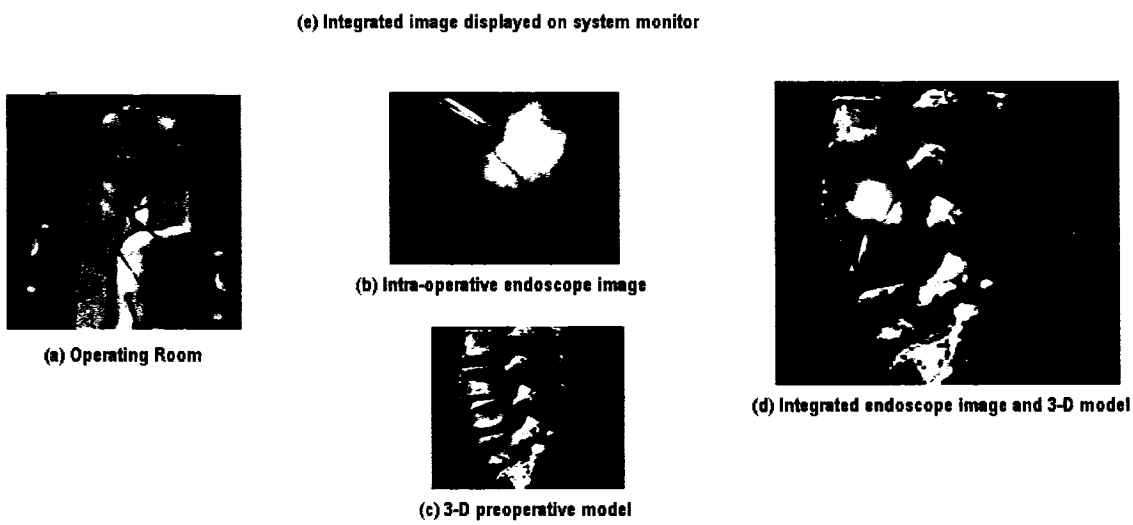


Figure 0.2: Operation of an enhanced reality system for MIS of the spine. The system integrates real-time intra-operative endoscope images (b) with 3-D preoperative models of the patient (c) to provide context and depth information about the surgical field (d).

The critical component of this system is the integrated video and preoperative model; the value the system adds over regular video-based MIS will depend on the nature of this

integration in two respects. First, context information will only by useful if the 3-D models and endoscope images are properly aligned. Second, the utility of the depth information will increase if, instead of thinking only in terms of restoring depth *perception*, the system is developed to provide actual numerical depth *values* in the integrated image. In vertebral fusion MIS procedures, this information could be used to verify exactly how much inter-vertebral disc had been excised.

To provide excellent alignment and depth information the system must be able to map objects in the endoscope frame from image units (pixels) to real world units (such as millimeters). This mapping requires a mathematical model for the camera.

Camera models

Camera models are based on a series of transformations that map real-world points to image points. Affine transformations are commonly used, and include translation, rotation, scaling and shear, resulting in the following camera model:

$$\mathbf{p}^{\mathbf{I}}=\left[\mathbf{S}(d_x, d_y)\Pi(f)\mathbf{T}(t_x, t_y, t_z)\mathbf{R}(\alpha, \beta, \gamma)\right]\mathbf{p}^{\mathbf{W}}. \tag{0.1}$$

Assuming the camera and real world coordinate systems are aligned, (0.1) models the viewing of a real world point $\mathbf{p}^{\mathbf{W}} = [x, y, z]$ with a camera as follows: The camera is rotated according to the rotation matrix **R**, and translated by the transformation **T** in order that it will be positioned to correctly align the world point and image point $\mathbf{p}^{\mathbf{I}} = [u, v]$. The 3D world point $\mathbf{p}^{\mathbf{W}}$ is then projected to the image plane of the camera via $\mathbf{\Pi}$ using the well known perspective projection model:

$$\begin{aligned} x_i &= f\frac{x}{z} \\ y_i &= f\frac{y}{z} \end{aligned} \tag{0.2}$$

where f is the focal length of the camera. At this time, the point on the image plane will still be in real-world coordinates given by $[x_i, y_i]$, which need to be scaled by the transformation **S** to get the pixel coordinates $[u, v]$ of the image point. Therefore, (0.1) fully describes a potential mapping between image and world coordinates that could be used to combine the endoscope image and 3D preoperative model. To apply such a model, methods to calibrate the camera's internal and external parameters are needed.

Camera calibration

The external parameters of the camera model include the translation and rotation parameters t_x, t_y, t_z, α, β, and γ. These values will change constantly as the endoscope moves; to update their values automatically, the enhanced reality system will include an optical tracking camera system capable of returning the positions of instruments within the world/global coordinate system.

Unfortunately, the optical tracking system alone cannot be used to determine the internal parameters of a camera model, which in model of (0.1), are those corresponding to the projection and scaling transformations. A general set of internal camera parameters includes:

- A principal point $[u_0, v_0]$: The intersection of the camera's optical axis and the image plane (often positioned at [0, 0], as in the affine camera model above);
- Scale factors $[d_x, d_y]$: The real-world dimensions of the camera's pixels;
- An aspect distortion factor τ: A scale factor that represents distortion in the aspect ratio of the camera;
- The focal length f: The distance between the image plane and camera center;
- A lens distortion parameter κ: This factor models any radial lens distortion.

These parameters are often calibrated during initialization of a system by imaging a set of points with precisely known 3D locations, after which time the parameters are considered fixed for the application. In the MIS context though, the endoscope may be re-focused, and its distance to the object being imaged can change dramatically. This reality demands a way to update the internal parameters in real-time if the enhanced reality system is to effectively map pixel-based distances to real world units.

One way to achieve an "on-the-fly" re-calibration is to identify and track objects in the image whose pixel-based movements relative to one another are also known in real world units, independently of the camera model. The instruments visible within the endoscope field-of-view during an MIS procedure should be useful for this task. Instrument shapes/contours, or features such as joints and etched text, create image objects that can be identified and tracked in the endoscope sequence. At the same time, the instruments themselves will be tracked by the optical positioning system. Even though the exact 3D

coordinates of the image objects will be unknown, their relative motions can be associated with instruments whose relative movements in real world coordinates are known via the optical positioning sensor. This link between the relative movements in pixel and real world coordinates exists independently of the camera model, and can be used to determine how the intrinsic parameters need to be updated in order to fully correlate the observed pixel movements to the known real unit movements. With the intrinsic parameters thus updated, the camera model can then be used to extract precise 3D location information for all points in the image.

Tracking and segmentation

For this calibration scheme to be successful, instrument objects must be unambiguously identified and tracked within the image. Schemes to isolate instrument objects by removing stationary objects cannot be applied because the camera also moves, so that the background also has an apparent motion. Furthermore, the surgical environment is complex, resulting in endoscope images that are characterized by (see figure 0.3):

i. Rapid intensity fall-off with radial distance from the image center;
ii. Strong specular reflections from metallic instruments and moist tissues, which change unpredictably under endoscope and instrument movements;
iii. Large changes in the appearance of objects due to motion and focusing of the endoscope;
iv. Blood, exposed tissue, and steam that blur and occlude boundaries; and
v. Cut or exposed tissues, which cause strong variations in background appearance.

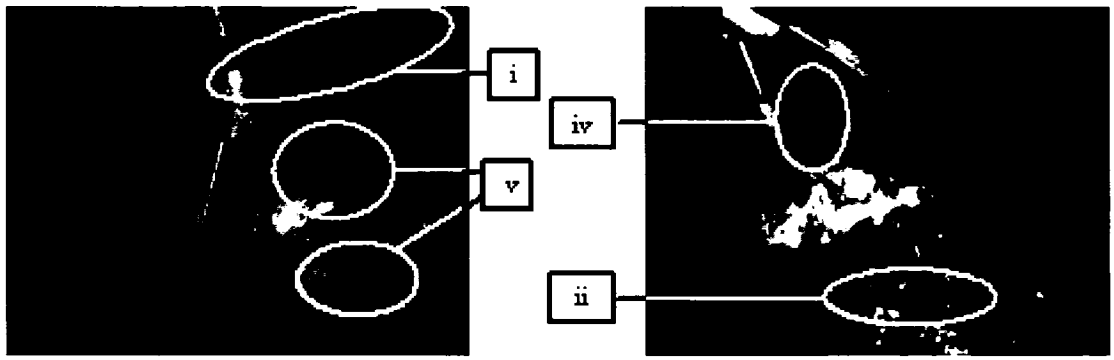


Figure 0.3: Problematic characteristics of endoscope images. Numbered areas correspond to the list of problems above. Note the difference in appearance caused by endoscope movement: while the two images are of the same instrument, only the prongs of the instrument are visible in the second image.

As a result of these complications, images can be heavily cluttered with noisy objects, making it difficult to isolate instrument objects by looking for those with similar motions.

Ultimately then, a way to segment instrument regions from an endoscope image is needed that is relatively immune to the various imaging difficulties outlined above. Once these regions have been segmented, a set of appropriate objects can be extracted for tracking and the ultimate camera calibration task.

Current segmentation methods are not well-adapted to the problems described by figure 0.3. Their limitations as well as a proposed probabilistic solution form the principle part of this thesis. The segmentation theory developed also forms the basis for an initial investigation into the final tracking solution needed to achieve the automatic parameter calibration goal. The tracking solution is by no means complete since effective development will require an investigation into what object types are most appropriate for tracking in endoscope images, and this investigation is still pending. However, the preliminary tracking results will help direct future work in this regard.

The thesis will proceed with an investigation into the state-of-the-art literature and research in the fields of surgical navigation and computer vision beginning in chapter 1. The focus of the computer vision discussions will be on the theory of regularization, and current work into segmentation and tracking. Following the literature review, chapter 2 will outline the limitations of current tracking and segmentation and tracking methods as

applied to endoscope images, and will detail a proposed solution. The results of the proposed algorithms will be analyzed and discussed in detail in chapter 3, followed by overall conclusions and proposal for future work in chapter 4. For the benefit of the reader, the specific topics of each chapter are outlined in more detail in figure 0.4 on the following page.

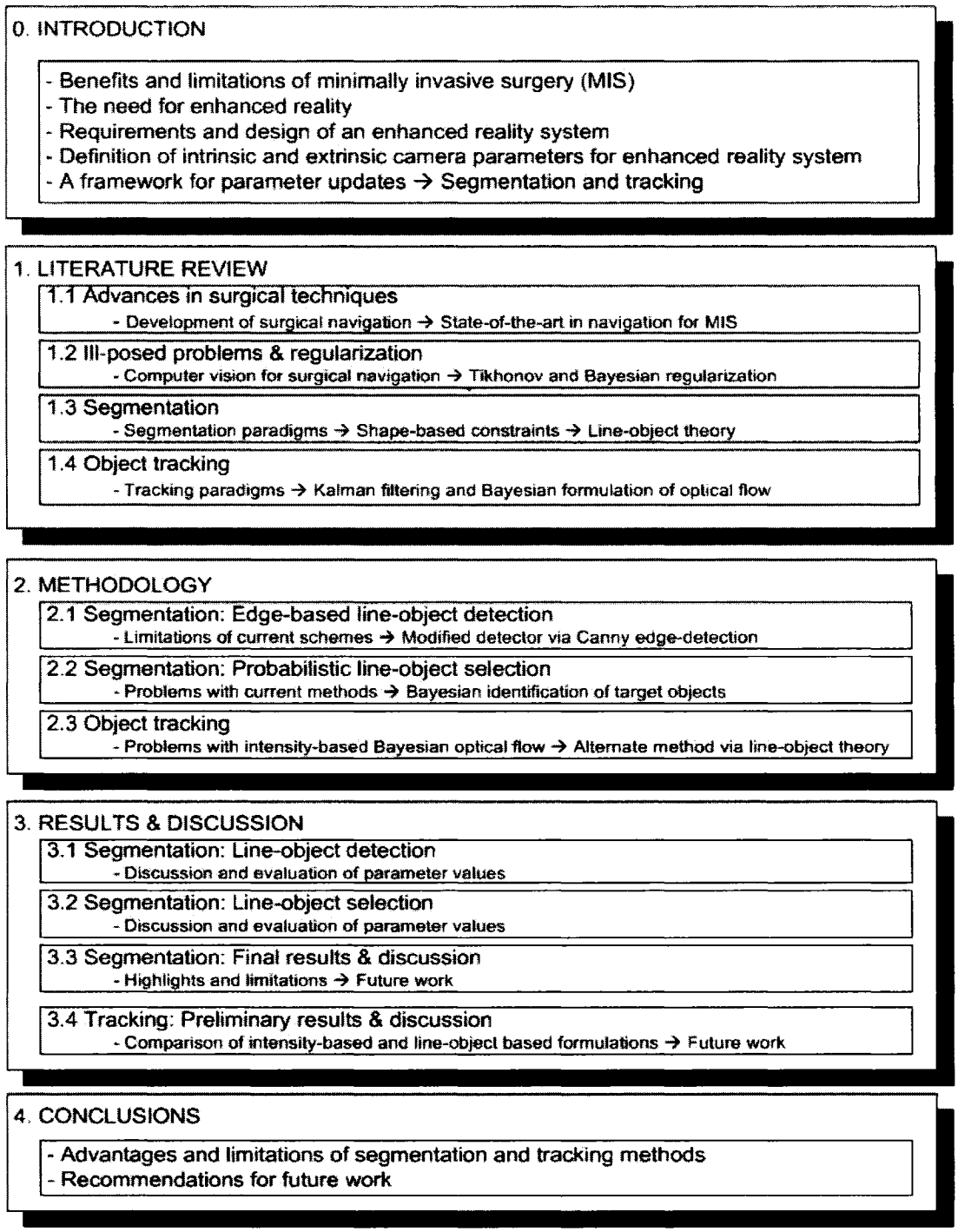


Figure 0.4: Outline of each chapter of the thesis.

CHAPTER 1 - LITERATURE REVIEW

Recall from the Introduction that the algorithms considered in this thesis for the automatic calibration objective will focus on the segmentation aspect of the object identification task and on temporal object tracking. To reiterate the importance of both, without a good set of instrument objects extracted from the first image in the sequence, it will be impossible to follow instruments in the endoscopic video so that their movements may be unambiguously correlated with the known 3D motions returned by the optical tracking system. This demands a reliable way to segment instrument regions from the image. Even with a quality set of objects extracted, the tracking step must still follow them robustly in order that their image movements are accurately understood.

These problems are not simple, and the research fields of segmentation and tracking have each been the subject of vast attention in computer vision. For the reasons outlined in the Introduction, the endoscope images used in this system are particularly challenging, and a thorough review of the existing research in each field will certainly help isolate algorithms, develop constraints, and generate ideas that will make the task of continuous recalibration less daunting. Underlying much of the current research in these domains, as well as the ultimate algorithms developed in this thesis, is the theory of regularization. Regularization concepts are importance enough to warrant distinct treatment, and this chapter will introduce the relevant theory before moving on to explore state-of-the-art segmentation and tracking research. Before moving on to any of these topics though, it will prove instructive to establish an overall context for them by reviewing the important elements of surgical navigation that have motivated the enhanced reality system in the first place.

1.1 Advances in surgical technique

The introductory discussions regarding the differences between open and minimally invasive techniques highlighted that improved outcomes of surgical procedures is a perpetual goal of research in the medical domain. Minimally invasive surgery has seen considerable development since Semm and Mouret introduced endoscopes into

appendectomies and cholecystectomies; following on their successes, researches and health care professionals quickly adapted image-based minimally invasive techniques into almost every surgical domain. However, the sensitivity of the thoracic anatomy coupled with the limited view of the surgical site slowed their introduction into spinal procedures, and spinal MIS has only advanced in any significant way since Obenchain [1991], Rosenthal et al. [1992], and Mack et al. [1993] first used MIS approaches in the early 1990s.

Of course, problems such as delicate anatomy are faced by surgeons in myriad domains: neurosurgery, vascular surgery, cardiac surgery, *etc.* are all fraught with complexities that have little margin for error. In many of these domains, navigated surgery has proved invaluable in helping surgeons improve results and these benefits can surely be extended to spinal MIS.

1.1.1 Surgical navigation concepts

Traditionally, the idea of surgical navigation is identical to the idea of the buoys used in sailing, which help ship captains navigate uncertain waters by providing a series of fixed reference points that they can compare to nautical charts. In surgical navigation, the charts may be considered preoperative patient data acquired through any modality such as X-ray or MRI scans. A series of landmarks identified on these models play the role of buoys: when they are located on the patient's anatomy intra-operatively, a correspondence with the preoperative models is established which the surgeon can refer to in order to better monitor the progression of the surgical procedure.

Just as the preoperative models can be created using a variety of technologies, the intra-operative data can be obtained in any one of many ways, from ultrasound to fluoroscopes. The landmarks used to register this intra-operative data with the prior models may also take a variety of forms – common options include significant anatomic structures, as well as fiducials affixed to the patient just before building the preoperative model. A positioning system in the operating room[1] is then used to identify the location of these markers and establish the patient's position relative to the preoperative models.

---

1 The Polaris optical tracking system manufactured by Northern Digitial Inc. is often used.

This approach to surgical navigation has wide-spread application in many surgical domains, including biopsies in neurosurgery [Nowinski et al. 2000], pedicle screw placement in orthopaedics [Kotani et al. 2003], [Sagi et al. 2003a], [Sagi et al. 2003b], *etc.* In fact, many commercial systems for surgical navigation exist from companies including General Electric [2005], ORTHO*soft* [2005], and Siemens [2005].

1.1.2 Surgical navigation applied to minimally invasive surgery

The navigation philosophy just described, seeking to register preoperative and intra-operative data via a series of landmarks, is typical of traditional surgical techniques but not video-based MIS. Nevertheless, in light of the challenges faced in MIS the potential benefits in adapting surgical navigation are significant.

The clear difference with MIS exposures is the presence of a camera in the surgical zone. Since this camera provides the only visual support for the surgery, navigation of MIS naturally implies registering this intra-operative image data with the preoperative models. As discussed in the Introduction, establishing this correspondence will allow the image data to be integrated with the preoperative models so that context and depth cues are re-established. Navigation systems in this spirit have already been established for laparoscopy (minimally invasive surgery of the abdomen) by Mourgues [2003] and Shahidi et al. [2002]. Both are limited in that the endoscope's internal parameters are calibrated once at the beginning of the procedure and are then assumed fixed. In consequence, the surgeon cannot alter the zoom or focus of the endoscope without having to recalibrate the system.

Zhang and Payandeh [2002] developed a system that allowed for dynamic and automatic parameter recalibration by affixing artificial markers to the instruments visible in the endoscope video. When these markers were identified in the image sequence, their relative movements could be correlated with other data reported by a global positioning system to update camera parameters. This approach is in exactly the same vein as the overall system outlined in the Introduction; however, artificial features are undesirable because they complicate the instrument sterilization process. Recall that one of the principle design objectives of the enhanced reality system is to integrate as seamlessly as

possible into existing surgical work-flows, without introducing additional complexity via surgery set-up. Accordingly, a preferred approach would be to use as large a set of natural instrument features as possible so that equipment can be used as-is. This desire motivates the object segmentation and tracking work of this thesis.

Unfortunately, the characteristics of endoscope images described earlier can make the process of isolating natural instrument objects for tracking extremely difficult. In fact, both segmenting and tracking these objects under such poor image conditions are impossible without introducing prior knowledge into the problem to constrain the solutions. Use of prior knowledge in this way falls under the theory of regularization, which is itself a deep field of research that underlies many methods in computer vision. Because regularization will not only underlie both the segmentation and tracking solutions developed later, but is also fundamental to many of the state-of-the-art segmentation and tracking methods to be discussed in this chapter, general concepts of the theory will be introduced next.

1.2 Ill-posed problems & regularization

The goal of any sensing system is to observe the "state" of the environment (i.e. world) in which it exists. Within this framework, a vision system encodes images of 3D scenes/objects via a sensing process R that specifies a set of data u to represent the real-world condition z, according to following model of image formation:

$$u = Rz. \tag{1.1}$$

In the computer vision domain, the opposite problem is of more interest: given the observed data only, what can be inferred about the real world conditions? This is the "vision problem," given by the inverse process:

$$z = R^{-1}u. \tag{1.2}$$

1.2.1 Ill-posed problems

A persistent and difficult problem faced throughout computer vision research is the amount of information lost during the image formation process when the 3D scene is encoded as a (typically) 2D image. The resulting ambiguity when attempting to apply the

inverse process in (1.2) means that vision problems are often *ill-posed.* A well-posed problem in the sense of Hadamard [1923] is one for which three conditions are met:

- *Existence:* For every element $u \in U$ there exists a solution $z \in F$, where U is the space of all possible data, and F the space of solutions;
- *Uniqueness:* The solution must be unique; and
- *Continuity:* The solution depends continuously on the data, so that if the error in the data tends towards zero, the error in the solution also tends towards zero.

Information loss in the imaging process or uncertainty about the exact form of the sensing process R can mean that one or more of the above conditions will be violated, such that a problem will be ill-posed and the inverse process cannot be directly applied. Moreover, discretization of a problem means that even a solution that otherwise depends continuously on the data may still not be robust to noise. This is the familiar problem of system instability under ill-conditioning of R [Betero et al., 1988].

To contend with these complications, methods are required to recover approximate solutions by relying on natural constraints and assumptions about the physical world. *Early vision* problems—those which correspond to conceptually independent preliminary steps such as edge detection, optical flow, shape from shading, etc.— are commonly ill-posed, and Poggio and Torre [1984] recognized the use of the regularization concepts applied in mathematics and mathematical physics to help constrain their solutions.

1.2.2 Regularization

Regularization methods are formed on the idea of using prior knowledge and constraints relevant to an ill-posed problem to restrict the domain of possible solutions. The classic regularization framework introduces a regularization operator Q exhibiting the property that, for a given metric ρ, if $\rho_U(u_t, u_\delta) \rightarrow 0$ then $\rho_F(z_t, z_\delta) \rightarrow 0$. The typical regularization functional is given by:

$$M^{\alpha}[z, u_\delta] = \rho_u(Rz, u_\delta) + \alpha \Omega[z] \tag{1.3}$$

In this equation, the first term on the right-hand side measures the consistency of the predicted and observed data; that is, the difference between Az (the data expected if the current solution z were true) and the actual observed data. The stabilizing functional

given by $\Omega[z]$ represents the likelihood of the current solution z with respect to any *a priori* knowledge of the problem at hand. α is a Langrange multiplier which weights the trade-off between the data-consistency and prior model terms. Values and forms for the metric ρ, stabilizer Ω, and multiplier α, are specified in advance. Thus, for a given set of data u the solution to the ill-posed inverse problem of (1.2) is found by testing numerous values for z and retaining the one that minimizes (1.3):

$$\hat{z} = \underset{z \in F}{argmin}\, M(z) \tag{1.4}$$

Tikhonov Regularization

The standard regularization framework is due principally to the work of Tikhonov and Arsenin [1977], for which the metric ρ applied is the standard quadratic norm, and the stabilizing functional is given by:

$$\Omega[z] = \int_a^b \sum_{r=0}^{p} q_r(x) \left(\frac{d^r z}{dx^r} \right)^2 dx \tag{1.5}$$

This term is a weighted sum of derivatives squared and represents the energy in an interval. When $q_r(x)$ is constant it is called a *Tikhonov stabilizer*. It is particularly common throughout work in early vision problems (see, for example, [Horn and Schunck, 1981], [Franko and Chellapa, 1988], [Hildreth,1984]).

Bayesian Regularization

While the Tikhonov regularization formulation has seen wide-spread application, it is also somewhat limited. In particular, as discussed in the paper by Marroquin et al. [1987] there are two particularly problematic issues. For one, too high an order of the regularization operator may cause the loss of important data, such as image discontinuities (i.e. occluding boundaries). Further, the solutions of regularization problems formulated in this way frequently lead to linear Euler-Lagrange equations which cannot be combined with output from other early vision modules in any way but linear. Both these problems imply choices for the stabilizing functional $\Omega[z]$ in (1.3) are limited, constrained either by concerns of retaining useful data for later processing, and/or by issues of mathematical convenience.

A different and somewhat more intuitive approach to regularization follows from probability theory. Given two simultaneous events A and B that have no (explicit) temporal order, the probability of event B happening, conditional on event A, is given by Bayes Theorem:

$$P(B|A)=\frac{P(A|B)P(B)}{P(A)} \tag{1.6}$$

If event B represents the solution z in the (inverse) image problem (1.2) and A is taken as the observed data, u, then a solution to the inverse problem is given by the value of z that maximizes the conditional probability given in (1.6): $P(z|u)$ The terms in this equation can be interpreted as follows (nb. $z \equiv B$, $u \equiv A$) :

- $P(u|z)$ – The first term in the numerator of the right-hand side reflects the forward problem given by $Az = u$, and noise. Given a candidate solution (i.e. real world conditions), this term reflects the data expected from the imaging system.
- P(z) – The second term in the numerator is the *a priori* probability density of the solution, and encompasses any prior knowledge about the solution before measurements are made. This prior model distinguishes Bayesian techniques from others. For example, the maximum likelihood (ML) estimate of the solution is the value of z that maximizes simply the conditional probability, P($u|z$).
- P(u) – The denominator is the *a priori* probability density of the data before any measurements are made. This term is typically ignored since it is independent of the value of z, and does not impact which value of z will optimize (1.6).

The spirit of this and the Tikhonov regularization schemes are the same: test various candidate solutions subject to data expectations and prior knowledge about the solution, ultimately choosing the one that, in the case of (1.3), minimizes the functional $M[z]$ or, in the case of (1.6), maximizes the conditional probability P($z|u$). However, the Bayesian approach is more general; in fact, the Tikhonov approach is a specific case of the Bayesian formulation. The two techniques are equivalent if $P(z|u)$ is cast in the form of a Gibb's distribution [Geman and Geman, 1984]:

$$P(z|u)=\exp\frac{(-\beta E(z,u))}{Z} \tag{1.7}$$

In this equation, $E(z,u)$ is the free energy of a system corresponding to a configuration z, and $P(z|u)$ is the probability of the system configuration z, given measurements u. The equivalence develops if the regularization functional, $M''(z,u)$ is associated with the energy functional $E(z,u)$, such that minimizing the system energy will maximize the probability in (1.7).

More details respecting Bayesian probabilistic methods in general can be found in textbook by Duda et al. [2001]. Specific ideas will also be addressed as needed in the following discussions of segmentation and tracking research, since Bayesian regularization is an important element of many of them. In particular, the shape-based segmentation method proposed by Leventon et al. [2000], and the optical-flow tracking scheme of Mansouri [2002] are deeply rooted in this theory. Moreover, since both methods have had strong influence on the proposed segmentation and tracking methods presented in the Methodology chapter, Bayesian concepts permeate the work of this thesis in an intimate way and future discussions will return to these concepts frequently, helping give them more context.

1.3 Segmentation

The preceding regularization discussion was intended to introduce general concepts that will be important to many of specific methods considered to solve the actual object identification and tracking problems. This section now initiates an investigation into these specific methods via a focus on the current state of image segmentation research.

1.3.1 Contour-based segmentation paradigms

All segmentation methods aim to partition an image into meaningful/descriptive regions. There are myriad segmentation approaches, and when the goal is to isolate a target object or area in the image, the use of segmenting contours is extremely popular (although other techniques certainly exist). In this case, a segmenting curve is evolved until some measure of its agreement with expected object boundary locations is optimized.

In contrast, if the objective of the segmentation is to achieve a descriptive measure of similarity between all areas of an image, then the segmentation is typically implemented

via an iterative clustering scheme. While contour evolution is effectively a form of clustering, pure clustering will continue until the regions in the image maximize some metric on intra-region homogeneity and inter-region heterogeneity. Compared to contour-based segmentation, which will produce two region *types*—background and foreground/target—there may be a large number of a regions once the clustering completes, and these will not necessarily have meaningful labels.

For the endoscope segmentation task, the objective is unquestionably to locate a specific (set of) target object(s), and there is little use trying to segment the background tissue into anything more descriptive than "non-instrument." Therefore, the focus of this section is on contour-based schemes for isolating target regions in an image, beginning with a discussion of the contour-evolution mechanics that underlie many segmentation methods. Two approaches will be considered in detail: active contours and geometric based evolution schemes.

Particle methods: Active Contours and Snakes

From a functional perspective, the form of the initial segmenting boundary in an image is usually quite arbitrary, and could be the entire image, a small region inside the target object, or otherwise. Regardless the initialization, the objective is to develop an evolution scheme that can evaluate the quality of the growing contour with respect to expectations about the real region boundary, stopping or continuing the evolution accordingly.

Active Contours

Blake and Zisserman [1987] proposed using energy minimization for boundary deformation. Evolving a segmenting contour via energy minimization leads to the (general) concept of *active contours*. In this framework, an initial contour is broken into a set of equally spaced particles (see figure 1.1).

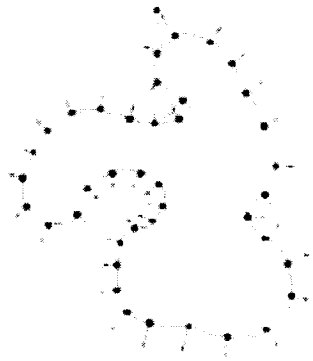


Figure 1.1: Parameterized representation of a boundary. The boundary is represented by a set of particles that break the curve into equal length segments. In the active contour model, interactions between adjacent particles influence how each particle is moved over a time step. (Source: Sethian [1999]).

The contour is then approximated via a spline fitted to the marker particles. In a purely particle-based evolution, the parameterized contour will be evolved by moving each particle according to some scheme: discretized equations of motion are typical, with finite difference equations computed via neighbouring particles. Active contours use a more sophisticated evolution scheme, where neighbouring particles are not just used to calculate finite differences, but also to constrain motion by, for example, penalizing high curvature of the spline joining several particles. Therefore, the movement of particles during a time step Δt depends on a combination of internal and external forces:

- Internal forces are those arising from interactions with neighbouring particles, such as smoothness requirements, perimeter limits, or the property of being a simple curve [Shapiro and Stockman, 2001].
- External forces might include designated points past which the contour cannot move, regardless the penalties incurred (e.g. from high curvature).

The active contour will evolve towards a minimum energy configuration, subject to these internal and external forces. If the forces are well set-up, the minimum energy state should occur when the contour coincides, or closely approximates, the actual boundaries of the target object.

Snakes

The classic implementation of the active contour approach is the "Snake" model proposed by Kass et al. [1988], which has also become the defacto definition of active contours in the vision community as opposed to the more general definition just described. The Snake energy minimization scheme evolves an initial parameterized contour C_0 towards a target object's boundaries via the following functional:

$$E(C)=\alpha\int_0^1 |C'(q)|^2 dq+\beta\int_0^1 |C''(q)|^2 dq-\lambda\int_0^1 |\nabla I(C(q))|dq \tag{1.8}$$

This functional makes explicit the (Tikhonov-like) regularization achieved by the internal forces of the active contour approach:

- The first two terms, which are squared magnitudes of the first and second derivatives of the parameterized contour, control the smoothness of the evolving contour. They represent the internal forces and play the role of the stabilizing functional $\Omega[z]$ in (1.3), helping attenuate the influence of local imperfections such as noise by penalizing the development of sharp corners.
- The third term acts as the data consistency term, working to attract the deforming contour towards the object boundaries by seeking the high gradient values typical of object edges.

Therefore, the final image segmentation is given by the contour C that minimizes E, and will be a close, but somewhat smoothed, approximation of the real object boundary when the multipliers α, β, and λ are chosen appropriately.

Before concluding this sub-section, it is important to note that image measures other than the intensity gradient can be used to constrain the active contour approach. For example, Xu and Price [1997] use a field of intensity gradient *vectors* to increase the likelihood that a contour will move into locally convex boundary areas where high curvature might stop an evolution controlled by the original scheme in (1.8).

Geometric methods - Level sets and Geodesic Active Contours

A significant difficulty encountered with particle-based boundary segmentation methods is the need for an explicit parameterization. As the segmenting contour moves in time,

the spacing between the marker particles will change, either increasing or decreasing depending on whether local image characteristics cause the boundary to expand or shrink. If the movement of particles along the boundary over a time step is based on quantities that depend on the inter-particle spacing, such as numerical differences, then the points need to be regularly redistributed to maintain the accuracy and stability of the numerical algorithms. Already this is a complicated process; the problem is compounded if there are also topology changes in the segmenting contour. Consider using an active contour energy minimization scheme to segment two target objects with the segmenting boundary initialized as a single closed contour surrounding both objects. Not only will the marker particles need to be redistributed as the contour evolves, but the contour will eventually meet itself[2], at which point certain particles need to be eliminated from the parameterization—not just redistributed—if the desired pinching-off into two objects is to happen. Without explicit bookkeeping schemes to manage the particle redistribution and elimination the topology change between the initial contour and targets cannot be handled. Effective and reliable schemes to accomplish the particle management are seldom obvious or intuitive to design.

Level Sets

An alternative to marker-based contour evolution schemes was proposed by Sethian and Osher [1988] and analyzed extensively in the textbook by Sethian [1999]. In their framework, the contour C to be evolved on the xy image plane is embedded within a higher dimensional surface $\varphi(x,y,t)$, called a level set function. This function is typically initialized to the signed distance function: at time $t = 0$, each grid point (x,y) is assigned a height $z = \varphi(x,y,t = 0)$ that is equal to the distance of that point to the initial boundary C_0. If the grid point is inside the boundary, the distance is set negative. Therefore, C_0 is the zero level set of $\varphi(x,y,t = 0)$. This initialization is represented in figure 1.2.

2 Assuming other constraints do not stop it at the point where the contour takes the form of a convex-hull that approximates the aggregate shape of the two objects.

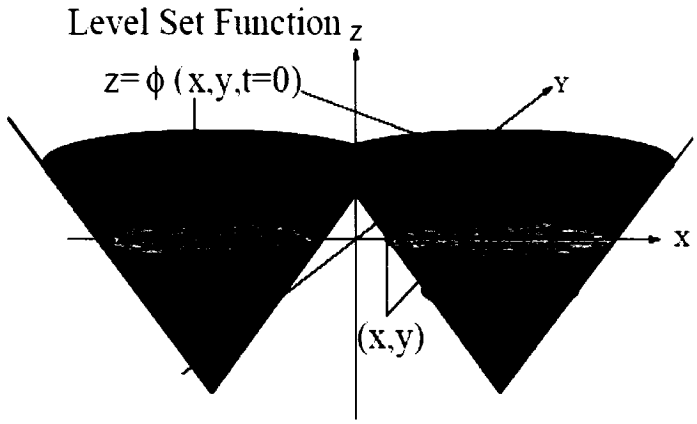


Figure 1.2: Initialization of the level set function to the signed distance function. Higher dimensional level set surface shown in red with the zero level set corresponding to C_0 in blue. C_0 is initially two distinct contours. (Source [Sethian, 1999].)

Conceptually, the initial contour is propagated in time by letting the higher dimensional function (red/dark grey in above figure) expand, shrink, rise or fall. As the surface moves, the value of φ at each point (x,y) on the grid is simply its current distance to the interface. Therefore, the interface is evolved not by tracking and bookkeeping marker particles but by simply monitoring the value of φ at each image pixel and returning those that are zero. Since different topologies of the contour do not imply different topologies of the higher dimensional function $\varphi(x,y,t)$, contours are also able to freely change topology as shown in figure 1.3.

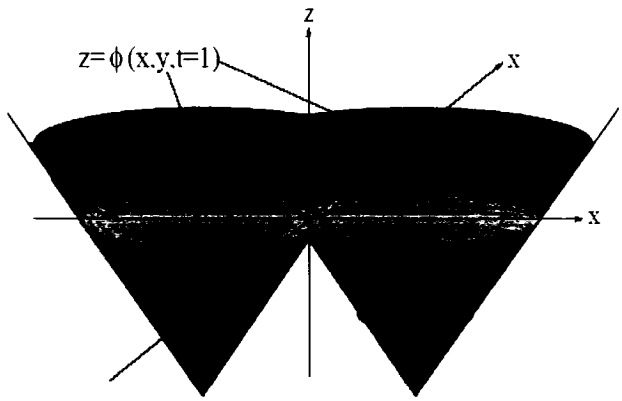


Figure 1.3: Topological independence of level sets. Time evolution of the level set surface from figure 1.2. Since the evolving contour C is given simply by the zero level set, the topology change resulting from the merge of the two initial contours occurs naturally. (Source: [Sethian, 1999].)

To actually control the evolution of φ, and hence the contour C, Osher and Sethian use a speed function F which acts outward normally at each point along the curve. Because a point on the initial contour must always remain on the contour as it moves along a path ($x(t)$,$y(t)$), and since the evolving contour is always given by the zero level set of φ, it follows that for points on the contour: $\varphi(x(t),y(t),t) = 0$. The time evolution of φ follows immediately from the chain rule:

$$\varphi_t + \nabla\varphi(\boldsymbol{x}(t),t)\cdot\boldsymbol{x}'(t)=0 \quad with \quad \boldsymbol{x}(t)\equiv(x(t),y(t)). \tag{1.9}$$

As stated, the contour propagates in its outward normal direction, so that $\mathbf{x}'(t)\cdot\mathbf{n} = F$, with $\mathbf{n}$ given by $\nabla\varphi/|\nabla\varphi|$. Therefore, the evolution of the surface φ is given by the solution of the following time-dependent initial value problem:

$$\begin{aligned} &\varphi_t + F|\nabla\varphi| = 0 \\ &given \quad \varphi(x(t),y(t),t=0) \end{aligned} \tag{1.10}$$

At any time the position of the evolving contour C_t is obtained by simply extracting the zero level set: $C_t=\{(x,y) \mid \varphi(x(t),y(t),t)=0\}$.

This scheme applies to evolving contours in higher dimensions as well, and has received a high degree of attention in the literature since its introduction (see for example the recent review of medical imaging applications by Suri et al. [2002]). The popularity of the approach stems from its parameter free, hence intrinsic, representation of the

contour; its numerical stability; its topological independence; and its high adaptability via different formulations of the speed term F, which can depend on any sort of complex physics or other constraints.

Geodesic Active Contours (GAC)

A well-known level set-based segmentation scheme was proposed by Caselles et al. [1997] as an alternative to the Snake model of active contours. As previously described, the active contour model is limited in the sense that it cannot naturally handle topology changes. To address this problem, Caselles used the concept of geodesic contours in a Riemannian space to establish an equivalence between the Snake model in (1.8) and the following contour evolution functional:

$$\frac{\partial C(t)}{\partial t} = g(I)\kappa\vec{N} - (\nabla g \cdot \vec{N})\vec{N} \tag{1.11}$$

In this functional, g is constrained only to be a strictly decreasing function on the image with the property that $g(r) \to 0$ as $r \to \infty$. A common form is $g = \frac{1}{1+|\nabla \hat{I}|^p}$ with $\hat{I}$ a smoothed version of the image and $p = 1$ or 2. In homogeneous image regions $g(I)$ will be high, and the contour will evolve according to the magnitude of the contour curvature in its normal direction. At edges, $g(I)$ will tend to 0, stopping the evolution. Unfortunately, edges are rarely ideal so, at weak edges especially, $g(I)$ cannot be guaranteed to be 0, and the contour may evolve past real object boundaries. The second term in the right-hand side of the (1.11) is added to address this shortcoming: the gradient of g will be characterized by valleys centered on edge-like objects, and this regularizing term will tend to push the contour into these valleys regardless of the edge strength.

By defining a curve-normal speed function $F = g(I)\kappa - \nabla g \cdot \vec{N}$, Caselles et al. were able to implement (1.11) in the level set framework given in (1.10), yielding the well-known GAC model for contour evolution:

$$\frac{\partial \varphi}{\partial t} = g(I)|\nabla \varphi|\kappa - \nabla g \cdot \nabla \varphi. \tag{1.12}$$

Like the Snake model in (1.8), the GAC method evolves a contour according to

gradient strength and curvature constraints (with curvature given by κ). Yet unlike the Snake model, a particular advantage of (1.12) is that the segmenting curve $C_t=\{(x,y) \mid \varphi(x(t),y(t),t)=0\}$ now evolves with an intrinsic parameterization (i.e. its geometric curvature) rather than an explicit parameterization of the curve. Moreover, since the level set technology naturally handles topology changes, the evolving curve is free to split and merge as the situation requires, and multiple objects can be easily segmented.

With the fundamentals of the two most popular contour evolution methods now in place, it is worth emphasizing that the basic forms of the Snake and GAC models both apply intensity gradients act as data-consistency terms to control the curvature evolution. However, other image measures can be applied, and the issue now becomes whether any of these are better suited to the endoscope image segmentation task.

1.3.2 Intensity and colour segmentation constraints

The choice of the image data used to control the contour evolution is vitally important so that the curvature-based regularization aspects of either the Snake or GAC evolution model do not over-constrain the evolving contour, creating too smooth boundaries.

Intensity-based constraints

As might be inferred from their use in both the basic Snake and GAC formulations, image gradients are intuitive choices for segmentation constraints since one expects an object's occluding boundaries to generate sharp intensity discontinuities (i.e. edges). Unfortunately, in many applications, especially endoscope imagery, intensity gradients alone as a segmenting constraint can be misleading, if not wholly unreliable. Consider the endoscope image in figure 1.4(a). Even after strong blurring via Gaussian convolution (a filter with $\sigma = 3$ has been applied to generate the image in (b)), there exist strong gradients in the textured background (as represented by the edge content shown in part (c) of the figure) that will confound gradient-based measures of the object's true occluding boundaries. This will make it difficult, if not impossible, for a segmentation scheme based on gradient information alone to accurately extract the target region.

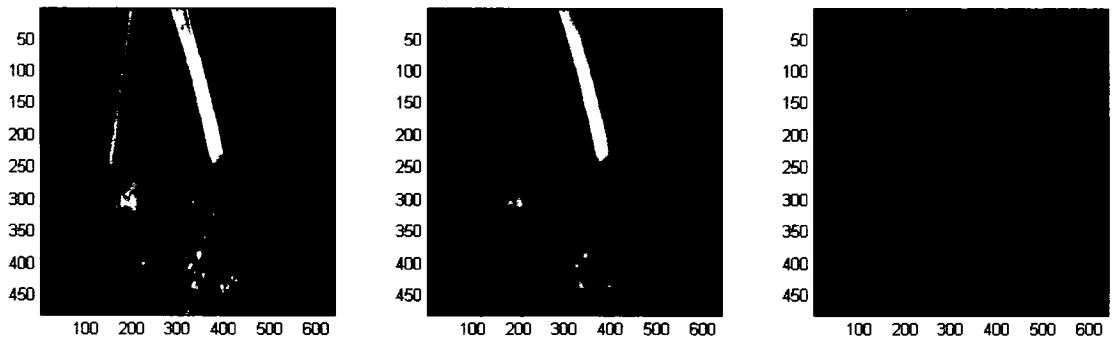


Figure 1.5: Problem of poor instrument contrast for segmentation. If the segmenting contour is initialized to a point inside the instrument, the final result is acceptable (a). However, when the contour is initialized outside the instrument, the poor contrast of the real instrument edge allows the contour to enclose both instrument and background regions (b).

In the case just illustrated, the principal difficulty is the presence of strong background gradients, but at least the object itself has good contrast with its background. In endoscopy the opposite extreme is also frequently encountered, and the object itself may have very weak contrast. This situation is shown in figure 1.5, and serves to underscore the general problems expected when using constraints based only on intensity. In figure1.5a, a segmenting Snake using the Gradient Vector Flow approach [Xu and Prince, 1997] was initialized to the interior of the instrument region, and eventually stabilized at the boundaries of the strong reflection on the instrument surface. In (b), the contour was initialized to a region outside the instrument, and the evolution was stopped after 150 iterations because the result was unacceptable – the contour had leaked over the real instrument boundary, part of which has been superimposed in the image for clarity. The problem here is the strong dependence of the contour evolution on intensity/edge content in the image: the real instrument edge is weak, and evolution is only constrained by the apparent edge caused by the reflection.

Moreover, this case is also illustrative of the problem of textured backgrounds. The evolution of the lower portion of the segmenting region has prematurely stopped due to the edge between the background tissues at different depths.

Finally, the weak instrument contrast of figure 1.5 compared to the good contrast of

figure 1.4 suggests that even if an intensity-based measure could be adapted to overcome background texture problems, it will be exceedingly difficult to find a solution that can also handle the polar extremes of contrast expected.

Colour-based constraints

Continuing with the analysis of the two previous figures, other segmentation constraints like colour descriptors may also be difficult to use reliably. Textured backgrounds will be particularly difficult to handle, since reflections and exposed tissues mean these backgrounds frequently contain similar colours as the instrument regions. Consider the exposed cartilage just under the left instrument prong in figure 1.4, or the numerous areas of apparently homogeneous intensity around the instrument in figure 1.5. In this latter case, the perceived structure of the instrument makes clear where the boundaries of the instrument exist, but colour statistics alone cannot exploit such structure and would have a hard time differentiating the lower tip of the instrument from the surrounding tissue.

1.3.3 Shape model constraints for contour-based segmentation

The concluding remark of the last discussion brings to focus an important idea for endoscope image segmentation: the perceived structure of the instrument helps define where the boundaries of the object exist, even under poor contrast. This prompts an investigation into how segmentation can be aided by applying models of target object shape.

Bayesian shape constraints for GAC contour evolution

Leventon et al. [2000] accomplished shape-based control for contour evolution by fusing Bayesian expectations of target object shape into the GAC level set model. The shape constraint was developed by first embedding the boundary of a target object as the zero level of a signed distance map. This was done for a set of training images yielding a series of distance maps φ_i that encoded typical shapes of the target object. These training images were combined to build a mean surface $\mu = \frac{1}{n}\sum \varphi_i$ that was subtracted from each training image to produce a mean offset map $\hat{\varphi}_i$. The series of mean offset maps

were used in a singular value decomposition (SVD) to establish a matrix Φ representing the orthogonal modes of shape variation, and a diagonal matrix Σ of principle components.

An estimate of a novel shape φ_n from the same object class could thus be obtained by first building a coefficient vector corresponding to the first k-principle components describing the shape: $\alpha=\Phi_k^T(\varphi-\mu)$. The shape estimate was then reconstructed from Φ_k and μ as $\tilde{\varphi}=\Phi_k\alpha+\mu$. Therefore, assuming a Gaussian distribution of object shape, the probability of an evolving curve representing the actual object segmentation at a given instant was given by its agreement with the prior shape model:

$$P(\alpha)=\frac{1}{\sqrt{(2\ pi)^k|\Sigma_k|}}\exp\left(\frac{-1}{2}\alpha^T\Sigma_k^{-1}\alpha\right) \tag{1.13}$$

This *a priori* expectation of shape was then integrated into a Bayesian regularization scheme, with the goal of obtaining the MAP estimate of the shape parameter α and pose parameter p of the final segmenting contour:

$$\langle\alpha_{\text{MAP}},\ p_{\text{MAP}}\rangle=\underset{\alpha,\text{p}}{\text{argmax}}\,P(\alpha,\ p|\varphi,\ \nabla I)=\frac{P(\varphi|\alpha,\ p)P(\nabla I|\alpha,\ p,\varphi)P(\alpha)P(p)}{P(\varphi,\ \nabla I)} \tag{1.14}$$

Recall from section 1.2.2 that the likelihood terms in this equation represent data consistency measures. Leventon based the first term on image gradients, and the second on a design constraint that the segmenting contour must be initiated at a point inside the target region. The third term is the prior model just developed, and the fourth is a prior distribution on object pose, taken to be a uniform distribution.

The final Bayesian probability (1.14) on the evolving contour was finally added as an additional constraint to the basic GAC curve evolution method described in (1.12):

$$\frac{\partial\varphi}{\partial t}=g(I)|\nabla\varphi|\kappa-\nabla g\cdot\nabla\varphi-\lambda_B(\varphi^*-\varphi). \tag{1.15}$$

where φ^* represents the MAP estimate of the final curve, and φ is the evolving contour.

Towards more general shape representations

While the probabilistic use of shape models as a constraint for contour evolution is certainly desirable, there are some drawbacks to Leventon's method that seriously limit

its applicability to endoscope images. For one, via the likelihood term, the approach is still rooted strongly in gradient measures. Tsai et al. [2003] used the same shape model, but addressed the gradient issue by exploiting region-based intensity statistics in a gradient descent search for the optimal shape and pose parameters; the use of region-statistics made their method less sensitive to local variations in image intensity characteristics. However, more serious limitations of shape-constraint methods such as that used by Leventon and Tsai exist.

First of all, these shape-models rely on a good set of training data to build a comprehensive set of principle components representing expected variations in the target object shape. Given the dramatic changes in the appearance of an instrument that occur in endoscope images due to camera movement and changes in the actual instruments used, building a good set of representative principle components is intractable.

Furthermore, in both of these methods, as with many other segmentation methods, the segmenting contour is assumed to be initialized within the target object. Again due to instrument movement, the location of target objects in any particular endoscope image is unpredictable. Requiring manual initialization would violate earlier design objectives to minimize required interactions from the surgical team. An automatic initialization would be desirable, but this demands a way to (coarsely) locate the object in the image. If a more general shape model could be developed, then this could be used to simultaneously identify the approximate location of the target object and constrain a segmentation refinement via contour evolution.

1.3.4 Line structures

Toward such a general shape model, observe that the instruments used in minimally invasive surgery must be both narrow enough to fit through the incision on a patient's body, yet long enough to reach the surgical site. This long-but-thin characteristic will exist at a very high perceptual level compared to other object identifiers, likely resulting in a high degree of stability under changes in reflections, intensity, colour and illumination/contrast. Accordingly, a shape model for endoscopy segmentation that is not only specific enough to effectively identify target object regions, but also sufficiently

adaptable to the strong changes in appearance of the target instruments, might exist within the theoretical framework of line-structure segmentation.

Line-structures at multiple scales

Line-structure detection concepts have been researched for some time[3], but applications were limited because early work in the field did not adequately consider that line-structures can exist at multiple scales. Like a river on a map that will appear large or small based on the map's scale, endoscope instruments can appear wide or narrow depending on their relative distance to the camera. This multi-scale property must be accounted for if line-structure concepts are to serve effectively in the segmentation task.

Koller et al. [1995] developed a multi-scale line-structure segmentation method independent of a structure's profile (e.g. bar, roof, or smooth ridge). They noted that if a line-structure of width w (assume for this discussion that the structure has a bar profile) is convolved with the first-derivative-of-Gaussian kernel at a scale of σ_f across its profile, the resulting filter response will have two sharp peaks centered at $\pm\, w/2$. One peak will be negative, and Koller proposed the following left and right detectors which are applied to the result of the convolution: $E_l = -G'_\sigma(x+s)$ and $E_r = -G'_\sigma(x-s)$. In these equations, s is assigned the scale of the Gaussian kernel, and is set to keep the support of the filter within the line-object: $s = \sigma \approx w/2$. The resulting left and right filter responses are shown in the following figure:

3 See the introduction of the paper by Koller et al. [1995] for a comprehensive set of references.

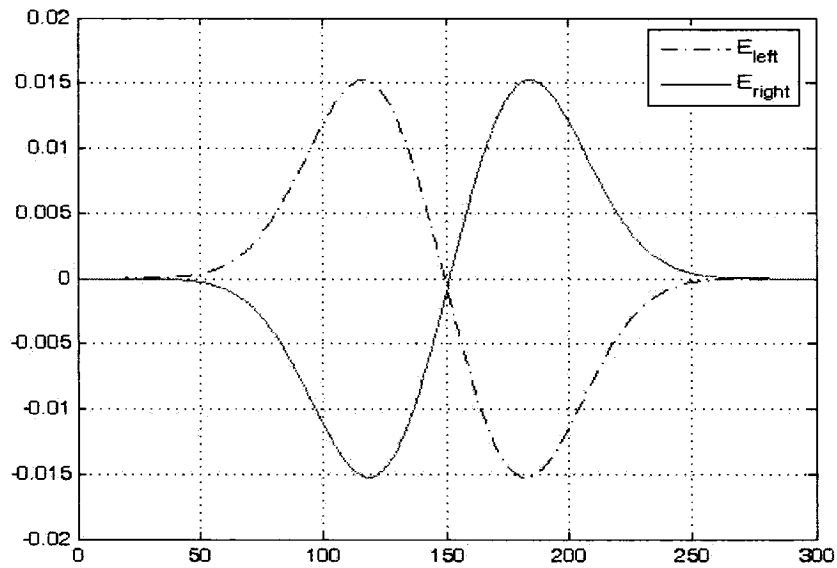


Figure 1.6: Line-object convolution response to first-derivative-of-Gaussian kernel. The two edges of the line-structure at opposite extremes of its profile will generate the left and right responses shown.

These filter responses are then combined in a nonlinear way to produce the following line-structure detector:

$$R_s(x)=\min\left(Pos\left((E_l * f)(x)\right), Pos\left((E_r * f)(x)\right)\right) \tag{1.16}$$

where the 'Pos' operators are applied to keep only the positive part of each response, accounting for the fact that the response to an ideal line-structure will have equal positive and negative peaks. Further, the min operator reflects the requirement that a good line-structure has both a good left and right response.

The multi-scale characteristic of the detector follows from the application of this filter at numerous Gaussian filter scales. The best line-structure at a given point is given as:

$$R_{multi}(x)=\max_s\left[R_s(x), s_L \leq s \leq s_H\right]. \tag{1.17}$$

This multi-scale scheme can be applied directly to images by realizing that the response given in (1.16) will be high across the image profile of a line-structure (the direction orthogonal to its length). Therefore, second derivatives are calculated at each image point, and the multi-scale line response (1.17) is evaluated in the direction of the maximal second derivative. After applying a non-maximum suppression scheme to the

set of filter responses in the second derivative directions, the final line-structure segmentation of the image is represented by a set of pixel chains centered along the line objects in the image – each pixel in each chain encodes the scale of the maximal response (i.e. width of the corresponding line object) and the local orientation.

Although Koller showed the successful application of this method to a variety of images, including extensions to the segmentation of vascular trees from 3D MRI volumes, it encounters a fundamental difficulty when used with endoscope images. The line-responses returned at each image point and scale will depend on the direction in which the filter (1.16) is applied, and this steering is accomplished via the second derivatives. Because the filter considers two discreet points left and right of each image point, if the second derivatives at an instruments actual scale s are slightly misaligned with the object profile due to local[4] variations in the image intensity, then the filter may not produce the desired across-scale maximum.

Eigenvalue line-structure response

This limitation of the Koller method was addressed by Lorenz et al. [1997] by using an eigenvalue formulation for line-structure detection. Like Koller, they exploit the fact the variation in the second derivatives about an image point should be maximal across the profile of a line-structure. However, they use this information directly to evaluate the line-like quality of the underlying image data, rather than use it only as a steering mechanism.

If the eigenvalues of the Hessian matrix $H = |\partial^2 f / \partial x_i \partial x_j|$ at a point $\mathbf{x} = (x_i, x_j)$ are ordered by magnitude such that $|\lambda_1| > |\lambda_2|$, then the relationship between these eigenvalues can be used to develop a measure of "line-structure-ness." In particular, since the first eigenvalue will correspond to the eigenvector in the direction of maximal variation (i.e. across the profile of a line object), then a high ratio of these eigenvalue magnitudes is a good indicator of an underlying line-structure. Further, a negative λ_1 indicates that underlying object is bright, while a positive values indicates a dark line. Finally, the direction of the line object should be well approximated by the eigenvector corresponding

4 At the scale being considered.

to the smaller eigenvalue.

Lorenz combined this information into the following scale-dependent line-structure detector for 2D images:

$$R_{2D}(\mathbf{x},\sigma)=\sigma^{\gamma}\cdot\lambda_1(\mathbf{x},\sigma) \tag{1.18}$$

where the scale, σ, is the standard deviation of a smoothing Gaussian kernel, $\gamma > 0$ and σ^{γ} compensates for the decrease in the response due to increased smoothing at higher scales.

Like the Koller method, this response is calculated at a series of scales and the maximum response across all scales is retained at **x**. Again, non-maximum suppression of the set of extreme responses generates a set of skeletons centred along the image's line-objects. As in the Koller results, these skeletons encode scale, contrast and local direction information and therefore represent the final segmentation. However, unlike the Koller method, the eigenvalue formulation is a much more robust response since it uses more global information and thus depends less on small imperfections in directional information.

1.3.5 Segmentation: Next steps

Line-structure segmentation concepts appear to be a very promising segmentation tool for endoscope images. They are based on general concepts of target object shape which should be quite robust when applied to minimally invasive surgery instrumentation. They offer an automatic interpretation of the segmentation result: clearly if the target objects are truly line-like in their structure, then their positions in an image should correspond directly to those image areas with high line-structure responses. Finally, if the approximate segmentation offered by the line-object detection procedure is too crude, then the initial estimate of shape and position should be readily integrable into a segmentation refinement scheme based on technologies like geodesic active contours. In fact, this exact principle has been applied with success in recent work to quantify 3D vascular structures in MRA volumes [Chen et al. 2004].

Despite all the expected advantages of line-structure-based shape models for the segmentation tasks, they cannot be immediately applied to endoscope images. There are numerous reasons, one of the most problematic being the difficult background conditions

of endoscope images. These problems, as well as a robust solution rooted in Bayesian regularization concepts, will be discussed in detail in the next chapter. Before moving onto that development, the discussion will overview some of the state-of-the-art tracking methods, with particular attention paid to a Bayesian-based optical flow method.

1.4 Object tracking

The segmentation aspect of the object-identification task is important towards the success of the automatic parameter calibration project since it will help isolate which objects in the image should be considered for the purpose of correlating image-based and global instrument movements. Once these objects have been identified, the critical task of tracking them begins. Preliminary investigations into this topic are initiated in this section.

1.4.1 Transformation-based trackers

Object tracking solutions in the computer vision literature can be cataloged into three general classes [Yilmaz and Shaw, 2004]: transformation-based, correspondence-based, and contour-based. In transformation-based approaches, the strategy is to transform a planar object region from one frame to the next; common transformations include translation, translation with scaling, and affine mapping. There are also many approaches for determining the appropriate transformation, such as template-matching or mean-shift tracking [Comaniciu et al. 2000]. Tracking via this strategy runs into several problems though, none-the-least-of-which is computational expense (especially for template matching), and strong limitations imposed by the assumed motion model. This could be a particularly problematic issue for endoscope video sequences, since even affine motion models may be too restrictive to adequately represent the complex relative movements between the instruments and endoscope.

Moreover, even if frame-to-frame motions are small or relative motions are stable, occlusions, blurring and highly variable contrast can create strong variations in apparent object shape. In general, the transformations will be only be approximate, the degree of the approximation depending heavily on the nature of the change in the visible object

shape between frames. As a result, the motion estimates produced by a transformation-based tracking scheme are likely to be too imprecise for the calibration task.

1.4.2 Correspondence-based trackers

As opposed to transform-trackers, which identify the target object once and then attempt to use motion estimates to follow the object into the next frame, correspondence-based trackers rely on object-detection in each frame and then seek to establish the appropriate correspondences between frames. The complexity of the process is mitigated by using the objects' states in the current frame to constrain the set of possible correspondences with the next frame. A state vector might include features such as texture and colour, or the position and/or motion of the object in frame I_k. In the latter case, the position and/or motion model can be applied to predict where in frame I_{k+1} the object will exist; actual object detection in I_{k+1} either supports or contradicts this prediction.

Any kind of object can be used for these methods, and corner features are particularly interesting for the current application. Corner features are tied to very precise object locations and are generally quite stable, making them overall predictable and a good input to this class of tracker. The precision and stability of corners also suggests that they might offer good features for the requirements of the calibration task. Smooth instrument surfaces offer only a few natural features such as those at joints, but these are often cluttered with noise and blood; a feature type like corners[5] that captures well the few existing real features, and represents a high structural level that will be robust to noise removal and other processing, is likely to be the most useful. By extension then, correspondence-based tracking with corner features appears highly suited to the calibration system and is expected to be the most feasible paradigm for development.

Kalman filters

Kalman filtering is a classic framework for correspondence tracking (see for example [Broida et al., 1990]) that is built upon the rigor and flexibility of the Bayesian approach to regularization. For tracking applications, the fundamental Kalman filter scheme begins

5 Note that this expectation must still be investigated via rigorous research into actual feature detection schemes.

with an initial estimate[6] $\hat{z}_k^+$ of the state of a set of measurements (in this case, positions of feature points). This estimate is passed through a model of the system dynamics Φ_k (in this case, a motion model) at the current time to produce a revised estimate $\hat{z}_{k+1}^-$ for the state of the measurements at the next time step, k+1. The final estimate of the state of the measurements $\hat{z}_{k+1}^+$ at time k+1 is then based on the revised estimate, as well and the actual data from the system u_{k+1} compared to expected data $H\hat{z}_{k+1}^-$, with H a model for the measurement process (in this case, the camera). The process is summarized in the following recursive set of equations:

$$\begin{aligned}
\hat{z}_{k+1}^- &= \Phi_k \hat{z}_k^+ \\
P_{k+1}^- &= \Phi_k P_k^+ \Phi_k^T + Q \\
\hat{z}_{k+1}^+ &= \hat{z}_{k+1}^- + K_{k+1}\left[u_{k+1} - H\hat{z}_{k+1}^-\right] \\
P_{k+1}^+ &= \left[I - K_{k+1}H\right]P_{k+1}^-
\end{aligned} \tag{1.19}$$

The essential element of the Kalman filtering process is the computation of the Kalman gain matrix K_{k+1}, which controls how much influence the data have on the final estimate of the state of the measurements $\hat{z}_{k+1}^+$. In other words, it balances how much consideration is given to actual measurements compared to prior expectations/models. This is exactly the standard regularization idea, but the particular power of the Kalman method comes from the fact that the Kalman gain is determined from the statistical nature of the system at the current time rather than fixed weighting terms. More specifically, the term P_k represents the aggregate error covariance of the system, including both measurement noise covariance R (not shown in above equations), and state extrapolation noise covariance Q. When the uncertainty in the prior model for the system (as measured by the covariance) is high but the measurement uncertainty is low, then $\hat{z}_{k+1}^+$ will be dominated by the data. In contrast, if the uncertainty in the data is high, perhaps due to noise, then the prior term will take precedence, helping maintain a stable solution.

Unfortunately, a significant difficulty experienced with basic Kalman filtering is the need for both a linear measurement process H, and predictable system dynamics, Φ.

---

6 The estimate can be simply a measurement, or a more complex output of an initialization process.

These problems have been addressed through extensions such as the well-known Extended Kalman Filter (EKF), which first linearizes the system at each time step through a first-order Taylor expansion before applying the recursive filtering. This method can lead to large approximation errors though, and alternate methods that are accurate to higher-order have been developed, including the Unscented Kalman Filter (UKF), [Wan and van der Merwe, 2000]. In this approach, a carefully chosen set of control features are first passed through the non-linear system to get better descriptions of the underlying system statistics and behaviour. While advanced techniques like the EKF and UKF may hold promise for the complexities of the endoscope tracking problem, the system is currently at a too-early stage of development to pursue this topic rigorously. The principle difficulty is the need for an investigation into what type of feature-points can be reliably extracted from endoscope images. From previous discussions, corner features certainly seem to hold a lot of promise, but many uncertainties exist. Most importantly, which of the innumerable corner detectors will be applied, and will it be able to return a sufficient set of quality corner points? Since the success of a feature-tracker will rely fundamentally on the quality and stability of the extracted features, until these questions have been addressed, the tracking problem remains under-constrained.

1.4.3 Contour-based trackers

Even if the feature-detection problem limits the depth with which a correspondence-based tracking scheme can be developed at this time, it is interesting to consider how the other classes of trackers might help a correspondence approach. In particular, approximate object motion estimates between frames might serve to constrain paradigms such as Kalman filtering. If a good motion estimate could be developed, the need for more complicated Kalman approaches such as the UKF might be avoided; at the very least, motion estimates may add another degree of constraint when verifying predicted feature locations with measured data, helping increase the stability of a Kalman filter if there are few real features and/or numerous false-features arising from reflections, blood, etc. The class of contour-based trackers should be particularly useful for this task.

A particular advantage of contour-based trackers for this purpose compared to

transform-based methods is their natural integration into the segmentation framework discussed earlier. This class of tracker is a straight-forward extension of contour/boundary evolution segmentation schemes: the object boundary in frame I_{k+1} is represented by a contour that has been evolved from an initial location, with the initialization in this case the position of the object's boundary in frame I_k rather than a more arbitrary position as in pure segmentation.

The mechanisms for contour evolution discussed previously still apply, with boundary-based or region-based measures used to continue of halt contour evolution. For the tracking application here, region-based energy measures are of more interest, since they offer more global motion information. More specifically, if the goal is to extract an approximation of a (mostly) rigid object's movement, then aggregate motion information will be easier to obtain from the set of all region pixels rather than trying to correlate motion information from boundary contours only, whose shape will be susceptible to occlusions and other effects—in short, region measures should provide an overall more stable motion estimate.

These measures may be integrated into segmentation-based or motion-based energy functionals for contour evolution. The majority of motion-based functionals are based on the concept of optical flow, wherein a field of motion vectors is computed using two sequential images. Contour evolution between the frames is constrained by measures of motion vector consistency in image regions. This will be explained in more detail next.

Optical flow – Tikhonov formulation

Optical flow extraction of the motion fields corresponding to a moving object are based on the assumption that the intensity of an object remains constant over time [Horn and Shunk, 1981]: $I(x_1, y_1, t_1) = I(x_2, y_2, t_2)$. This is known as the "convected irradiance" constraint, and through a Taylor series expansion can be represented as:

$$\nabla I \cdot \vec{v} + \frac{\partial I}{\partial t} = I_x u + I_y v + I_t = 0 \tag{1.20}$$

The objective of the optical flow formulation is to solve for the velocity field (u, v) that satisfies the above equation. The result is a motion estimate at every point in the

image that can be used to help isolate target objects by searching for regions with consistent velocity information. Yet (1.20) alone cannot be used to solve for the motion field, since the problem is an ill-posed one: (1.20) provides a set of N equations, but there are $2N$ unknowns (i.e. u and v, the x and y components of the velocity field), such that there are an infinite number of solutions. To deal with the ill-posedness of the problem, Horn and Shunk developed a regularized energy functional by using a Tikhonov stabilizer; the velocity field was then the one that minimized the following motion-based functional:

$$\int \left(I_x u + I_y v + I_t\right)^2 + \lambda\left[\left(u_x^2 + u_y^2\right) + \left(v_x^2 + v_y^2\right)\right] dx\, dy \tag{1.21}$$

This functional imposes an assumption of smooth velocity fields, and falls from the idea that points on an object should move with near the same velocity between two image frames. Any large discontinuities in the gradient of u or v will be penalized strongly, so that solutions which minimize the energy given by (1.21) will be those with near constant intensity and very smooth velocity fields. The importance of the smoothness constraint compared to the observed data (i.e. the convected irradiance term) is controlled as normal via the Lagrange multiplier λ, and this value plays a very critical role. Specifically, the more influence one gives to the smoothing term, the more that real discontinuities in the velocity field will be suppressed. These discontinuities always occur at the occluding boundaries of objects, and smoothing therefore risks loosing precise boundary information. For the tracking objective under consideration here, which is only to obtain an approximate motion estimate, this is not a serious flaw. Still, solutions that reduce this effect are available via a Bayesian formulation for the optical flow problem.

Optical flow – Bayesian formulation

Unlike motion-based energy functionals such as (1.21), segmentation-based functionals use earlier classifications of background and region when controlling contour evolution between frames. Mansoui [2002] proposed a method to bridge the gap between the two frameworks, formulating a method using the level set framework that embedded the convected irradiance constraint of the optical flow scheme into a more robust Bayesian

expression which was then used as the level set propagation speed, F. While retaining the assumption of stable intensity he eliminated the need to calculate the velocity field (u,v) by integrating previous region and background classifications, which helped avoid the problem of blurring at motion boundaries.

Mansouri's method begins from a MAP probabilistic estimate of the target object region R_1 in the current image, where the target object's position in the previous image is known and represented by R_0:

$$\begin{aligned}\hat{R}_1 &= \underset{R}{\operatorname{argmax}}\, P\left(R_1 = R | I^n, I^{n+1}, R_0\right) \\ &= \underset{R}{\operatorname{argmax}}\, P\left(I^{n+1} | I^n, R_0, R_1 = R\right) P\left(R_1 = R | I^n, R_0\right)\end{aligned} \tag{1.22}$$

The second expression follows from Bayes' rule.

Via some simplifying assumptions, the maximization scheme of (1.22) can be formulated as a minimization one using negative log likelihoods:

$$\begin{aligned}\hat{R}_1 = \underset{R}{\operatorname{argmin}} & -\int_R \log P_{\text{in},\mathrm{x}}\left(I^{n+1}(\mathrm{x}) | I^n, R_0\right) dx \\ & -\int_{R'} \log P_{\text{out},\mathrm{x}}\left(I^{n+1}(\mathrm{x}) | I^n, R_0\right) dx \\ & +\lambda_L \oint_{\partial R} d\rho\end{aligned} \tag{1.23}$$

In this expression, the first term represents the probability of a pixel being a region pixel in the current frame, given its intensity and classification in the previous image. The second term represents the probability that the same pixel is a background pixel. The third term falls from the prior term of (1.22), but the conditional dependencies on the previous image and region have been removed to be as general as possible. The resulting term in (1.23) acts as a regularizer to enforce a degree of smoothness in the evolving contour. Therefore, this functional is very similar in form to a Tikhonov regularization scheme except that the data consistency measures, as represented by the first two terms of (1.23), are motivated strongly by Bayesian regularization concepts.

To better understand the probabilistic underpinnings of the method, as well as its operation in general, consider the following expressions for the first two terms of (1.23):

$$-\log P_{\text{in},\mathrm{x}}\left(I^{n+1}(\mathrm{x}) | I^n, R_0\right) \approx \inf_{\mathrm{z}:\|\mathrm{z}\| \le \delta, \mathrm{x}+\mathrm{z} \in R_0} \left(I^{n+1}(\mathrm{x}) - I^n(\mathrm{x}+\mathrm{z})\right)^2 \tag{1.24}$$

and

$$-\log P_{\text{out},\mathrm{x}}\left(I^{n+1}(\mathrm{x})|I^{n}, R_0\right) \approx \inf_{\mathrm{z}:\|\mathrm{z}\|\le\delta,\,\mathrm{x}+\mathrm{z}\in R_0^c} \left(I^{n+1}(\mathrm{x})-I^{n}(\mathrm{x}+\mathrm{z})\right)^2. \tag{1.25}$$

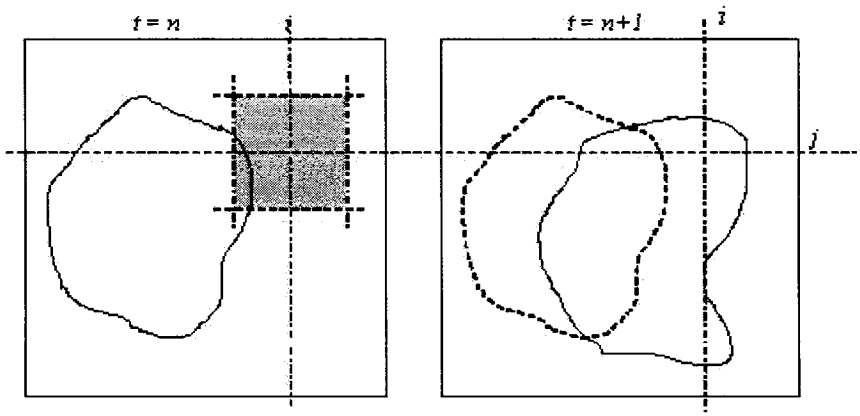


Figure 1.7: Probabilistic optical flow scheme. The intensity of the pixel* x *(denoted (i,j) in the figure) in the frame n+1 is compared to the intensities of all pixels in the neighbourhood of* x *in the previous frame; differences with pixels classified as region and background are recorded by (1.24) and (1.25) respectively.

In (1.24) the intensity of a pixel $I^{n+1}(\mathrm{x})$ in the current frame is compared to the intensities of the pixels classified as region in the neighbourhood of x in the previous frame, with the size of the neighbourhood used for comparison given by δ. Via (1.25), the same intensity comparison is made against all pixels in the neighbourhood classified as background. These comparisons are represented visually in figure 1.7.

The minimum differences produced by the region and background comparisons are recorded in the variables $V_{\mathrm{x}}^{\mathrm{in}}$ and $V_{\mathrm{x}}^{\mathrm{out}}$ respectively. In the case where there are no region or background pixels in the neighbourhood in the previous image, the appropriate variable is set to a high positive constant. Therefore, the difference $V_{\mathrm{x}}^{\mathrm{in}} - V_{\mathrm{x}}^{\mathrm{out}}$ represents the likelihood of a pixel in the current frame being region or background based on previous classifications and current data: a high positive value encodes a strong probability that the pixel is background, and vice-versa.

Armed with this probabilistic measure, Mansouri integrates (1.23), (1.24) and (1.25) into the basic level set scheme given in (1.10) to produce the following PDE for contour

evolution:

$$\frac{\partial \varphi(\mathrm{x})}{\partial t} = -\left[V_{\mathrm{x}}^{\mathrm{in}} - V_{\mathrm{x}}^{\mathrm{out}} + \lambda_L \kappa_u(\mathrm{x}) \right] \| \nabla \varphi \| \tag{1.26}$$

The general numeric implementation of the level set scheme is described in Sethian [1999], and results in the following, final form for Mansouri's tracking method:

$$\varphi_{i,j}^{n+1} = \varphi_{i,j}^{n} + \Delta t \left(-\left[\max(F_{i,j}, 0) \nabla^{+} + \min(F_{i,j}, 0) \nabla^{-} \right] + \epsilon K_{i,j}^{n} \left(\left(D_{i,j}^{0x} \right)^2 + \left(D_{i,j}^{0y} \right)^2 \right)^{1/2} \right) \tag{1.27}$$

where the designation of a pixel has been substituted with *i,j* for notational clarity. In this expression $F_{i,j} = V_{\mathrm{x}}^{\mathrm{in}} - V_{\mathrm{x}}^{\mathrm{out}}$ is as described previously. In the parlance of Sethian's text, the ∇ terms are designed to ensure the correct "upwinding scheme" is used for contour evolution. Simply put, these terms force the evolution process to use forward or backward differences when appropriate, and take the following definitions:

$$\nabla^{+} = \left[\max(D_{ij}^{-x}, 0) + \min(D_{ij}^{+x}, 0) + \max(D_{ij}^{-y}, 0) + \min(D_{ij}^{+y}, 0) \right]^{1/2}$$
$$\nabla^{-} = \left[\max(D_{ij}^{+x}, 0) + \min(D_{ij}^{-x}, 0) + \max(D_{ij}^{+y}, 0) + \min(D_{ij}^{-y}, 0) \right]^{1/2}$$

The various *D* terms throughout these two equations as well as in (1.27) are simple numerical differences:

$$D_{ij}^{-x} = \varphi_{i,j}^{n} - \varphi_{i-1,j}^{n}$$
$$D_{ij}^{+x} = \varphi_{i+1,j}^{n} - \varphi_{i,j}^{n}$$
$$D_{ij}^{-y} = \varphi_{i,j}^{n} - \varphi_{i,j-1}^{n}$$
$$D_{ij}^{+y} = \varphi_{i,j+1}^{n} - \varphi_{i,j}^{n}$$
$$D_{ij}^{0x} = \left(\varphi_{i+1,j}^{n} - \varphi_{i-1,j}^{n} \right) / 2$$
$$D_{ij}^{0y} = \left(\varphi_{i,j+1}^{n} - \varphi_{i,j-1}^{n} \right) / 2$$

The κ term is the central difference approximation to the contour curvature, and takes the form:

$$\kappa_{i,j} = \frac{\varphi_{xx} \varphi_y^2 - 2 \varphi_y \varphi_x \varphi_{xy} + \varphi_{yy} \varphi_x^2}{\left(\varphi_x^2 + \varphi_y^2 \right)^{3/2}}$$

with the subscripts indicating partial numerical derivatives:

$$\varphi_x = \left(\varphi^n_{i+1,j} - \varphi^n_{i-1,j}\right)/2$$
$$\varphi_y = \left(\varphi^n_{i,j+1} - \varphi^n_{i,j-1}\right)/2$$
$$\varphi_{xx} = \left(\varphi^n_{i+1,j} - 2\varphi^n_{i,j} + \varphi^n_{i-1,j}\right)$$
$$\varphi_{xy} = \left(\varphi^n_{i+1,j+1} - \varphi^n_{i+1,j-1} - \varphi^n_{i-1,j+1} + \varphi^n_{i-1,j-1}\right)/4$$
$$\varphi_{yy} = \left(\varphi^n_{i,j+1} - 2\varphi^n_{i,j} + \varphi^n_{i,j-1}\right)$$

Recall from the discussion of section 1.3.1 that the level set evolution scheme operates by using the speed function F to move up or down the higher-dimensional surface φ. If the initial contour is embedded as the zero level set of the surface, and the the surface is initialized to the signed distance function, then region points are those on the surface with a negative distance to the zero level set, and background points are those with a positive distance. In Mansouri's scheme, when a previous background pixel becomes part of the object area in the current frame $n+1$ due to object movement, intensity comparisons with the previous frame should generate a negative speed value F. As (1.27) is applied, the signed distance to the point from the previous zero level set will become progressively smaller. This is equivalent to the zero level set moving outward to englobe the point to follow the movement of the object.

Since the level set evolution applies to the entire surface φ, with the boundary simply embedded as the zero level set, all the region pixels of the target object are automatically moved between frames. Therefore, a global measure of motion should be readily obtainable. This is not as convenient as the original optical flow scheme, which produces a motion estimate directly via the velocity field (u,v); however, there are numerous reasons why Mansouri's approach is preferable. For one, the Bayesian formulation is more robust, relying on probabilistic constraints that are more adaptable to local image variations like real motion discontinuities. Further, the level set implementation is very stable, which will be advantageous for the difficult endoscope imaging environment. Moreover, the use of motion estimates to serve as additional constraint for a correspondence-based tracker is only speculative. The probabilistic approach embeds the adaptability of Bayesian methods without imposing the need to compute motion models, and this should offer more flexibility for development in the future – motion information can be extracted when, and *if*, needed.

Despite these advantages, the Mansouri scheme is still intensity-based. From earlier discussions, this might reasonably be expected to be a significant short-coming of the method for tracking across endoscope images. The scope of this problem, as well as a potential solution, will be addressed in the next chapter.

CHAPTER 2 - METHODOLOGY

In the preceding Literature Review state-of-the-art computer vision technologies for target identification and tracking were discussed. This chapter extends these ideas, giving particular focus to modifications and advancements of line-object technologies for segmentation. This is a critical system component: to re-iterate earlier observations, if image regions *strictly* corresponding to surgical instruments cannot be reliably identified it will be exceedingly difficult to differentiate moving instrument objects from the motion of the background. The resulting ambiguity will cripple the accuracy of 3D motion measurements, and by extension, undermine the reliability of the endoscope calibrations.

The importance of the segmentation step does not diminish that of the object tracking task. Without question, even if instrument objects are extracted, these must be accurately tracked if the calibration objective is to be achieved. Therefore, preliminary tests and developments of an instrument tracking algorithm will also be considered.

The discussion of the segmentation will be broken down into two pieces: line-object extraction and line-object selection. The first will focus on the actual detection of line-objects, beginning with the limitations of current methods followed by the proposed Canny-edge detection approach. Then, the shortcomings of an assumption of existing schemes—that all good line responses in an image arise from target objects—will be addressed by developing the Bayesian scheme used to select which line-objects in the image correspond to instruments. Following these developments, specifics of the tracking algorithm will conclude the chapter. To connect the development to the results of the next chapter, each section will also outline the methods that will be used to evaluate the algorithms. The structure of the chapter is summarized in the following figure:

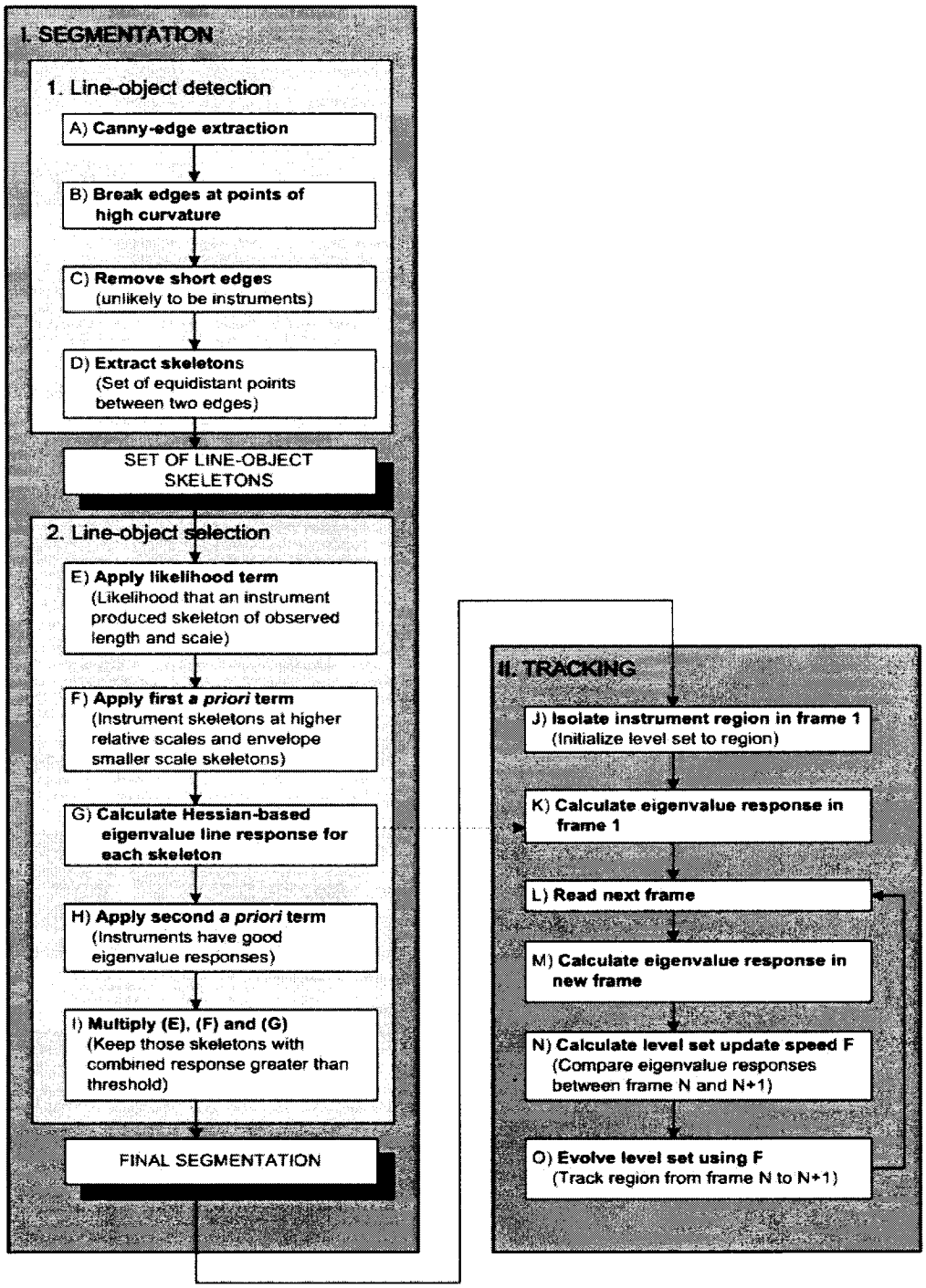


Figure 2.1: Structure of proposed segmentation and tracking methods.

2.1 Segmentation: Edge-based line-object detection

A robust and broadly applicable method to strictly segments surgical instruments should rely on identification clues that are as consistent as possible between different endoscopy procedures. As discussed previously, a particularly consistent instrument feature is a generally long and narrow shape. Current line-object methods are not immediately applicable to endoscope images because they implicitly assume that local across-scale extrema of detection operators such as (1.18), $R(\mathbf{x},\sigma_d)=\sigma_d^{\gamma}\cdot\lambda_1(\mathbf{x},\sigma_d)$, always correspond to target objects; they do not consider the possibility that extrema might also arise from non-target objects or textures. Two problems result: the first concerns how line-objects are actually detected in endoscope images. The second problem is determining which of the detected line-objects correspond to instruments. In this section, the former problem is considered.

2.1.1 Limitations of current line-object detectors

Line-object detection schemes like that proposed by Lorenz et al. [1997] have worked well for problems like vascular-tree extraction from 2D images and 3D volumes, where the objects of interest have good contrast against relatively homogeneous backgrounds. Detection is complicated in endoscope images for several reasons.

Potential to extract inaccurate line-objects

Recall from the literature review that the line-object detection schemes discussed operated by calculating a line-response function $L(\mathbf{x},\sigma_d)$ at each image point $\mathbf{x}$, for each of a set of scales. After all scales were processed, each $\mathbf{x}$ was assigned the extremum response across scales: $L_{\mathbf{x}} = \max\{\ |L_{\mathbf{x}}(\sigma_d)|,\ 0 < \sigma_d < \sigma_{\max}\ \}$. Finally, non-maximum suppression of the responses in the neighbourhood of each $\mathbf{x}$ yielded chains of pixels corresponding to line-object skeletons. These chains fully characterized the segmentations of their corresponding line-objects: not only did each point along the chain correspond to the position $\mathbf{x}_{\text{ext}}$ of the locally extreme response, but each point also encoded the detector scale σ_d at which the response occurred, as well as the orientation of the eigenvector at that scale.

Figure 2.2 demonstrates this behaviour for an ideal Gaussian-profiled line-object against a homogeneous background for the detector $L(\mathbf{x},\sigma_d)=R(\mathbf{x},\sigma_d)=\sigma_d^{\gamma}\cdot\lambda_1(\mathbf{x},\sigma_d)$. The plot in (a) shows the line profile in the direction given by the eigenvector corresponding to the strongest eigenvalue at the point $\mathbf{x} = 0$. In (b), the response in this direction has been calculated for three different detector scales: $\sigma_d = 1, 4, 7$. As expected, all scales have produced extrema at $\mathbf{x} = 0$, and the across-scale extremum is clearly given by $\sigma_d = 7$, which corresponds to the real (full-width half-maximum) scale of the line-object. Therefore, non-maximum suppression of this output across the different scales will correctly identify the underlying line-object.

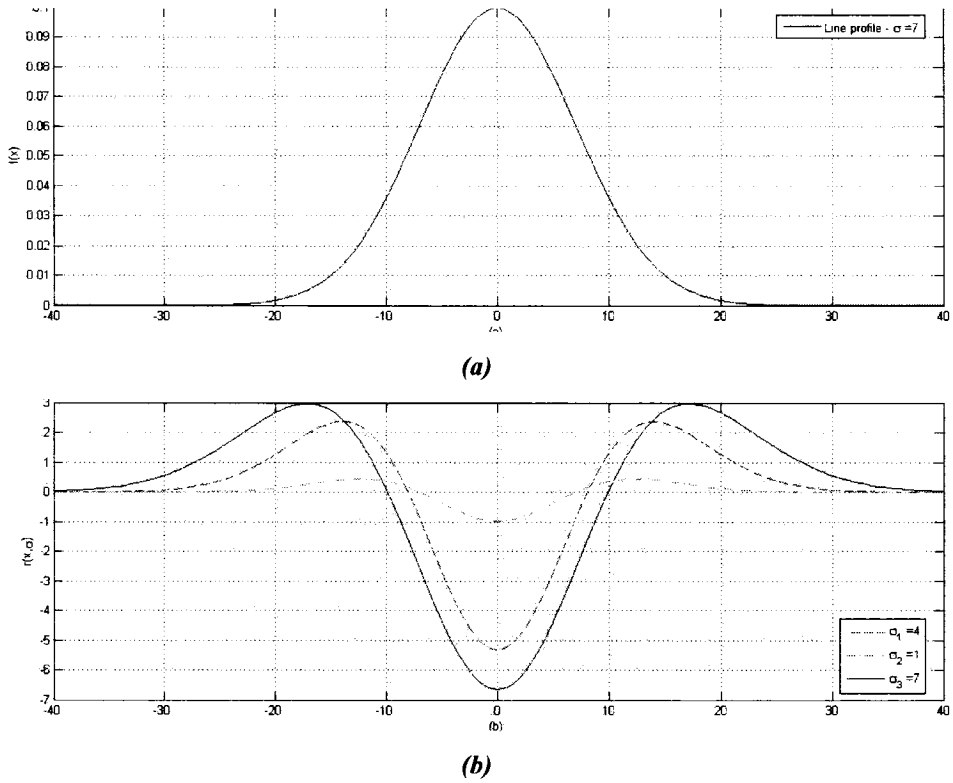


Figure 2.2: Response of the Hessian-based detector to an ideal line-object. (a) Cross-section of an ideal Gaussian-profiled line-object. The scale, $\sigma_l = 7$, of the line-object is given by the full-width-half-maximum point. (b) The responses given by (1.18) for three different detector scales: $\sigma_d = 1$, 4 and 7. The response for $\sigma_d = 7$ generates an across-scale extremum, as expected.

The problem for endoscope images is that coincidental texture in the background, or unexpected intensity profiles across the instrument surface arising from environmental

conditions such as blood smearing, can cause an across-scale extremum of (1.18)[7] at a different scale than the true instrument. This situation is illustrated in figure 2.3:

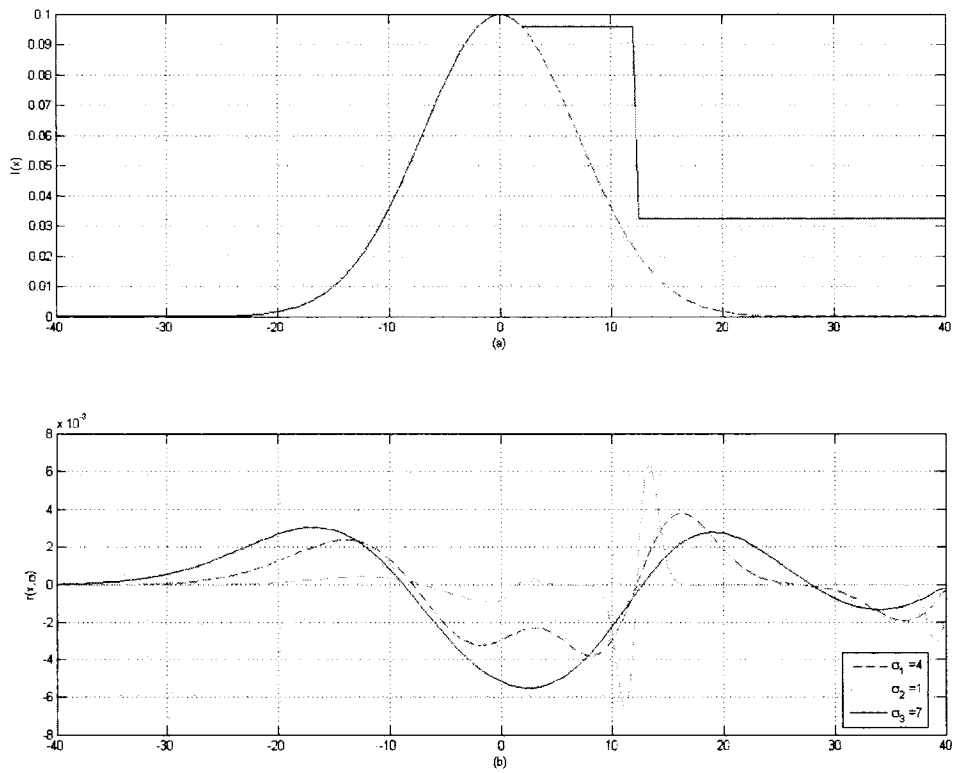


Figure 2.3: Response of the Hessian-based detector to a non-ideal line-object. (a) The profile of the same line as in figure 2.2(a), except that image conditions have washed-out the intensity variation to the right side, and background texture has created another edge. (b) Responses using the same detector scales as in figure 2.2(b). Non-maximum suppression applied to these responses risks incorrect segmentation.

The same profile from figure 2.2 has been plotted in (a), except that environmental conditions have washed-out the intensity variation across the right portion of the profile. Background tissue has simultaneously caused the appearance of another edge just past the scale of the ideal line profile. In (b), the response (1.18) is shown for each of the three scales considered earlier. If the across-scale non-maximum suppression scheme is used

7 Note that for the sake of clarity, the discussion here will focus on this specific detection scheme. However, the vast majority of the discussion is equally applicable to other detection schemes. Specific mention will be made when this is not the case

with the results in (b), one of two non-ideal results might occur. In the best case, the extremum of the $\sigma_d = 7$ response will be found, in which case the skeleton will embed the correct instrument scale, but will not follow exactly its true shape since the extremum does not occur at the true centre of the line-object profile. In this case, the resulting segmentation will likely englobe a small amount of background tissue to the right of the true instrument edge. In the worst case, the non-maximum suppression scheme might easily choose the peak of the $\sigma_d = 1$ response as the across-scale extremum in the neighbourhood of $\mathbf{x} = 0$; not only will the position $\mathbf{x}_{ext}$ encoded by the skeleton be far from the real instrument centre, but the embedded scale will also be wrong.

If problems like this were assumed to occur only over isolated, short regions, a robust detector could be built easily. However, the real endoscope image in figure 2.4 illustrates that this is a dangerous hypothesis. Consider the lower left prong of the instrument, which is enlarged in the image inset. At the extreme lower tip of the prong the instrument exhibits good contrast and the segmentation (and the associated skeleton) corresponds well with the actual shape. However, a problem quickly develops as the segmentation/skeleton follows the prong upwards. In particular, the contrast towards the right instrument edge falls, while at the same time a significant edge in the background tissue develops that is difficult to isolate because it begins near the instrument before gradually diverging. The response (1.18) to the intensity profile between this tissue and the left instrument edge is stronger than to the profile between the left and right instrument edges, and the non-maximum suppression scheme keeps only the response at the higher scale. Thus the skeleton is led astray of the true instrument shape and encodes too-high scale information, resulting in the observed over-segmentation.

Figure 2.4: Poor line-object segmentation due to coincidental texture and blurring. Inset – Detail of the lower left instrument prong with segmentation corresponding to the non-maximum suppression of the across-scale extrema given by (1.18). Because the real instrument contrast in this area is weak, the background tissue has led to an inaccurate skeleton and improper segmentation. Also, the upper-right instrument prong in the larger image is slightly over-segmented, due in part to Gaussian blurring applied during line-structure detection.

This problem is rooted in the assumption that the maximal line-responses in an image are always generated by target objects. Since object edges, contrast and intensity profiles in endoscope images can be weak or highly variable, and since backgrounds frequently give rise to difficult textures like that shown, this assumption makes line-object schemes as they exist inadequate for the current application.

Reduced response resolution at high-scales

Another detection problem arises from movements of the endoscope and instruments themselves which can cause dramatic changes in the apparent scale of an instrument. The resulting problem lies in the heavy dependence on Gaussian filtering of existing detection schemes: while blurring is necessary to appropriately determine eigenvectors or other direction indicators, it will also reduce the resolution of the non-maximum suppression

schemes with increasing scale. Consider for example the graph shown in figure 2.5.

The figure plots the response (1.18) to a line-object of scale σ_l for two different detector scales σ_d. The solid line represents the response when $\sigma_d = \sigma_l$, and the dashed line is the response when $\sigma_d = \sigma_l - 10$. At lower scales, there is a clear difference between the two responses, and a non-maximum suppression scheme will not have difficulty extracting the appropriate line-object scale. However, as the scale of the line-object increases, the difference between the two responses quickly diminishes due to the increasingly heavy blurring. Especially in the presence of weak boundaries and/or background textures this reduced response resolution at higher line-object scales may easily lead to a non-maximum suppressed extremum at a scale inconsistent with the true object.

Target objects usually existed at lower scales in previous line-object applications, but large changes in apparent scale arising from endoscope and instrument movements indicate the potential for seriously flawed segmentations in the current context. A real example of this problem is visible in figure 2.4 from before, where the upper portion of the right instrument prong is over-segmented. In this region, there is little background texture, which means that the principal reason for the over-segmentation is a lack of precision in the exact instrument scale due to the strong blurring at the relatively high scale of the prong.

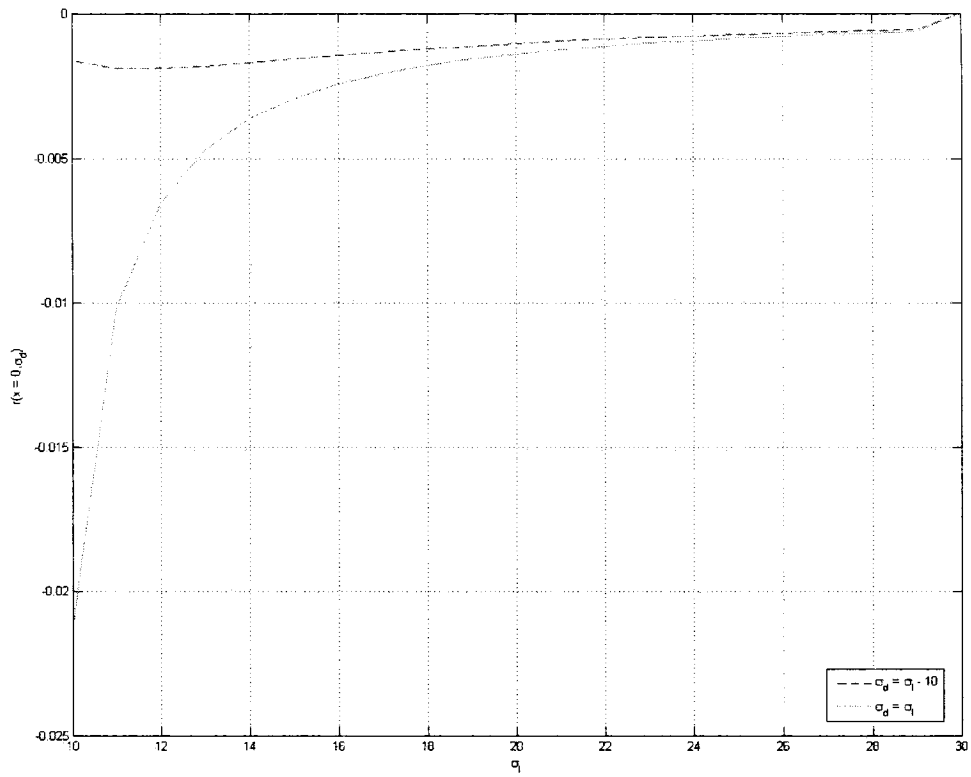


Figure 2.5: Relationship between Gaussian blurring and response resolution. The lower (solid) line represents the response of the detector in (1.18) tuned to the scale of of a line-structure (i.e. $\sigma_d = \sigma_l$) whereas the upper (dashed) line is set to σ_l - 10. At lower scales, there is good separation of the responses; however, as the scale of the line-object increases, Gaussian blurring causes a loss of resolution. While the relative scales of the response functions have not changed, the risk of an incorrect segmentation is higher, especially in the presence of non-ideal image conditions such as weak object contrast.

2.1.2 Modified line-object detection with Canny edge extraction

To motivate the proposed solution to these detection problems, consider first that current detection methods, in effect, assume that all image points are potential skeleton pixels. Thus, the detection procedure amounts to searching the Gaussian-blurred image for good responses at each point to support this initial hypothesis.

Now observe that, in general, an object with an intensity profile that generates a good response to a function such as (1.18) also generates good edge content. Evidence of this idea is presented in figure 2.6, where a Gaussian profile (σ = 10) is shown at left, and the

edge content found by a Canny detector [Canny, 1986] at right (normalized low and high thresholds of 0.2313 and 0.5781, respectively). The detected edges correspond exactly to the scale of the profile.

Therefore, the segmenting boundaries derived from line-object responses like (1.18) (with the edges just the extracted skeletons offset by the skeleton scales) can reasonably be expected to be amongst the set of image edges directly extracted by a tool such as the Canny detector. In short, a Canny-detected set of edges will contain the same, or more, edge information that can be extracted via the eigenvalue-based line-detector.

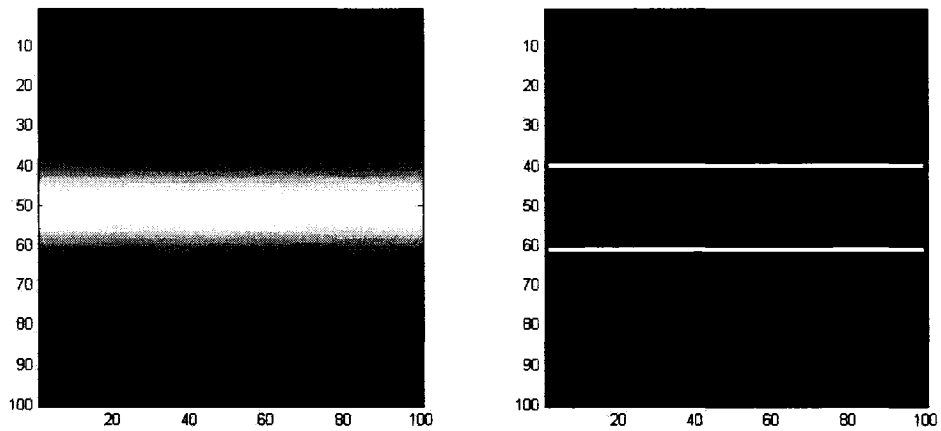


Figure 2.6: Gaussian line profile and Canny-detected edges. The edge positions in the right image correspond exactly to the scale of the Gaussian profile in the left image. Therefore, Canny-edge detection should return at least the same edge content that would be available by a standard line-object detection scheme.

Line-object edge extraction

These observations prompt a new line-object detection strategy: instead of beginning with an arbitrary pixel and then asking if there is support for the line-object hypothesis, treat the image edge-content as support for a line-objects, and use the presence or absence of edges to isolate immediately which pixels might be real skeleton points. By using edge content in this way, the Gaussian blurring is avoided, and the precision of the ultimate line-object boundaries will not depend on the scale of the line-object detector. From the perspective of the system accuracy need identified in the Introduction, this is a good design choice.

Obviously though the scheme needs to be refined, and the glaring questions include not only how to extract the set of edges that might correspond to line-objects, but also how to use this edge set to determine which pixels are skeleton points. As suggested by the previous discussion, the Canny edge detector has been used to address part of the first problem.

Canny edge detection

For this application, the hysteresis thresholding integrated into the Canny detector is particularly useful. In the first place, there are normally some areas along an instrument's length with sharp, well-defined boundaries. This implies that a target instrument should have edge content somewhere along its length that is amongst the best edge content at a particular scale. Indeed, for this not to be the case is a contraindication for the task of segmentation, since the instrument is probably not well imaged. The upper limit of the hysteresis thresholding filter can enforce this assumption by retaining only those edges containing at least some near-optimal edge content.

While the high limit can help prune the set of extracted edges, the low threshold works to ensure that the remaining ones represent as much of an instrument as possible, even under strong fluctuations in the boundary contrast. The choices for the high and low thresholds (τ_h and τ_l respectively), as well as the standard deviation of the Gaussian filtering, σ_{canny} will be discussed in the results section.

Edge removal

After extracting a set of edges in the image, the next task is to use that information to actually determine where in the image skeletons corresponding to line-objects exist. The solution to this problem revolves centrally around the parallelism expected between the occluding boundaries of line-objects, and proceeds in two stages: edge removal followed by final skeleton extraction.

Despite expectations that instruments should contain at least some good edge content, it is dangerous to rely too heavily on the thresholds of the Canny detector. The high limit is a hard threshold and, under variable illumination, a broadly applicable definition for "good edge content" is hard to define. To make the line-object detection scheme as

transferable as possible between images, the hysteresis thresholds of the Canny detector should be set lower rather than higher. This will invariably result in too many edges detected; however, many can be discarded by applying general assumptions about the problem at hand. The goal of this approach is to avoid rejecting edges by relying too heavily on intensity thresholds, instead using more stable knowledge of the problem to retain the most informative edges.

Curvature and minimum length

The shape and minimum visible length of instruments are two relatively stable characteristics. At an extremely general level, most instruments are straight – they may have a gradually curved structure, but are unlikely to have sharp changes in shape except at joints (see the set of images in figure 2.10 of section 2.2.4 for illustrations). This is just a limiting case of the general curvilinear structure expected of most line-structures. Furthermore, while the low thresholds of the Canny detector will produce edges corresponding to instrument edges and arbitrary connections between background textures, the background edges are less likely to be straight. Therefore, if image edges are broken into straight segments, the overall shape and length of instrument edges should be better preserved than background edges. It follows immediately that the resulting set of straight edges can be thresholded according to a minimum length criterion appropriate for expected instrument visibility. This should be a more constant characteristic than image intensity, such that the set of edges for skeleton detection will be reduced using a more predictable and stable criterion than Canny thresholds alone.

Actual breaking of edges into straight segments is performed by first parameterizing all edges during the Canny detection according to $\vec{e}(s)=(x(s), y(s))$. Then, the curvature at each point along the extracted edge is calculated as:

$$k(t)=\frac{\dot{x}(t)\ddot{y}(t)-\dot{y}(t)\ddot{x}(t)}{\left(\dot{x}(t)^2+\dot{y}(t)^2\right)^{3/2}}. \tag{2.1}$$

Edges are broken at points where this curvature exceeded a certain threshold. Values for this break threshold τ_k, and the minimum length threshold τ_e applied to the remaining edges, will be discussed in the results.

Skeleton identification

With the final set of extracted edges, the task now is to determine which pairs actually border real instrument regions, so that the skeleton for the line-object can be extracted as the set of equidistant points between them. The process is complicated by the inevitable presence of background edges (despite the last step), the possibility of more than one instrument, and the potential for instruments with jointed parts. As an example, consider a pronged instrument such as the one shown previously in figure 2.4. In the ideal case, four edges representing the occluding boundaries will be detected. Although only two pairs of these edges will bound instrument regions, there are six possible pairings to consider, each of which will generate a set of equidistant points. If there are edges in the texture between the prongs, or within the prongs themselves (discontinuities between facets, for instance), the set of possible pairings will grow even more. It is clear that the extraction process requires additional constraint, which can once again be found by resorting to expected instrument shape.

Parallelism

To begin, if the characterization of instruments as line-objects is valid, then occluding boundaries of the same instrument segment must exhibit a high degree of parallelism. Consequently, only near-parallel edge pairs need be considered further. Herein lies a benefit of the edge-breaking scheme from the edge-removal stage: straighter edges mean more accurate orientations, which improves the accuracy of the parallelism evaluation. Restricting the search for skeletons to parallel edges will help overcome difficulties such as the multiple lines of the instrument prongs just mentioned. The two true boundaries of each prong should always be parallel to one another, but all four boundaries will only parallel when the prongs are closed, in which case it is not a problem if the maximally offset edges are paired.

Maximum separation

Yet with parallelism as the only constraint, two parallel but far-apart edges (e.g. at opposite image edges) will still be considered, and it is unlikely that instruments will ever appear that wide. Accordingly, edge pairs that are more than a threshold distance apart

are also eliminated from further consideration.

Skeletons might now be extracted as the set of equidistant points between each proximal and parallel edge pairing, except that one final issue must be addressed. The distance from an edge is defined in all directions, and the set of equidistant points between two parallel edges has an infinite extent in the parallel direction. However, the two parallel edges themselves will have finite lengths, either because an instrument has ended, an edge has changed direction, or the contrast has fallen below the low Canny threshold. To avoid over-segmenting, it is important to constrain the parallel extent of an extracted skeleton to be consistent with the extents of the corresponding edges.

The final set of extracted skeletons has now been achieved. The overall process of skeleton identification can be thought of as first building a distance map around each edge remaining after the initial edge removal stage. Then, the normal direction **n** of each edge is determined. Given the maximum expected image size of an instrument, W_{max}, the normals are projected through the endpoints of the edges to distances of $\pm W_{max}$, defining top right (tr), top left (tl), bottom right (br) and bottom left (bl) points. These points define an envelope around each edge, and skeletons are extracted as the equidistant points falling within the overlap of the envelopes of two near-parallel edges. This process is depicted graphically in the following figure:

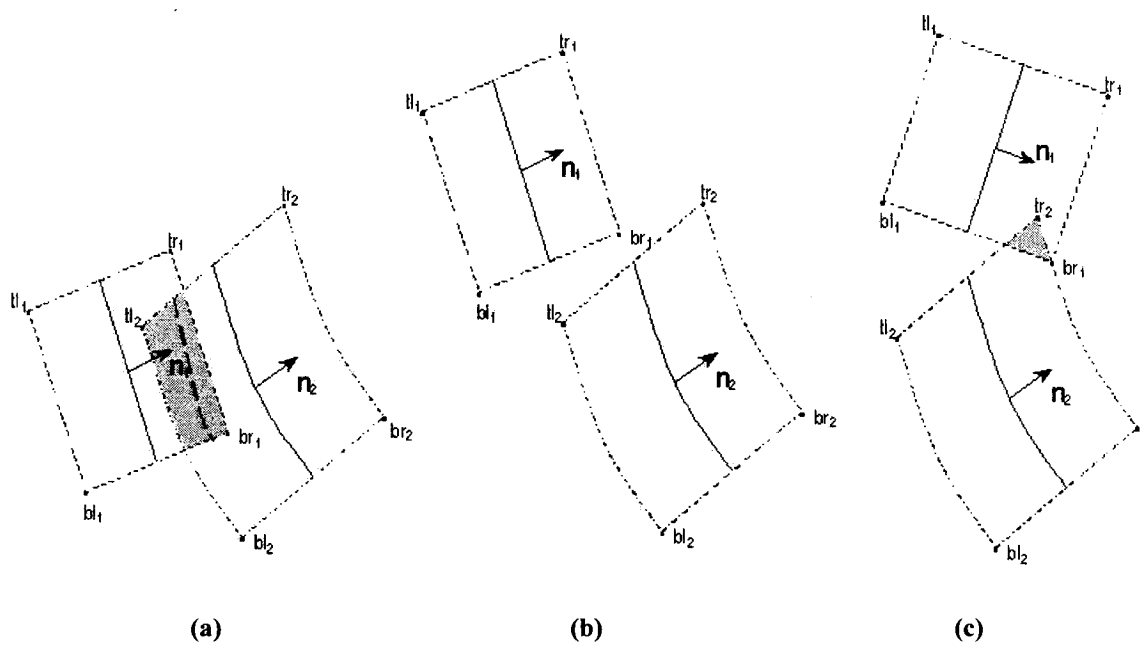


Figure 2.7: Examples of the proposed skeleton extraction scheme. In (a), a good skeleton has been extracted (dashed line). In (b), parallelism and proximity requirements have been met, but no skeleton has been extracted since the envelopes do not overlap. In (c) the envelopes overlap, and proximity requirements are met, but the lines are not sufficiently parallel to generate a good skeleton.

In the case depicted in (a), a good skeleton has been extracted. In (b), although parallelism and proximity requirements have been met, the vertical misalignment creates uncertainty about the true nature of the underlying structures – do the edges arise from a single instrument with weak boundaries, or is some other factor present? Perhaps only one edge corresponds to an instrument, and the other is nearby texture. To be safe, the set of equidistant points has been suppressed since there is no overlap between the edge envelopes. In (c), although the edges are close to one another and the envelopes have some overlap, it is hard to imagine that the two lines really enclose an instrument region. The overlap results from the relative orientation of the edges, which are sufficiently far from parallel that no skeleton will be extracted.

2.1.3 Final comments on proposed detection method

To conclude this section, several other aspects of the proposed skeleton extraction scheme are highlighted. For one, the parallelism requirements offer a natural guard

against the possibility of coincidental background textures leading instrument skeletons astray of true object shape. Problems like that illustrated earlier in figure 2.4 are less likely since edges arising from the background texture will violate the parallelism constraint. The parallelism requirement is not infallible, however, and it is not unreasonable to expect parallel textures to exist in the background. In these cases, the proximity and length parameters will provide an additional buffer.

It is also worthwhile to note that the proposed scheme preserves the dual nature of skeleton extraction. On one hand, if the approximate scales of the target objects in an image are known in advance, one need only retain skeletons who correspond to points that are appropriately separated from the underlying edge pair. On the other hand, if the scales are not known *a priori*, numerous scales can still be processed. However, an advantage of the new method is that all scales can be processed simultaneously. Current detection methods must consider each scale individually, and the processing requirements will quickly grow as the number of scales increases, especially as stronger blurring is applied. With the proposed method, edges are extracted all at once, with the associated skeletons/line-objects given by the set of equidistant points between edges. The distances of these points from their respective edges encode the scale of associated line-objects, such that the relevant scales of image line-objects are immediately available.

2.2 Segmentation: Probabilistic line-object selection

Compared to other line-object detection methods, the set of skeletons returned from the extraction method just proposed is treated as a comprehensive set of candidate skeletons, rather than as the final image segmentation. Therefore, the requirement now is to determine which of these skeletons really correspond to the target instruments in the image. The necessity of this step is illustrated by briefly returning to investigate another problem of existing line-object detection schemes.

2.2.1 Errors in current line-object segmentation methods

Recall the implicit assumption made by existing line-object segmentation schemes: any good responses arising from a line-object detector such as $R(\mathbf{x},\sigma_d)=\sigma_d^{\gamma}\cdot\lambda_1(\mathbf{x},\sigma_d)$

correspond to the type of object to be segmented. Additional to the problems this causes in the presence of weak instrument contrast and coincidental background textures, difficulties related to this assumption also arise because background tissues frequently contain line-like objects, both accidental and real. For instance, cut tissue can give the *appearance* of line-like structures, while anatomical structures such as veins actually *are* line-objects. The former case is illustrated in figure 2.4, where cut tissues and specular reflections have all created isolated regions of incorrectly segmented background. The latter case is detailed in the figure 2.8 below. In this figure, the skeletons detected by the segmentation scheme proposed by Koller et al. [1995] have been overlaid on the original image. As stated previously, these skeletons all encode scale information and thus represent the final image segmentation. The presence of blood vessels in the background has led to an unacceptable result – the objective is to segment only instruments.

By assuming any local extrema arising from a function such as (1.18) correspond to the objects to be segmented, current schemes ignore the possibility of non-target line-structures, and responses arising from target and non-target objects cannot be differentiated. As shown in figure 2.8, this represents a fatal limitation for images with complicated backgrounds.

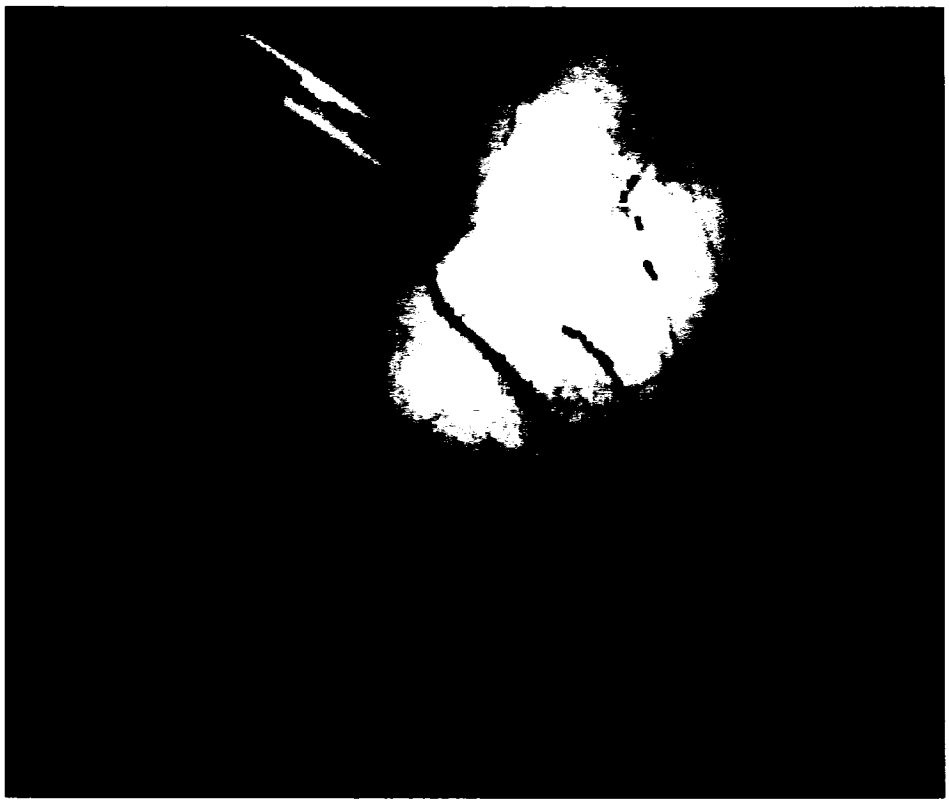


Figure 2.8: Problem of background line-objects in current segmentation methods. The skeletons are those detected by the method proposed by Koller et al. [1995]. Current methods do not consider the possibility of non-target line-objects.

2.2.2 Bayesian differentiation of target and non-target line-objects

Unfortunately, the Canny line-object detector, while designed to help reduce the number of background skeletons extracted, cannot be guaranteed to return only skeletons corresponding to instruments. Therefore, the segmentation method must now be extended by introducing the capability to discriminate between extracted line-objects. The extension again applies highly general assumptions about instrument shape, embedding these assumptions into a robust Bayesian regularization framework.

General Bayesian model

To initiate the development, note that after the extraction step, each skeleton will have a scale and length. This information can be used to identify which skeletons correspond to instruments by recalling that the earlier line-object characterization of instruments was

based on the observation that they must be narrow enough (small scale) to fit through the small incisions on a patient's body while long enough to reach the surgical site. The objective of the following development will be to embed these characteristics into a Bayesian expression for the probability that a skeleton corresponds to an instrument, given the skeleton's observed scale σ and length l:

$$P(Inst|\sigma,l)=\frac{P(\sigma,l|Inst)\cdot P(Inst)}{P(\sigma,l)}. \tag{2.2}$$

Referring to the discussion in section 1.2.2 of the Literature Review, (2.2) can be interepreted as follows: the left term in the numerator of the right-hand side represents the likelihood of a skeleton having the observed length and scale, assuming the skeleton corresponds to an instrument. The right term in the numerator reflects any *a piori* information regarding instruments in endoscope images. The denominator can be ignored since it does not depend on instruments and acts only as a weighting. Therefore only the terms in the numerator need to be defined more rigorously.

Likelihood term

Since the long-but-thin characteristic of instruments implies that corresponding skeletons should have a high length-to-scale ratio, the first term in the numerator of the right-hand side of (2.2) should assign higher probabilities to skeletons that reflect this characterization. However, any scheme that assigns probabilities based on this observation must respect that skeletons with high ratios are more likely at smaller scales. An instrument with a high image scale will most often result from close proximity to the camera. The increased proximity will mean less of the instrument's length will be visible, and the ratio of skeleton length to scale will necessarily decrease. Therefore, any likelihood term that looks for high ratios as a basis for isolating instrument skeletons should allow lower ratios as the scale of the skeleton increases. The following non-linear weighting can be applied to each skeleton, which assigns a weight on the interval [0 1) as illustrated in figure 2.9:

$$P(\sigma,l|Inst)=\frac{1}{2}+\frac{1}{\pi}\arctan\left(\alpha_\sigma(1-\beta_\sigma)\right)-P(\sigma=0,l=0|Inst). \tag{2.3}$$

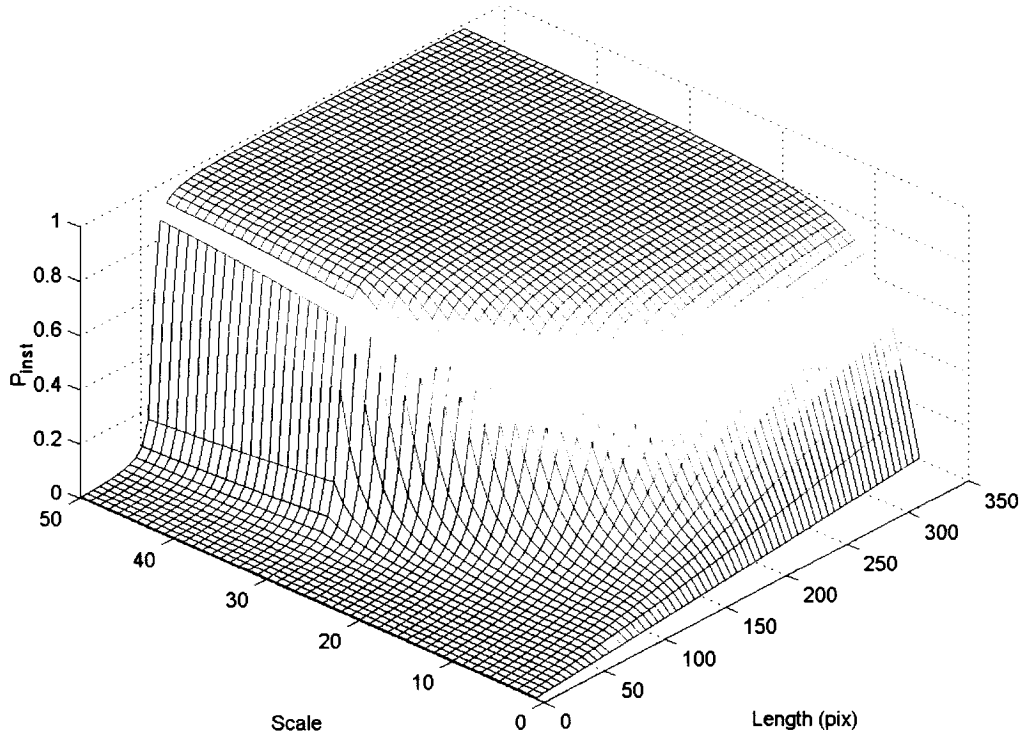


Figure 2.9: Likelihood term of the Bayesian line-object selection model. This functional assigns values nearer 1 to skeletons with larger rations of length-to-scale, but considers that at higher scales, this ratio will inevitably decrease.

The shape of this functional is controlled by the functions β and α. β positions the half-maximum point of (2.3) at the minimum length a skeleton must achieve before it is considered more likely to represent an instrument than a background texture. In other words, the function defines the shape of the P = 0.5 level of figure 2.9. This half-maximum length varies inversely with scale:

$$\beta(\sigma)=\begin{Bmatrix} l_{min} & \text{if} & \sigma>\sigma_h \\ \exp\left\{-\dfrac{\sigma-\sigma_l}{\eta}\right\}\cdot l_{min} & \text{if} & \sigma_h\geqslant\sigma\geqslant\sigma_l \\ l_{max} & \text{if} & \sigma<\sigma_l \end{Bmatrix}. \qquad (2.4)$$

In this equation, η controls the shape of β as it transits from high to low scales. It is specified according to:

(i) σ_h, the scale at which it is reasonable to assume an instrument may give rise to a

skeleton only l_{min} pixels long, and

(ii)σ_l, the smallest scale at which skeletons corresponding to instruments are expected, which may be up to l_{max} long.

Some further insight is useful: σ_h is related to the W_{max} parameter applied during skeleton extraction, which encoded the maximum expected width of instruments in an image. In fact, W_{max} is effectively the limiting value of σ_h. As this limit is approached, l_{min} reflects the expected decrease in line-ratio due to the reduction in the visible length of the instrument. At the other extreme, instruments are unlikely to be the smallest objects in an image, and even when the camera is far away, should have an image width of at least σ_l. The l_{max} parameter will be limited by the dimensions of the image itself. A detailed discussion of actual values for these parameters will be postponed until the results chapter, but it is worthwhile noting that for the segmentation method to be as general as possible, the aim will be to set these parameters to cover the extreme cases of instrument length and scale that might arise in any endoscope image.

The impact of the α function must also be addressed. This function specifies the slope of the transition of (2.3) through the $P = 0.5$ length. For most endoscope images, textures and tissues that generate longer skeletons are more common at lower scales. By extension, background skeletons that exceed the l_{min} threshold specified in (2.4) are less likely at higher scales. It follows that skeletons exceeding the $P_{0.5}$ length threshold at higher scales can be more confidently assumed to correspond to instruments than skeletons that exceed the threshold at lower scales. The α function embeds this reasoning into (2.3) by allowing a sharper transition to the high probability asymptote at higher scales:

$$\alpha(\sigma)=\frac{\sigma}{\beta_\sigma}. \tag{2.5}$$

The last term on the right hand side of (2.3) simply enforces that $P(\sigma,l \mid Inst) = 0$ for a zero-length chain.

Finally, to conclude the discussion and better appreciate how the likelihood term in (2.3) compensates for lower line ratios at higher scales, consider that the function in figure 2.9 was obtained assuming a 320x240 image, with the parameters set as $l_{min} = 60$,

$\sigma_h = 40$, $l_{max} = 240$ and $\sigma_l = 2$. These values generate a β function that sets the half-maximum point at a line ratio of only ~1.5 at higher scales, and at 100 or more for smaller scales. At the same time, the transition between low and high probabilities is slower at lower scales, reflecting decreased confidence that skeletons extracted at these scales correspond exclusively to instruments.

Prior term

As the reasoning for the form of the α function in (2.5) demonstrated, the likelihood term alone cannot be guaranteed to suppress all non-instrument skeletons: especially at lower scales, very long background objects may still exist and generate skeletons longer than the threshold assigned by the $P_{0.5}$ length. To increase the probability of isolating only instruments, *a priori* expectations about the target instruments can be applied. This probabilistic term will complete the development of the new line-object segmentation scheme by making explicit many of the general assumptions regarding instrument shape that have been used to guide the development of the segmentation scheme thus far.

Surface structure and object size

Regardless the *actual* scale at which instruments appear in an image due to varying distances from the endoscope, their scale *relative* to background objects is typically large. This stems from two factors: instruments are usually larger in a real physical sense than structures such as veins, and they are always as close, or closer to, the endoscope than background tissues. This implies an *a priori* weighting to favour those line-objects that are amongst the largest in an image. In fact, this kind of scheme has been used successfully in preliminary work [Windisch et al., 2005] (see Appendix A), but to make the differentiation task more robust against the possibility of accepting occasional high-scale background line-objects, the prior term can be formulated in a slightly different way.

Note that instrument surfaces generally give rise to line-object responses at numerous scales. The surfaces are often faceted and the resulting real discontinuities between faces will cause changes in the intensity profile that lead to either a response in a function such as (1.18), or will generate edge-content. The result is a set of closely clustered line-object

skeletons at numerous scales. Even continuous and smooth instrument surfaces will typically generate these clusters, since instruments are metallic and give rise to reflections. Changes in surface depth (e.g. curved profiles) and incident light due to reflections from nearby tissue will both generate changes in the intensity profile across an instrument, and these changes will give the appearance of line-objects at numerous scales. Compared to instruments, natural line-objects like veins are not as reflective and will not give rise to the same number of line-objects.

This difference can be exploited to improve the robustness of the characterization for instruments through a prior term that is built by first isolating the highest-scale skeletons extracted during the detection stage. Call these the focus-skeletons, which exist at the focus-scale. For each focus-skeleton, the scales of all smaller-scale skeleton that fall within its envelope[8] are recorded and the set of overlapping skeletons are considered to correspond to the same line-object. All skeletons in the resulting cluster are removed from further consideration. Once all skeletons at the focus-scale are processed, the procedure then considers the next lowest scale at which skeletons were found. If any skeletons remain unclustered at this scale, the clustering procedure is repeated. This process is iterated until all skeletons have been processed. At the end of the procedure, there will be some skeletons that remain isolated, while there will be others that are part of a larger group – these larger groupings are more likely to represent instruments, and a weighting is applied to favour these clusters:

$$P(Inst) = \frac{n_{skel}}{N_{max}} \tag{2.6}$$

In the above function, n_{skel} is the number of skeletons in a cluster, and N_{max} is the number of skeletons in the largest cluster of skeletons found. As desired, this function naturally incorporates a bias in favour of larger scale line-objects since it is easier for a larger object to envelope numerous skeletons than it is for a smaller object.

Eigenvalue response

One final detail must be addressed in the development of the prior term. Specifically, a limitation of the Canny extraction method applied earlier (see section 2.1.2) is the lack of

8 The idea of envelopes was illustrated in figure 2.7.

a way to determine from edge content alone whether two parallel edges actually straddle a real instrument region. The $W_{\max}$ parameter was developed to address this uncertainty by setting a limit on the instrument width expected in an image, but its use implies a trade off: to guard against edge pairs that straddle non-instrument regions, $W_{\max}$ should be set small. However, to keep the skeleton extraction step adaptable to the extremes of apparent instrument size arising from camera and instrument movements, larger $W_{\max}$ values are desirable.

At the onset of this chapter, considerable analysis was conducted into the eigenvalue-based line-object response function $R(\mathbf{x},\sigma_d)=\sigma_d^{\gamma}\cdot\lambda_1(\mathbf{x},\sigma_d)$. While poor resolution and sensitivity to background texture meant that a detection scheme based on this response was ill-suited to the endoscope segmentation problem, the eigenvalue analysis still provides a wealth of information that is useful for the current selection task. In the first place, it is important to understand that the poor resolution of the response only means it may be difficult to accurately locate instrument boundaries – it does not imply that the response itself is wrong. This information can be used to help resolve the $W_{\max}$ trade off because a skeleton generated by two edges that really bound a line-object should overlap image pixels that have a strong eigenvalue response. In the opposite case, two edges separated by a scale inconsistent with any underlying line-object can be expected to bound an image region with only weak line responses at that scale.

This observation is used to condition the *a priori* term by calculating the value $R_\sigma(\mathbf{x}_{skel})=\sigma_{max}^{\gamma}\cdot\lambda_{1,\sigma}(\mathbf{x}_{skel})$ at each point along the highest-scale skeleton (i.e. $\sigma_{\max}$) in each cluster identified in the previous step. The average response over the length of the skeleton is calculated and normalized by the maximum average value obtained from all clusters. Combined with (2.6) this results in the final form of the prior term:

$$P(Inst)=\frac{n_{skel}}{N_{max}}\cdot\frac{\bar{R}_{skel}}{\bar{R}_{max}}. \tag{2.7}$$

In this development, the natural question that arises is, since the eigenvalue detector was previously shown to respond to background textures (see section 2.2.1), how can it be used as a reliable indicator of instrument regions now? Although some responses may

still be wrong, these erroneous responses are generally associated with short edges and skeletons (consider figure 2.4), and the minimum length threshold τ_e applied in the extraction step should help eliminate these skeletons from consideration.

Therefore, the final form of the *a priori* term in (2.7) draws the development of the segmentation scheme to a close by embedding the fundamental expectations that motivated the line-object investigation in the first place; namely, that instruments have line-like characteristics, and should therefore have a decent response to a basic line-object detector function.

2.2.3 Final remarks on proposed line-object differentiation

Equations (2.3) and (2.7) are substituted into (2.2) to arrive at the final expression for the Bayesian line-object discriminant function:

$$P(Inst|\sigma,l)=\left\{\frac{1}{2}+\frac{1}{\pi}\arctan\left(\alpha_\sigma(1-\beta_\sigma)\right)-P(\sigma=0,l=0|Inst)\right\}\cdot\left\{\frac{n_{skel}}{N_{max}}\right\}\cdot\left\{\frac{\bar{R}_{skel}}{\bar{R}_{max}}\right\}. \quad (2.8)$$

The first term of this expression reflects the limitations of endoscope imaging: while an instrument is expected to be long and narrow, it is impossible to know *a priori* whether it will appear as such. This term is used to compensate for dramatic changes in instrument size as the situation dictates.

Unlike the length-to-width ratio of the visible portion of an instrument, expectations of large instrument scales *relative* to other objects, constant surface structures, and good overall responses to line-object detectors should not be influenced by camera movements etc. The second and third terms encode these *a priori* expectations.

To actually apply this probabilistic term to differentiate instrument skeletons from other objects in the image, it is important to note that (2.8) has not been developed as a *maximum a posteriori* detector. While it certainly could be used as such, it is risky to do so since there is no way to know the number of instruments that will be visible in an image. The alternative is to apply a threshold value for (2.8), and consider any skeletons that exceed it as instruments. The danger with this scheme is that the probabilistic model will be useless if the threshold needs to be constantly adjusted. With this potential limitiation in mind, every attempt has been made to design the Bayesian detector to work

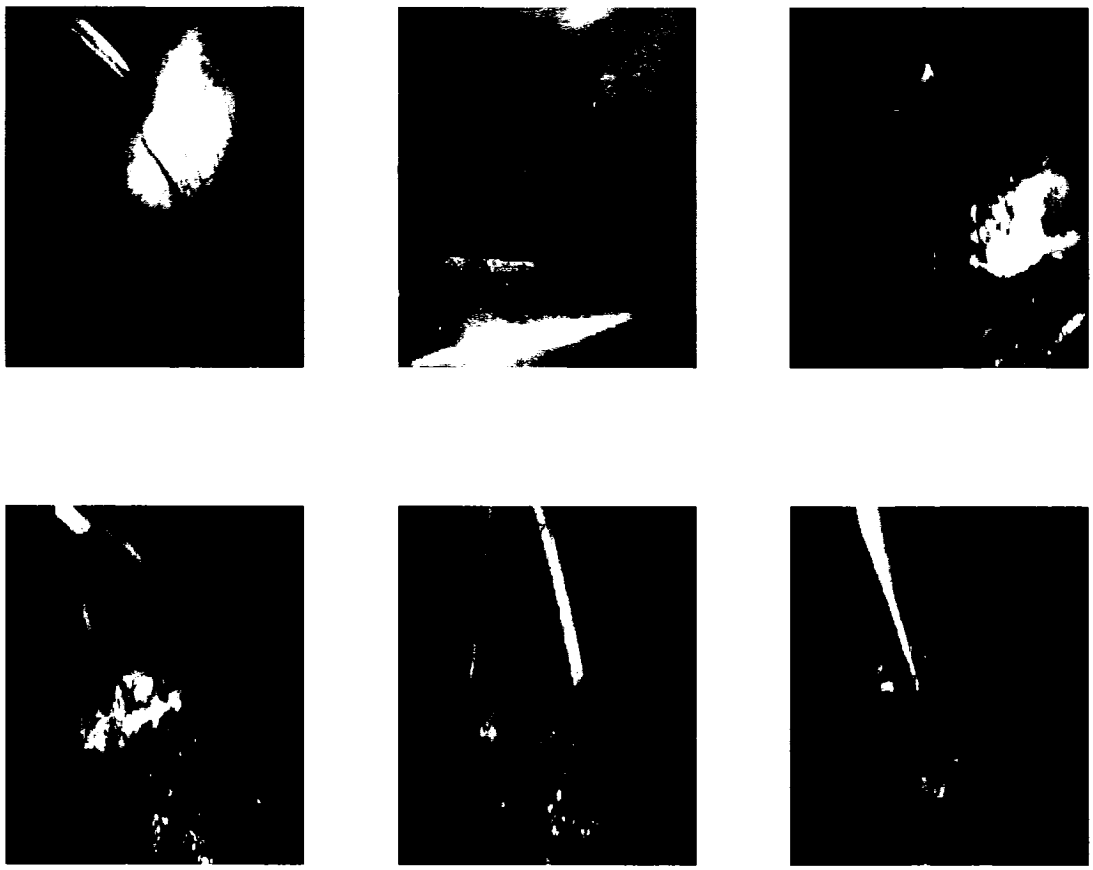


Figure 2.10: Endoscope images used to test the proposed segmentation scheme. The manually segmented "true" segmentations used to evaluate the proposed method are shown. Images will referred to as "Test Image 1" through "Test Image 6" according to a clockwise progression beginning in the upper left of the grid.

Each image was processed using MATLAB code. In all cases, the final segmentation was evaluated according to four criteria:

(1)The proportion of the ground truth segmentation returned by the segmentation;

(2)The proportion of the final segmentation that correctly included instrument region;

(3)The proportion of the final segmentation that incorrectly included background content, but was immediately connected to region content. In other words, this serves as a measure of any "leaking" of the segmentation past instrument boundaries, and provides an indication of the segmentation precision.

(4)The proportion of the final segmentation that incorrectly segmented background that was disconnected from any part of the region content. Isolated background like this

serves to indicate whether background line objects are being suppressed by the Bayesian scheme.

Table 2.1: Aspects of segmentation algorithm validated by different test images.

Test Feature	*Image Index*					
	1	**2**	**3**	**4**	**5**	**6**
Multiple instrument regions					X	
Textured background – without line objects			X	X	X	X
Textured background – with line objects	X	X				
Varying instrument scale		X	X			
Poor instrument contrast	X					X
Change in instrument appearance			X	X	X	
High scale instrument	X	X				X
Low scale instrument				X		

2.3 Object Tracking

Recall from the literature review that three broad categories of tracking methods exist: correspondence-based, transform-based, and contour-based. While a correspondence-based tracking scheme may turn out to be the most practical for the calibration objective, the current state of the system is still too preliminary to allow for its effective development. However, a coarse contour-tracking scheme to follow frame-to-frame instrument regions might be useful to constrain eventual correspondence-based feature trackers like the Kalman filter. A region-based contour-tracker requires an initial segmented region as input, which is now available via the extensive work done into the segmentation problem earlier. Accordingly, the system development is at a stage where preliminary investigations into contour-tracking can occur.

The tracking investigation is also motivated by a natural connection between Mansouri's Bayesian optical flow scheme and the Bayesian segmentation framework just developed. Accordingly, Mansouri's method forms the basis of the proposed tracking algorithm. In fact, the algorithmic aspects are identical except that the inter-frame instrument tracking will be implemented using the eigenvalue response generated during the segmentation step instead of the original intensity information. The overall structure of the resulting segmentation algorithm was outlined previously in figure 2.1, but before discussing this implementation in more detail, it will be worthwhile to revisit some of the specifics of Mansouri's method, especially the advantages of the Bayesian formulation.

2.3.1 Benefits of Bayesian tracking

The Bayesian framework for tracking employed by Mansouri has a couple of appealing aspects in the context of contour-tracking in endoscope video:

- Independence from motion models. Because the tracking itself is accomplished via level set contour evolution, there is no need to estimate frame-to-frame motions or transformations. The motion/transformations may need to be estimated eventually to provide an *a priori* estimate for a Kalman filter, but this information can be obtained from the already tracked regions *if* needed. Since the estimates are not required to *accomplish* the tracking, the tracking will be more adaptable to overall system needs.

- A stable implementation. The use of level-set technology also offers the advantage of numerical stability and topological independence. Considering the dramatic changes in instrument shape that might occur over an endoscope image sequence, including the possibility of multiple instrument regions (see figure 2.4 earlier), the topological robustness of level set methods is particularly appealing.

Unfortunately, these advantages are somewhat diminished for the current application by the fact that Mansouri bases his level set propagation speed F on differences in pixel intensities between frames. Remember from the Literature Review that the propagation speed was defined as $F_{i,j} = V_{\mathrm{x}}^{\mathrm{in}} - V_{\mathrm{x}}^{\mathrm{out}}$, where:

$$V_{\mathrm{x}}^{\mathrm{in}} = \inf_{\mathrm{z}:\|\mathrm{z}\|\le\delta,\,\mathrm{x}+\mathrm{z}\in R_0} \left(I^{n+1}(\mathrm{x}) - I^{n}(\mathrm{x}+\mathrm{z})\right)^2 \tag{2.9}$$

and

$$V_{\mathrm{x}}^{\mathrm{out}} = \inf_{\mathrm{z}:\|\mathrm{z}\|\le\delta,\,\mathrm{x}+\mathrm{z}\in R_0^c} \left(I^{n+1}(\mathrm{x}) - I^{n}(\mathrm{x}+\mathrm{z})\right)^2. \tag{2.10}$$

Considering the attention paid to the problems of highly variable intensity characteristics of endoscope images throughout this thesis, one might suspect that an intensity-based tracking formulation like this is not particularly well-suited to tracking instruments.

Problems with the intensity-based method

Suspicions regarding the limitations of the intensity formulation are confirmed by the following tracking results obtained using Mansouri's method:

Figure 2.11: Problems with intensity-based contour-tracking. Images from the first 12 frames of an endoscope video sequence are shown. Results of the region tracking method proposed by Mansouri [2002] are superimposed on each image. The strong blurring over portions of the instrument boundaries has caused the track to lose the instrument. See text for more details.

In areas of good contrast, the region has followed the target well, but in areas of reduced contrast and high blurring, the same cannot be said: by the 9th frame of the sequence, the tracked region has completely lost the right tip of the lower instrument prong. The principle difficulty is that the speed function in these areas takes a very low value, such that the region is not evolved. Recall from the Literature Review that the speed estimation based on (2.9) and (2.10) operates by first determining the minimum intensity difference V_x^{in} between the intensity at pixel $\mathbf{I}(i,j)$ at time n, and the set of region intensities in the neighbourhood of (i,j), $\mathbf{I}(i \pm \delta, j \pm \delta)$, in the previous frame. The same minimum difference measure is also calculated using the background intensities, producing V_x^{out}. If $V_x^{in} < V_x^{out}$ the pixel in image n+1 is classified as region, and the level set scheme evolves the region to include the pixel. If $V_x^{in} > V_x^{out}$ the opposite effect occurs. The problem with this scheme is that poor contrast may mean $V_x^{in} \approx V_x^{out}$, in which case the update speed will be near 0 and the region will not evolve even if the underlying object does move. This is exactly the behaviour observed in figure 2.11 above.

Although the curvature term is designed to resolve some of these uncertainties by penalizing the development of strong curvature along the boundary of a tracked region, the low-contrast problems present in endoscope images are generally beyond its capabilities to constrain. As a result, an alternate formulation for the speed functional has been developed.

2.3.2 Eigenvalue-based Bayesian tracking

The proposed solution returns to the idea of line-objects applied earlier in the segmentation development. Unlike local intensity characteristics, the overall structure of instruments should only vary slightly between two sequential frames in an image sequence. Accordingly, the value of a response function like $L(\mathbf{x},\sigma)=\sigma^{\gamma}\cdot\lambda_1(\mathbf{x},\sigma)$ should have good frame-to-frame consistency for the same object. Also note that the sign of the eigenvalue gives another description of underlying object: if it is negative, the object is brighter than its background, while if it is positive, it is darker. Therefore, even

though diminished contrast will affect the magnitude of $R(\mathbf{x},\sigma)$, its sign is much less likely to change.

The modified tracking algorithm flows directly from these observations. First, as before, the first image in the stream is segmented with the previously developed line-object method to initialize the tracked region. The segmentation will return the line-object response for the target object, as well as its scale. For the next image in the sequence, the line-object response at the scale of the target object is calculated over the entire image. The resulting line-responses of the two sequential images are then simply substituted for the intensity information in Mansouri's original tracking framework. The resulting speed functional and tracking algorithm are obtained immediately by replacing the intensity matrices in (2.9) and (2.10) with the eigenvalue response function $L(\mathbf{x}, \sigma)$.

2.3.3 Evaluation procedure

The tracking investigation is still preliminary, and the modified tracking algorithm has to-date only been tested with one image sequence from real endoscope video. The image sequence consisted of 25 frames – some of the first 12 are shown in figure 2.11. Clearly the tracking algorithms must ultimately work over much longer periods than this; however, given the difficult nature of endoscope images and the preliminary stage of investigation, these short sequences will suffice to evaluate whether the proposed tracking algorithm shows comparable or improved results compared to the intensity based method.

To make this assessment, the tracking results for both methods were evaluated qualitatively and quantitatively. For the quantitative analyses, every frame of the sequence was manually segmented as in the evaluation of the segmentation results. The same test criteria were also applied with the single exception that the two measures of segmented background were combined. These measures were separated for the segmentation algorithm so the precision of the segmentation and the efficacy of the Bayesian selection term could be evaluated; this distinction is not useful for the tracking results, since the interest is simply knowing how well the methods track the initially identified region.

CHAPTER 3 - RESULTS & DISCUSSION

The discussion now shifts from the development of the segmentation and tracking algorithms to focus on their implementation and results. Consistent with the progression of earlier chapters, the segmentation algorithm will be discussed first, followed by the tracking methods.

3.1 Segmentation: Line-object extraction

Recall from the Methodology section that the segmentation algorithm has been broken down into two sub-components: skeleton extraction, followed by selection of those skeletons that are most likely to correspond to real surgical instruments. These steps require numerous parameters to be set, and their choices and associated implications will be the focus of this section and the next.

The first of the steps, skeleton extraction, itself involves several sub-processes: after initial Canny edge detection, any edges that can be confidently disassociated with instruments are removed. This final set of edges is used to extract skeletons that should, for the most part, represent the principle shapes of surgical instruments visible in the image.

3.1.1 Canny detection

The results of Canny edge detectors depend on the standard deviation of the Gaussian kernel, as well as the high and low thresholds. These last two parameters especially wield important influence; a too-low high threshold will allow extraneous background content into the results, while a too-high low threshold will crop otherwise good edges that appear weak on account of locally poor contrast. To ensure instrument content is not prematurely lost, it is preferable to avoid aggressive removal of edge content now. Accordingly, relatively low values for the thresholds have been explored.

Having made this decision, the actual choices of the parameters need to be addressed. There are two possible approaches: setting the parameters adaptively and automatically for each image, or setting values that remain fixed for all images. An adaptive approach seems the most intuitive for the endoscope application; however, a rigorous analysis is

worthwhile.

To begin, each test image was processed with the automatic parameter selection scheme implemented in MATLAB. This sets the standard deviation to 1 and the high and low thresholds to proportions of the total image pixels that are classified as edge pixels. The parameters returned by the automatic scheme were used as a starting point to investigate different fixed values for each parameter. Across all the different parameter values explored, including the automatic values, the two combinations that produced the best results in a qualitative sense are summarized in table 3.1. To provide some quantitative insight, the returned edges were further compared against the ground-truth segmentations discussed in the Methodology chapter. Measures on the number of region and background edges, as well as their average and median lengths were computed, and these results are included in the table.

The quantitative results help highlight which parameter settings give rise to the generally desirable characteristic of long region edges and short background edges. Even so, the qualitative analysis was used as the primary selection criterion since quantitative measures are difficult to use to evaluate the effect of marginal changes in parameter values: an importantly positioned edge with average length may be deleted without dramatically influencing quantitative measures, yet if the edge is a defining instrument edge and its removal means a skeleton will be lost, then its impact might be dramatic for overall results. This kind of effect is easier to predict qualitatively. Still, the quantitative results are useful to verify that the qualitative analyses makes sense, and to differentiate between qualitatively similar results.

Table 3.1: Canny edge detection results for different parameter combinations. For each endoscope image, the two best results (in a qualitative sense) are shown, with the better of the two in bold.

	Canny Parameters			*Region Edges*			*Background Edges*		
Img	σ_{canny}	τ_{low}	τ_{high}	*N*	*Av. len.*	*Mdn. Len.*	*N*	*Av. len.*	*Mdn. Len.*
1	2	0.04	0.09	32	42.0	19	279	32.3	15
1	**2**	**0.04**	**0.25**	**18**	**60.8**	**32.5**	**23**	**67.5**	**50**
2	1	0.04	0.25	41	46.3	22	27	21.0	12
2	**3**	**0.04**	**0.30**	**18**	**89.7**	**68**	**14**	**49.3**	**46**
3	**1**	**0.04**	**0.15**	**62**	**35.6**	**19.5**	**343**	**20.5**	**15**
3	2	0.04	0.20	31	54.8	48	172	31.3	20
4	1	0.04	0.20	37	41.8	14	541	19.1	14
4	**2**	**0.04**	**0.20**	**23**	**64.0**	**37**	**302**	**27.8**	**21**
5	**1**	**0.04**	**0.20**	**22**	**97.2**	**84.5**	**474**	**18.3**	**14**
5	2	0.04	0.15	18	110.9	71.5	493	23.3	18
6	**1**	**0.04**	**0.15**	**62**	**28.1**	**18**	**702**	**16.8**	**12**
6	2	0.04	0.25	38	39.2	19.5	393	24.9	19

For all the test images considered, the fixed parameter cases out-preformed the automatic scheme. Yet this observation alone is not satisfactory, since parameters of the segmentation scheme cannot be fine tuned for each case individually. Therefore, it is necessary to see if the best results from each image can be combined into a single set of parameters that can be applied to all images with comparable or superior results to the other parameter combinations explored. To determine such an optimal set, the parameter values from table 3.2 across all images were first averaged:

Table 3.2: Best Canny detection parameters averaged across all endoscope test images.

	Best Result			*Second Best Result*		
Img	σ_{canny}	τ_{low}	τ_{high}	σ_{canny}	τ_{low}	τ_{high}
1	1	0.04	0.25	2	0.04	0.09
2	3	0.04	0.30	1	0.04	0.25
3	1	0.04	0.15	2	0.04	0.20
4	2	0.04	0.20	1	0.04	0.20
5	1	0.04	0.20	2	0.04	0.15
6	1	0.04	0.15	2	0.04	0.25
Avg.	***1.8***	***0.04***	***0.2083***	***1.5***	***0.04***	***0.2***

The two average values for each parameter were then combined with a bias towards the best measure according to equation (3.1) below to produce the final, global value for each.

$$PARAM_{\text{final}} = \frac{1}{2} * \left\{ 1.5 * \overline{PARAM}_{\text{best}} + 0.5 * \overline{PARAM}_{\text{second}} \right\} \tag{3.1}$$

This measure produced the parameter set (τ_{low}, τ_{high}, σ_{canny}) = (0.04, 0.2062, 1.725), and the results using this set on a subset of the six test images are shown in figure 3.1. For the purpose of a final comparison, the results obtained with MATLAB's automatic scheme and using the best parameters from table 3.1 are also shown. Results using all three parameter sets on all six images are also summarized in table 3.3.

From these results, several observations come to light. First of all, the visual results support earlier comments that the best image-specific parameters produce better results than the automatically determined parameters. In the latter case, too many noisy edges seem to have been admitted without any significant gain in instrument edge content. This observation is supported both by qualitative and quantitative results. Moreover, while more sophisticated automatic selection schemes might be explored, they are unlikely to produce significantly improved results over the global parameter set, since this set already gives comparable results for each image as the best parameter sets for each image. Therefore, as just mentioned above, the final Canny edge detection parameters for the segmentation algorithm have been taken as (τ_{low}, τ_{high}, σ_{canny}) = (0.04, 0.2062, 1.725).

Table 3.3: Final results for Canny edge detection parameters. The first row for each image corresponds to automatic parameter selection as implemented in MATLAB. The second row contains the image-specific best parameter set, and the third row the results using the globally optimal parameter set determined by (3.1).

	Canny Parameters			*Region Edges*			*Background Edges*			
Img	σ	τ_{low}	τ_{high}	*N*	*Av. len.*	*Mdn. Len.*	*N*	*Av. len.*	*Mdn. Len.*	*Mode*
1	2	0.038	0.094	32	42.0	19	279	32.3	15	3
1	1	0.040	0.250	19	68.7	32	74	39.0	21.5	3
1	1.725	0.040	0.206	21	58.9	26	54	62.3	46.5	18
2	3	0.081	0.203	33	55.6	25	24	48.1	43.5	43
2	3	0.040	0.250	28	63.4	30	16	50.4	46	33
2	1.725	0.040	0.206	37	54.8	27	29	39.2	25	3
3	3	0.056	0.141	34	49.9	21	222	29.6	16.5	3
3	1	0.040	0.150	62	35.6	19.5	343	20.5	15	3
3	1.725	0.040	0.206	28	59.5	37.5	176	30.1	20.5	3
4	2	0.038	0.094	62	30.7	13.5	362	26.1	14	3
4	2	0.040	0.200	46	34.9	14.5	106	40.9	24.5	3
4	1.725	0.040	0.206	49	33.2	21	112	37.3	24	3
5	2	0.056	0.141	20	100.3	61.5	521	22.5	18	3
5	1	0.040	0.200	22	97.2	84.5	474	18.3	14	3
5	1.725	0.040	0.206	15	135.7	134	359	24.4	17	3
6	2	0.063	0.156	43	35.3	19	502	21.9	16	3
6	2	0.040	0.150	38	39.2	19.5	393	24.9	19	3
6	1.725	0.040	0.206	43	32.4	22	383	25.1	19	3

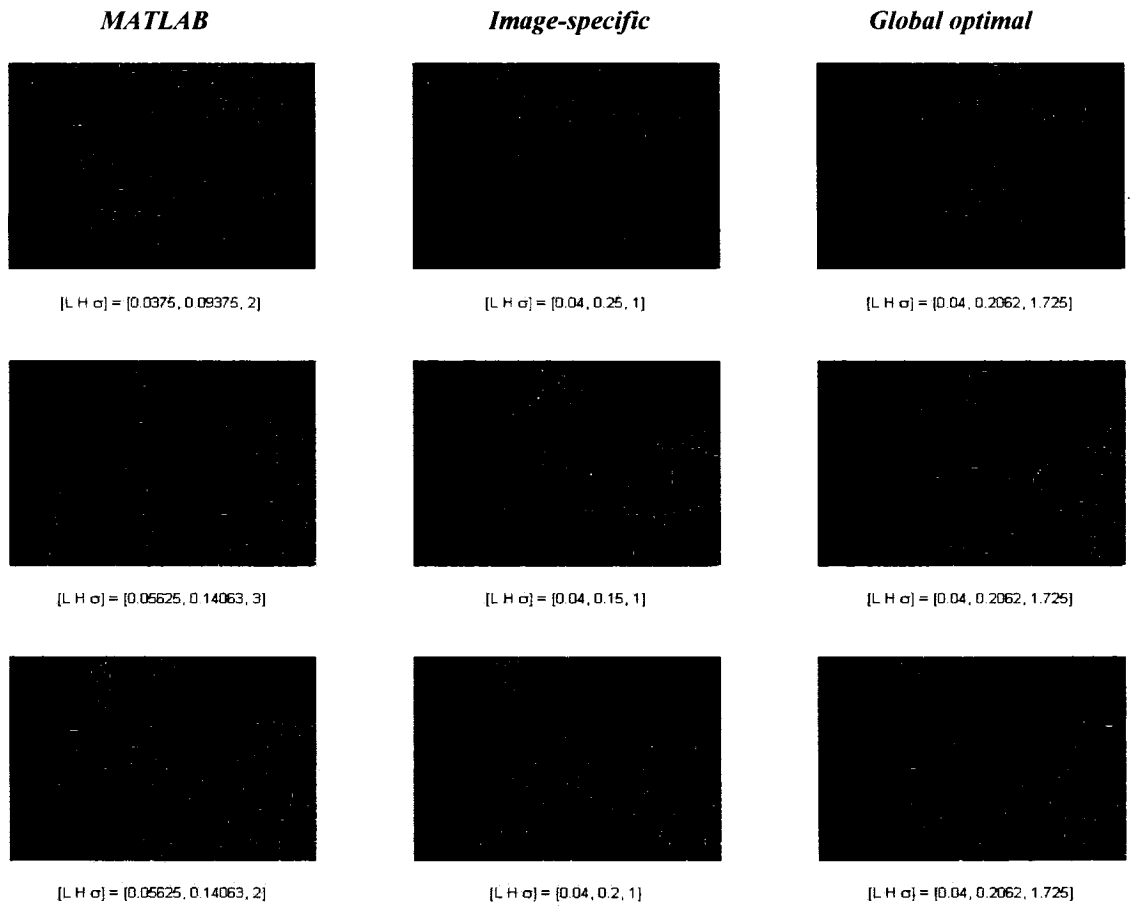


Figure 3.1: Examples of Canny edge detection results on endoscope images. The rows correspond to test images 1, 3 and 5 respectively. Columns correspond to varying detection parameters. Default parameters identified by MATLAB (first column) under-preform the other two cases. The global parameter set (third column) generated results comparable with the image-specific best parameters (second column).

3.1.2 Curvature-based edge removal

Armed with the results of the Canny edge detection, the next task of the skeleton extraction process is to establish a curvature-based measure to break edges into segments. The resulting segments will be used to enforce straightness assumptions, helping reduce the set of edges used for the final skeleton extraction to include only those with a high chance of corresponding to instrument content.

Minimum length

As discussed previously in the Methodology chapter, straightness constraints will be

enforced by imposing minimum length assumptions on the edge segments remaining after curvature-based breaking of the Canny results. Instruments are expected to be long and straight compared to edges arising from background tissue, and so should remain long even after the curvature-based segmenting.

To establish the minimum length threshold, the statistics presented in table 3.3 are instructive. As with the Canny parameter investigation, a globally applicable threshold is desired, which suggests that statistics corresponding to region edges should be ignored. Relative motions between the camera and instruments cause large changes in the image appearance of instruments, such that region edges are unlikely to offer stable length statistics across a wide range of cases. In contrast, background edge content should be more consistent since the appearance of edges arising from tissues does not depend as much on the particularities of instrument visibility. Therefore, background statistics will be explored in more detail.

The mean length statistic will not be used since there is no way of knowing the nature of the underlying edge length distribution. Consequently, it is difficult to understand whether a threshold based on the mean or standard deviation will represent a good differentiator between region and background edges. Take for example the case in image 1 (see figure 3.1): in this image, the background content is unusually long – the average background edge length is 77% of the average region edge length, compared to a maximum background edge length less than 60% of region length across the other test images. A threshold using the mean seriously risks cropping too much region content if it is not set well.

Instead, the mode of the background edge lengths appears to be more a descriptive and stable statistic. Specifically, the results of figure 3.1 suggest there are a lot of short edges that could be removed from the image regardless of any curvature-based breaking, since these will not produce good skeletons. These short edges are the most common ones, and not only does the mode gives a good sense of their length, but the statistic seems quite stable across the test images, usually about 3 pixels. The exceptions are images 1 and 2 (see figure 2.10), but both of these have quite smooth backgrounds which produce only a few longer edges. Longer edges are more likely to produce good

skeletons and should not be eliminated at this stage anyway, since there is not enough information yet to determine whether they represent region or background content. Therefore, they will not influence the value of the mode statistic used.

Further insight can be gained from the median statistic. Assuming that the best region edges are amongst the longest in the image, the median seems a good upper limit for any length threshold. Therefore, with the mode as a lower limit and the median as an upper limit, a length threshold of $\tau_e = 10$ pixels seems reasonable. A more restrictive threshold nearer the median edge length could be established, and/or a more rigorous analysis applied, but the objective at this point is only to remove edges that clearly do not correspond to instruments. Therefore, a safety buffer that errs on the side of retaining too many edges rather than too few is acceptable, and the subsequent Bayesian selection step can be used to make a more informed cropping of those edges that remain.

Maximum curvature

Before applying the minimum length criterion, the Canny edges will first be broken at points where their curvatures exceed a given threshold[9]. An appropriate curvature value can be determined by investigating how edge content is affected for different threshold values. The following figure highlights the gain in edge content as the curvature threshold is set at higher values (as the threshold increases, higher curvature in edges is permitted, and it will be less likely for an edge to break into short pieces). It reveals several distinct curvature values that produce dramatic edge content changes. The locations of these points are summarized in table 3.4.

9 The minimum length criterion previously established is used throughout the development of the curvature threshold to prune the broken edge segments.

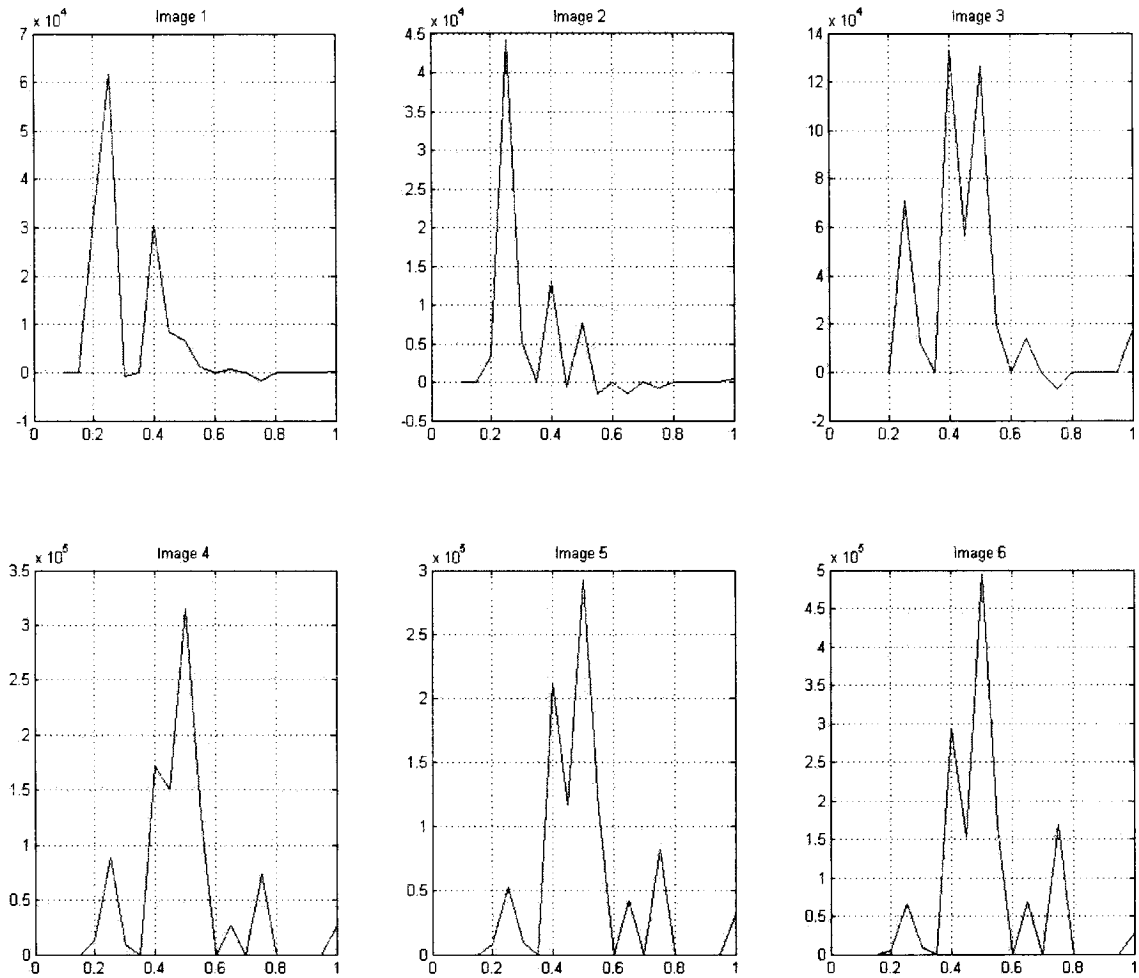


Figure 3.2: Edge pixel gain for different curvature thresholds. Each graph corresponds to the indicated test image, and represents the marginal increase in edge-content for increases in the curvature threshold. The figures illustrate that there will be a few distinct curvature thresholds that give large changes in edge content. These thresholds show a high consistency between images, as summarized in table 3.4.

Table 3.4: Curvature thresholds yielding largest changes in number of edge pixels.

Img	*Peak 1*	*Peak 2*	*Peak 3*
1	0.25	0.4	--
2	0.25	0.4	0.5
3	0.4	0.5	0.25
4	0.5	0.4	0.25
5	0.5	0.4	0.75
6	0.5	0.4	0.75

Table 3.5 summarizes the nature of the change in edge content around these threshold values by using the ground truth segmentations to tabulate the actual changes in the number of background and region edges across the peak values listed in table 3.4.

Table 3.5:Quantitative differences in edge content for changes in curvature threshold.

	$\kappa = 0.2$ *to* $\kappa = 0.3$		$\kappa = 0.3$ *to* $\kappa = 0.45$		$\kappa = 0.45$ *to* $\kappa = 0.6$	
Img.	ΔN_R	ΔN_B	ΔN_R	ΔN_B	ΔN_R	ΔN_B
1	2	21	0	-5	1	-4
2	21	15	-2	1	-3	1
3	17	46	2	49	-2	22
4	2	68	0	83	-1	52
5	-3	51	1	107	1	59
6	16	49	0	122	4	65
Avg	***9***	***42***	***0***	***60***	***0***	***33***

To analyze these data, consider that the objective of the curvature threshold is effectively to retain long instrument-region edges while eliminating short background ones that clearly do not correspond to instrument content. The results in table 3.5 indicate that a change in the curvature threshold from less than 0.25 to greater than 0.25 generally introduces more instrument edges than corresponding shifts across curvature values of 0.4 and 0.5. For these higher curvature values, the gain in edge content is primarily in the background. Therefore, a curvature threshold τ_κ in the range of $0.25 \leq \tau_\kappa < 0.4$ is probably appropriate.

This range is supported by figure 3.3. Between the images in the first two columns, which corresponds to an increase in τ_κ across 0.4, there is a noticeable increase in the background edge content, but not in the region edge content. Between the second and third columns' images, corresponding to an increase in τ_κ across 0.5, it is difficult to discern any significant change in the visual content since the edge gain is mostly in the unstructured background.

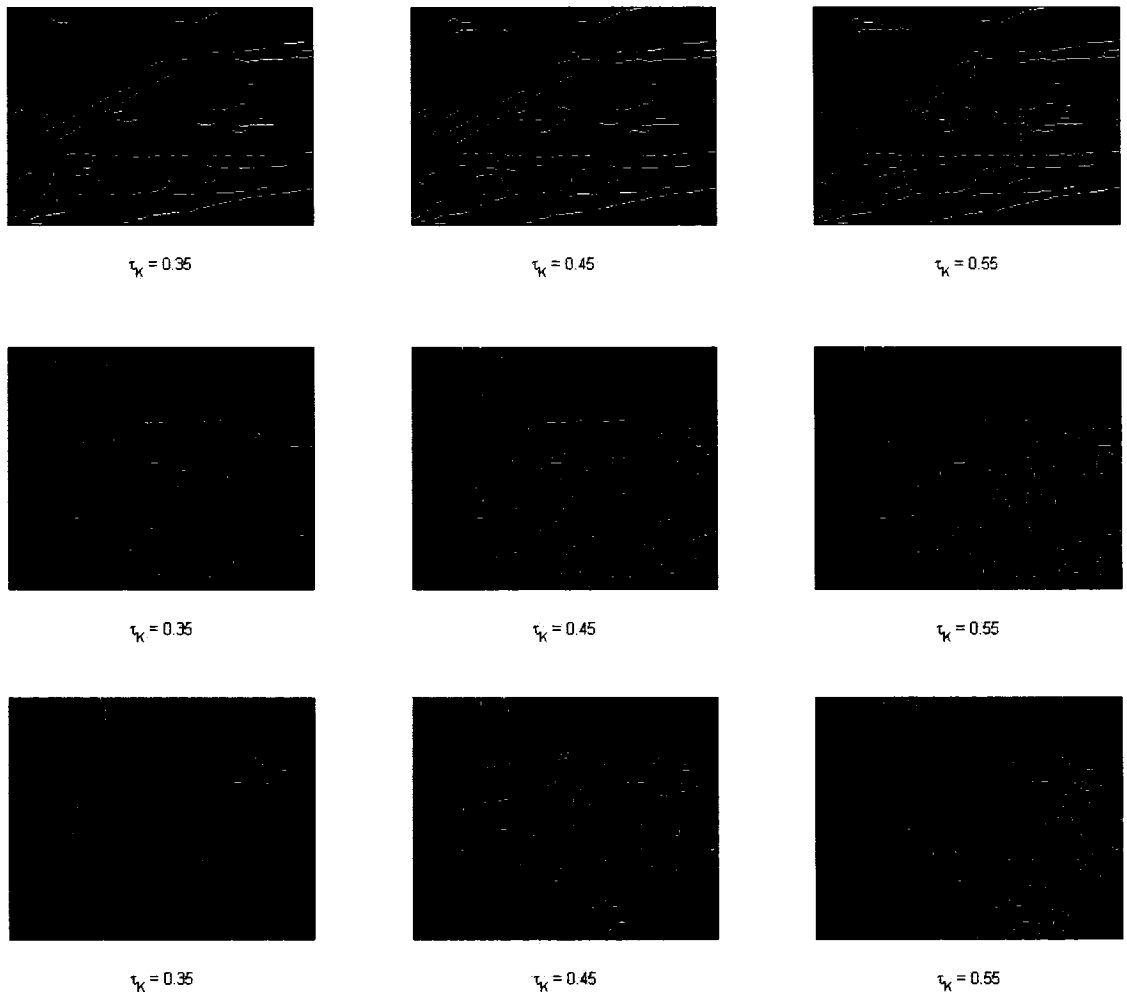


Figure 3.3: Qualitative differences in edge content for changes in curvature threshold. Between the images of the first two columns there is a noticeable gain in background edge content, but not in the instrument edges. Between the last two columns, the gain is principally unstructured background edges, and is difficult to discern visually. The first two columns straddle the curvature value of 0.4 and offer support for $\tau_\kappa < 0.4$.

Although instructive, the current threshold range still needs to be distilled to a single value that can be applied to all images. To that end, identical investigations into the change in edge content across different thresholds have been run iteratively to refine the threshold range, ultimately arriving at a curvature threshold of 0.358. To illustrate the efficacy of this threshold, consider the *second* column of figure 3.4. The first row shows the results of the Canny edge detection using the global parameter set established earlier. The second row shows the edges remaining after applying a curvature threshold of 0.35

and minimum length threshold of 10. In the third and fourth rows the curvature threshold has been increased to 0.358 and 0.55, respectively. Between $\tau_\kappa = 0.35$ and 0.358, there is a noticeable difference in the instrument edge content, even though the threshold value has changed only by 0.008. In contrast, for a change of 0.2 between the last two rows, there is not much visual difference, which is consistent with earlier observations regarding principle gains of unstructured background edges.

This figure also offers a final validation of the global parameter set for Canny edge detection established earlier. The *first* column images present results with the identical curvature thresholds used in the second column, but with the automatic parameter selection scheme implemented by MATLAB. Even after using curvature and minimum length criteria to prune bad edges, which has significantly improved the content, the results obtained with the global parameter set (second column) are markedly better.

Skeleton extraction

With the parameters for edge detection and preliminary edge removal established, an initial set of skeletons can now be extracted as the set of equidistant points between edges. This extraction requires additional parameters to be set, including τ_θ, the minimum degree of parallelism that must exist between two edges before any skeleton between them is considered, and W_{max}, the maximum allowable object width they can encode (i.e. the maximum perpendicular distance between them).

These are difficult parameters to evaluate without considering the final Bayesian segmentation results in parallel. Unlike previous steps, which removed edge content that could be confidently associated with non-interest objects, at this point it is assumed that the skeletons to be extracted will closely represent the desired segmentation. While there will certainly be non-target skeletons, there is too little information to isolate these via only the parallelism and maximum scale parameters. In short, the problem is ill-posed, and the significant risk of attempting to set the parameters now is choosing too-low values that will preclude both instrument and non-instrument edges from being used to generate skeletons.

Unlike Canny detection, these parameters cannot simply be set to permissive values

that err on the side of too many extracted skeletons with a total reliance on the subsequent probabilistic scheme to differentiate them. The problem is that the edge-based skeleton extraction is limited in that *any* two non-co-linear edges in the image will produce a skeleton, even if there is not underlying line-object content. The eigenvalue response *a priori* term of the Bayesian selection scheme is designed to help eliminate some of these pseudo-skeletons, but its chances for success will improve if skeletons that are clear artifacts of this extraction scheme can be removed first. Therefore, a good upper limit for the parallelism and maximum width/scale parameters should be established – one that is high enough not to suppress potential instrument skeletons, but low enough to eliminate clearly bogus skeletons that are mere artifacts of the edge-based extraction. The suitability of the parameter values for this objective cannot really be assessed without seeing their influence on the Bayesian selection scheme in the next section.

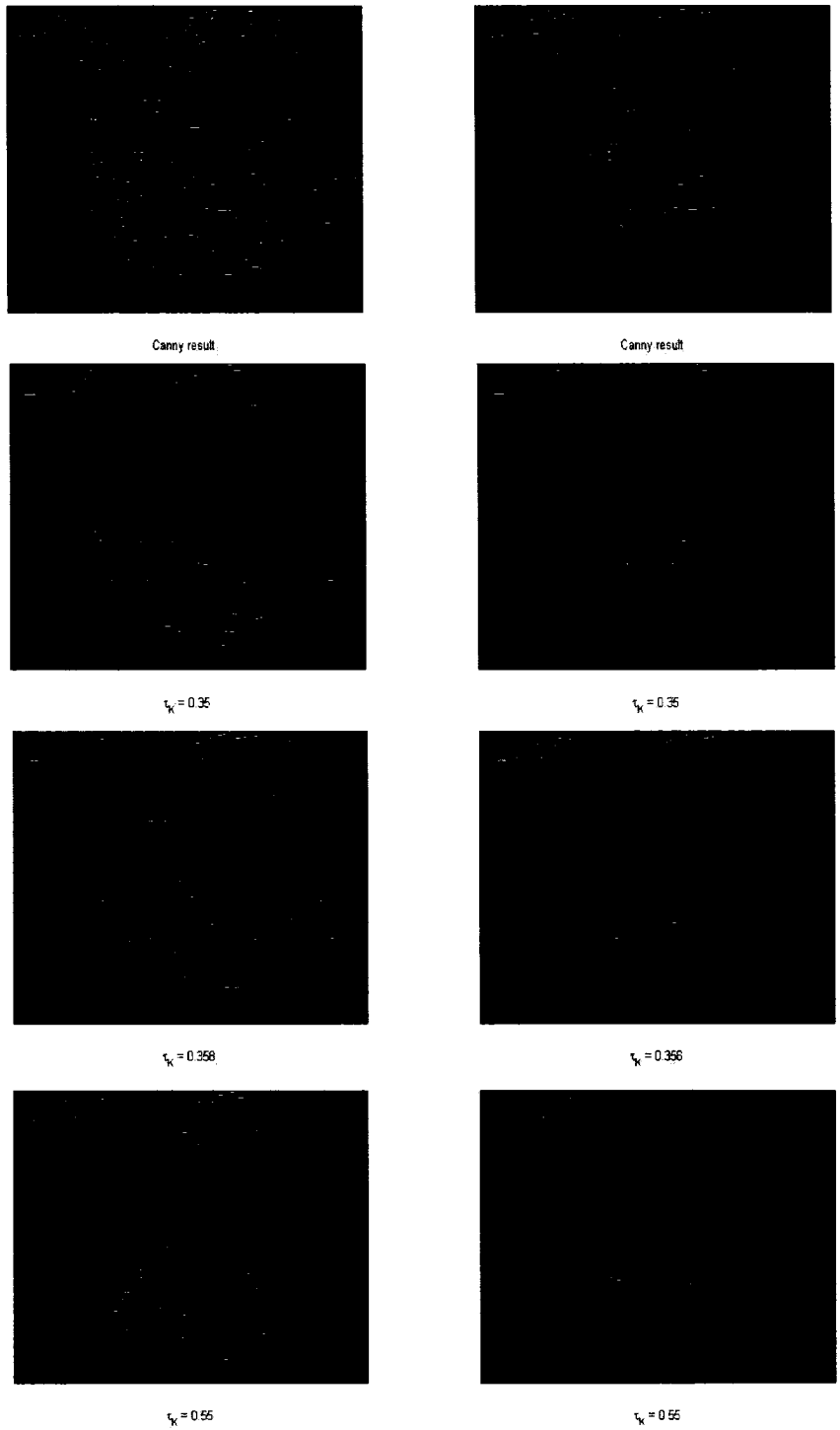


Figure 3.4: Final results of initial edge detection and removal. The global parameter set for Canny detection (second column) outperforms the automatic parameters set by MATLAB (first column), regardless the curvature threshold (rows). Edge changes between rows support $\tau_K = 0.358$ (third row).

3.2 Segmentation: Line-object selection

Parameters for the Bayesian selection scheme include the probability threshold used to select the final set of skeletons, those controlling the shape of the likelihood function, as well as the two skeleton extraction parameters. As the discussion at the end of the last section implied, the different parameters of the overall Bayesian selection scheme have a high degree of inter-dependence. For instance, too-high skeleton extraction parameters may produce large-scale pseudo-skeletons that compromise the assumption that background skeletons are less likely at-high scales, which is the cornerstone of the likelihood function. Therefore, more restrictive likelihood parameters will need to be set, and this will surely impact the results of the normalized *a priori* terms, and by extension, the final probability levels of the extracted skeletons. This will demand that the probability threshold be adjusted. The interplay of these parameters is difficult to evaluate analytically, so that the best approach for parameter selection at this time seems to be a brute-force test of different combinations. The ground truth segmentations will again be used to asses the quality of the results.

3.2.1 Review of probabilistic terms and parameters

Before proceeding, a brief review of the different probabilistic terms and parameters will be instructive.

Likelihood term

The five parameters to control the likelihood function are the high and low scales, the long and short lengths, and the transitional slope α. The scale and length parameters refer to the width and length of the detected line-objects, as encoded by each skeleton. These parameters determine the shape of the 0.5 probability level of the likelihood function, with the high-scale and minimum-length thresholds accounting for camera movements that are expected to cause instruments appearing at the 'high-scale' to produce skeletons only 'minimum-length' pixels long. At lower scales instruments should produce more traditional long-and-thin line-objects, and this expectation is embedded in the minimum-scale and maximum-length thresholds. At lower scales, background line-objects are also more likely, and these last two parameters are difficult to use to differentiate instruments.

Instead, they are most useful for controlling the logarithmic shape of the 0.5 probability level through intermediate scales and lengths.

The fifth parameter, α, determines how the likelihood function transits from low- to high-probability states. This parameter is determined automatically, and does not need further attention.

Prior terms

The prior terms are based on a parameter-free dynamic normalization scheme. Even so, remember that they are rooted in expectations that instruments (i) exist at large relative scales, so are likely to produce clusters of skeletons; and (ii) should generate overall good responses to an eigenvalue-based line-object detector.

Probability threshold

The probability threshold for the final skeleton extraction is needed since the use of a typical maximum a posteriori probability (MAP) selection scheme would assume that instruments are always present in endoscope images. Moreover, even were this not a dangerous assumption in-and-of-itself, a probability threshold would still be needed to extract more than one instrument region. A probability threshold of 0.5 was previously proposed based on the development of the likelihood term, but this value will need to be investigated in more detail: more important than achieving some specific target value is having one that can be applied reliably to all images.

3.2.2 Parameter investigation

Enumerating all the parameter combinations considered in the development of the Bayesian selection scheme is tedious and not particularly insightful. Rather, the following series of figures and tables present representative and descriptive summaries that highlight some of the important trade-offs, advantages, and limitations.

Skeleton-extraction parameters

As previously mentioned, the objective when setting the two extraction parameters is to leave as much of the skeleton selection process to the probabilistic measures as possible, while still removing pseudo-skeletons that only exists as a consequence of the edge-based

extraction. Results for restrictive, mid-range, and flexible values are presented in figure 3.5. All of these results apply fixed values for the likelihood parameters (σ_h = 40, l_{min} = 40 pixels, σ_l = 24, l_{max} = 80 pixels) and probability threshold (τ_B= 0.1).

In the first row of images, the maximum scale parameter W_{max} has been fixed at 90, and the parallelism requirement becomes more inhibitive from left to right. In the first case, edges are required only to be within 15° of parallel, and while most of the instrument region has been extracted, a significant portion of the background has also been segmented. The upper portion of the background segmentation does not contain much good line-object content, and is likely a result of pseudo-skeletons. In the next two images, the parallelism constraint is more restrictive, at values of 12° and 10°. In both cases, there is a slight reduction in the amount of background segmented, but the incorrect background region persists strongly anyway. More importantly, the demanding angular thresholds have reduced the amount of region retained. This implies that more lax angle requirements are desirable. However, before this can be concluded confidently, it is important to know whether the angle thresholds shown are already too lax, and therefore the cause of the segmented background, or if it is a result of the permissive maximum scale used in all these cases.

To investigate this issue, the second row of figure 3.5 highlights the characteristic influence of the W_{max} parameter seen throughout the Bayesian investigation. In all the images, the angular threshold has been set lax while the maximum scale parameter decreases (becomes more exclusive) from 90 to 46 to 20, left to right. Unlike the angular threshold, the results demonstrate a dramatic improvement for smaller scale values: a complete removal of the background content and strong retention of region content. This implies that the segmented background may have included pseudo-skeletons due to the high maximum scale value, and this parameter needs to be set smaller. Yet, as with the angular threshold, a too-small value causes a loss of region content. Overall then, a lax maximum scale parameter is quite detrimental, while larger angular thresholds are useful to help ensure maximum region content is retained.

Based on this analysis and all the combinations tested, the final value for the maximum scale parameter is taken as 46, and the value for the parallelism constraint, 15°.

The parameters correspond to the results shown in the second image of the second row of figure 3.5.

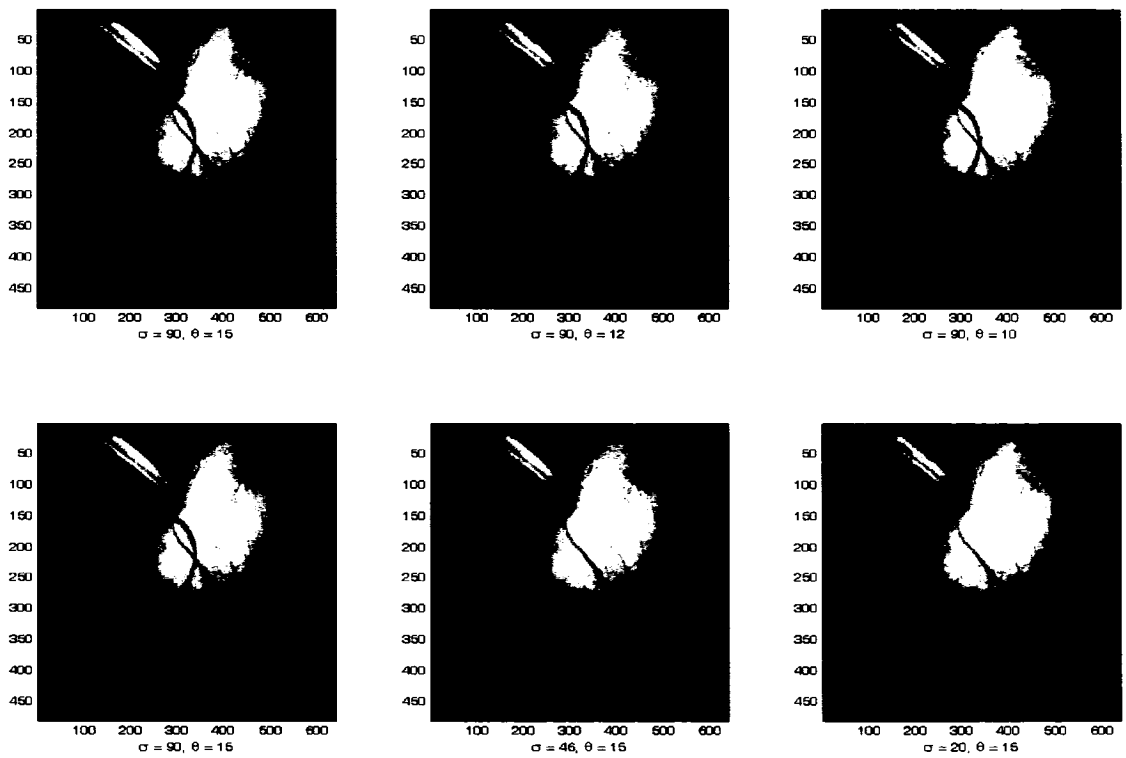


Figure 3.5: Qualitative results for varying skeleton extraction parameters. First row: fixed maximum scale constraint (90) with increasingly strong parallelism criterion left to right (15°, 12°, 10°). As parallelism is more strongly enforced, region content is lost with little change in the incorrectly segmented background. Second row: fixed parallelism constraint (15°) with decreasing maximum scale left to right (90, 46, 20). As the maximum scale is reduced, incorrectly segmented background disappears. Too-small maximum scale causes loss of region content.

Likelihood parameters

The investigation of the likelihood parameters occurred in exactly the same manner as with the skeleton extraction ones: with certain parameters fixed, adjust other ones and evaluate the result against the ground-truth segmentations presented in the Methodology chapter. For the skeleton extraction, the results with respect to these ground truth segmentations were not discussed or shown – they would have cluttered the example images. To shine light on how these ground-truth references were used, example results

for image 3 and different likelihood parameters are summarized in table 3.6. The specific results shown apply fixed values for the probability threshold (0.1) and the two skeleton extraction parameters (the values just determined); however, it is important to reiterate that this table presents only a subset of all the parameter combinations considered, and that during the actual testing these last three parameters were adjusted along with the likelihood parameters.

The columns of table 3.6 can be interpreted as follows: the "%Instrument" heading represents the proportion of the ground truth instrument regions from figure 2.10 that were returned within the final segmentation. "% Correct" corresponds to the proportion of the final segmentation that includes ground-truth instrument regions. The "% Adjacent" column returns the proportion of the final segmentation that included background that was *directly* connected to correctly segmented instrument. In other words, it reflects how much of the segmentation "leaked" over real instrument boundaries. Finally, "%Isolated" is the amount of the final segmentation that included completely distinct background content. In figure 3.5 above, this term would encode the background regions shown in the images of the first column. The "%Correct," "% Adjacent", and "%Isolated" values must sum to 100.

Table 3.6: Quantitative results for varying skeleton extraction parameters. These results correspond to test image 3. See text for interpretation of column headings.

l_{min}	σ_{max}	l_{max}	σ_{min}	*%Instrument*	*%Correct*	*%Adjacent*	*%Isolated*
40	40	80	2	56	75	3	22
40	40	80	10	56	75	3	22
40	40	80	18	56	75	3	22
40	40	80	24	38	98	2	0
40	40	60	24	56	75	3	22
40	40	70	24	56	75	3	22
40	40	80	24	38	98	2	0
40	40	90	24	38	98	2	0
10	10	80	2	56	75	3	22
10	70	80	2	53	96	4	0
70	10	80	2	56	75	3	22
70	70	80	2	56	96	4	0

Precision

Before discussing the parameter values suggested by these results, it is interesting to focus on the interpretation of the "%Adjacent" values. Specifically, the low values of this column—regardless the parameter settings—support the expected precision of the edge-based line-object detection scheme. In the Methodology chapter, two of the deficiencies of earlier line-object segmentation methods for endoscope images were revealed to be over-segmentation, rooted in poor contrast and coincidental textures, and compromised precision at higher scales, arising from Gaussian blurring. As illustrated in the Methodology discussion (see figure 2.4) these problems run the risk of positioning boundaries of real instrument regions within surrounding background tissues, and will cause an increase in the "%Adjacent" measure. Since the Bayesian-based line-object scheme relies on precise edge content, this kind of problem becomes much less a concern. In fact, the low proportions quoted in the table are more likely artifacts of the imprecision of the hand-segmented ground-truth references than they are errors of the segmentation method itself.

Parameter values

Given the need for as many instrument objects as possible—but instrument objects only —to accurately achieve the calibration objective of the system, parameter values should be set with the goal of maximizing the "%Instrument" measure, while minimizing the amount of segmented background (i.e. "%Adjacent + %Isolated."). In general, the likelihood parameters had more impact on the "%Instrument" and "%Isolated" measures.

From the results of table 3.7, the maximum length and low scale parameters appear to have the most most influence on the results. Too-low values for either of these parameters tend to allow a lot of isolated background content into the results. This is expected since these values are used to control the likelihood function shape through the intermediate scale and length ranges. Lower values will move the 0.5 level closer towards the 0-length, 0-scale origin, and will therefore assign more skeletons higher likelihoods.

Further, consistent with earlier observations that higher scales will contain more

dominant instrument content, the values of the high scale and minimum length parameters are less influential—provided they are set high enough to maintain confidence in the instruments-at-high-scales assumption. From this analysis then, appropriate values for the likelihood parameters seem to be $\sigma_h = 40$, $l_{\min} = 40$ pixels, $\sigma_l = 24$, $l_{\max} = 80$ pixels. However, it is dangerous to make these conclusions too hastily.

The reason for the continued emphasis on the fact that all probability parameters were investigated in unison becomes clear when the results shown in figure 3.6 are analyzed. The images of the first row represent results obtained with the "optimal" likelihood parameters quoted above and different values of the probability threshold. In the second row, only the σ_l likelihood parameter has been changed, set to 2 instead of 24. The second image of the second row corresponds to the first row of table 3.6, and the second image of the first row corresponds to the fourth row of the table. These correspondences support the use of the higher parameter value: even though region is lost, no background content is retained.

Yet because the results in table 3.6 were obtained with fixed values for the probability and skeleton extraction parameters, subtleties of the relationships between all the parameters are lost. In particular, when the probability threshold is also adjusted and only slightly increased from 0.1 to 0.121 (second and third columns of figure 3.6), the results obtained with $\sigma_l = 2$ become superior to those with $\sigma_l = 24$. This lower value of the small scale parameter proved to be the best across all the parameter combinations tested.

This case of figure 3.6 is illustrative of the limitations of current results. The brute force scheme to evaluate the subtleties of the parameter interactions, while set-up to be comprehensive, may have missed certain combinations that could produce better results. Perhaps a more analytic scheme can be found to optimize the parameter choices, but this will have to be the subject of future work.

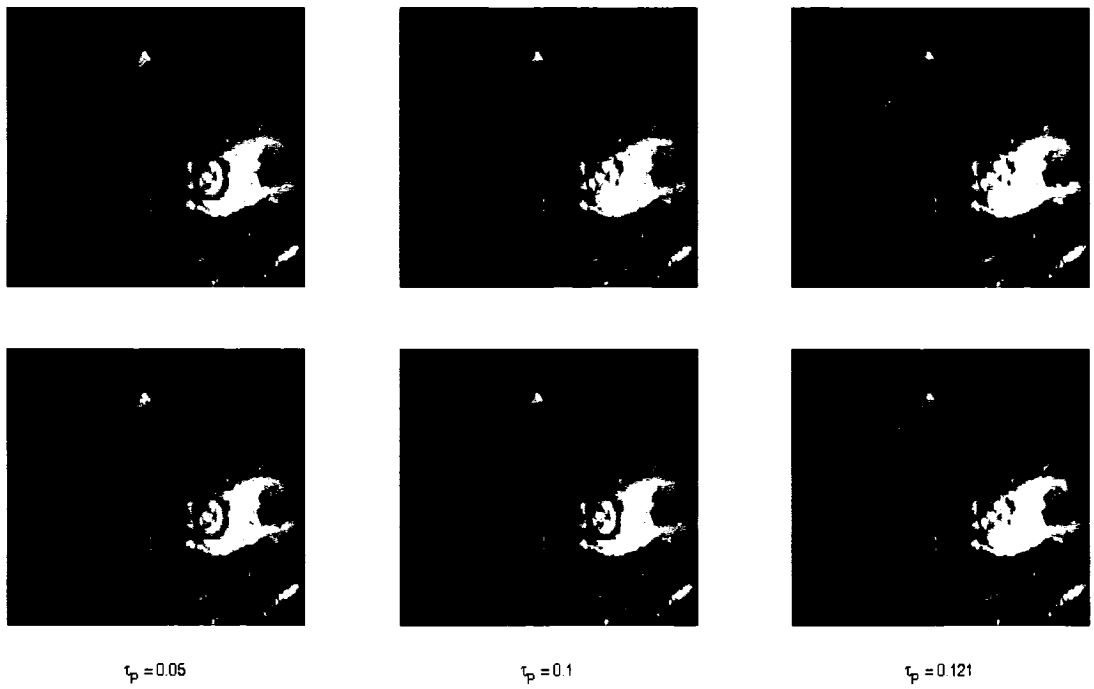


Figure 3.6: Impact of a too-strong low-scale likelihood parameter. First row: The σ_{min} likelihood parameter has been set at 24; Second row: the parameter has been set to 2. Columns represent increasing probability threshold, from 0.06 to 0.1 to 0.121. While results in table 3.6 suggest setting σ_l = 24, they do not reflect the impact of different probability thresholds. For only a slight increase in this threshold (0.1 to 0.121), the results for σ_l = 2 become superior than those with σ_l =24 (all else fixed).

Probability threshold

To conclude the parameter discussion, the probabilistic threshold applied to the skeletons after processing by the Bayesian functions needs to be set. As has been stated before, to reduce tracking ambiguity for the calibration task, the parameter value should be set to minimize segmented background, even at the expense of loosing instrument content. Since the threshold value needs to be set at a value that respects this trade-off for all images, some images suffer more instrument loss than others. This trade-off is clear in the results of figure 3.7. Test image six (last row) is particularly affected, but the threshold that would differentiate between the lost instrument and background is low enough that applying it would reintroduce the background content of image four (top row).

The final probability threshold established via the brute-force parameter search was 0.121. Although lower than the 0.5 originally suggested, the 0.121 threshold is not reason for concern: as the quantitative results in table 3.7 indicate, this threshold works well for all the test images considered, which was one of the principle objectives of the investigation.

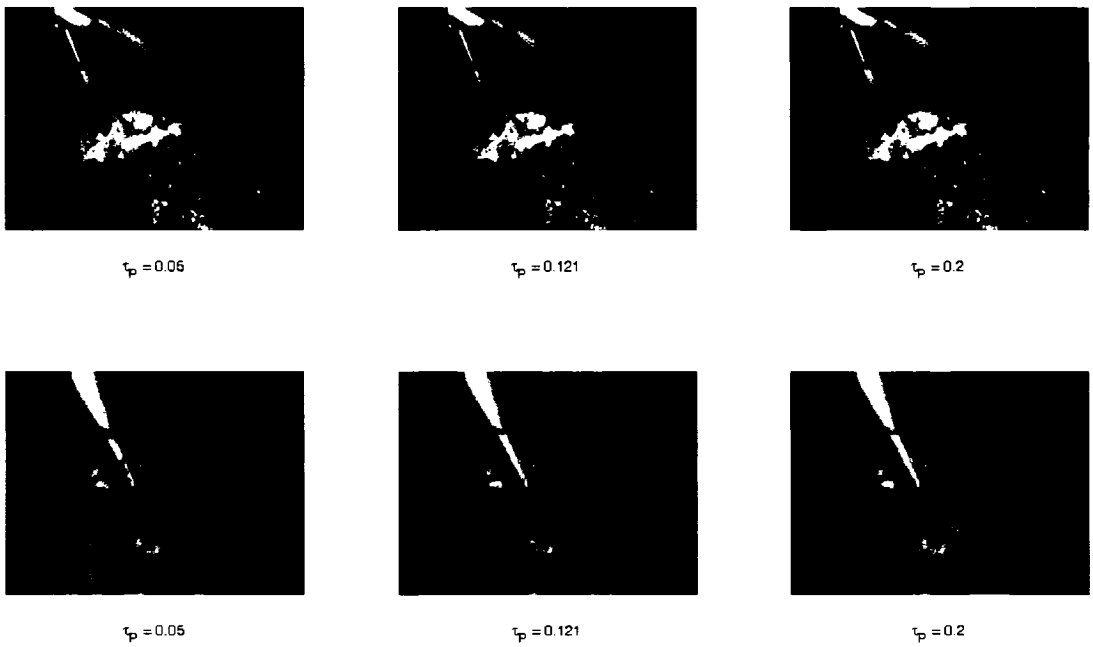


Figure 3.7: Qualitative results for varying probabilistic thresholds. First row: test image 4; Second row: test image 6. A threshold of 0.121 (second column) offers the best retention of instrument regions while ensuring all background content is removed.

Table 3.7: Quantitative results for varying probabilistic thresholds. A threshold of 0.121 gives the best instrument segmentations, subject to the constraint of minimizing segmented background. Graphic results for image 4 and 6 are in figure 3.7.

Image	*Prob. threshold*	*%Instrument*	*%Correct*	*%Adjacent*	*%Isolated*
1	0.050	0.72	0.42	0.01	0.56
1	0.121	0.72	0.97	0.03	0.00
1	0.200	0.72	0.97	0.03	0.00
2	0.050	0.63	0.48	0.52	0.00
2	0.121	0.56	0.46	0.54	0.00
2	0.200	0.44	0.39	0.61	0.00
3	0.050	0.48	0.57	0.03	0.40
3	0.121	0.41	0.95	0.05	0.00
3	0.200	0.08	0.96	0.04	0.00
4	0.050	0.44	0.85	0.01	0.14
4	0.121	0.36	0.98	0.02	0.00
4	0.200	0.26	0.98	0.02	0.00
5	0.050	0.95	0.82	0.05	0.14
5	0.121	0.95	0.95	0.05	0.00
5	0.200	0.95	0.95	0.05	0.00
6	0.050	0.54	0.71	0.05	0.24
6	0.121	0.38	0.99	0.01	0.00
6	0.200	0.38	0.99	0.01	0.00

3.3 Segmentation: Final results and discussion

The overall results of the combined segmentation steps are now discussed.

3.3.1 Final parameters and results

From the discussions and analysis of the previous two sections, the final set of parameters used to implement the Bayesian line-object segmentation method are:

Table 3.8: Final parameter values for the Bayesian line-object segmentation.

Module	*Task*	*Parameter*	*Description*	*Value*
Skeleton extraction	Edge detection	τ_{low}	Canny low threshold	0.04
		τ_{high}	Canny high threshold	0.206
		σ_{canny}	Canny standard deviation	1.725
	Edge breaking and removal	τ_{κ}	Maximum edge curvature	0.358
		τ_{e}	Minimum edge length	10
	Skeleton extraction	τ_{θ}	Minimum edge parallelism	15
		W_{max}	Maximum separation scale	46
Skeleton selection	Instrument likelihood	l_{min}	Short line-object length	2
		l_{max}	Long line-object length	80
		σ_{l}	Low line-object scale	40
		σ_{h}	High line-object scale	40
	Instrument selection	τ_{B}	Probabilistic threshold	0.121

The final results obtained using these parameter values for each test image are presented in the following figure:

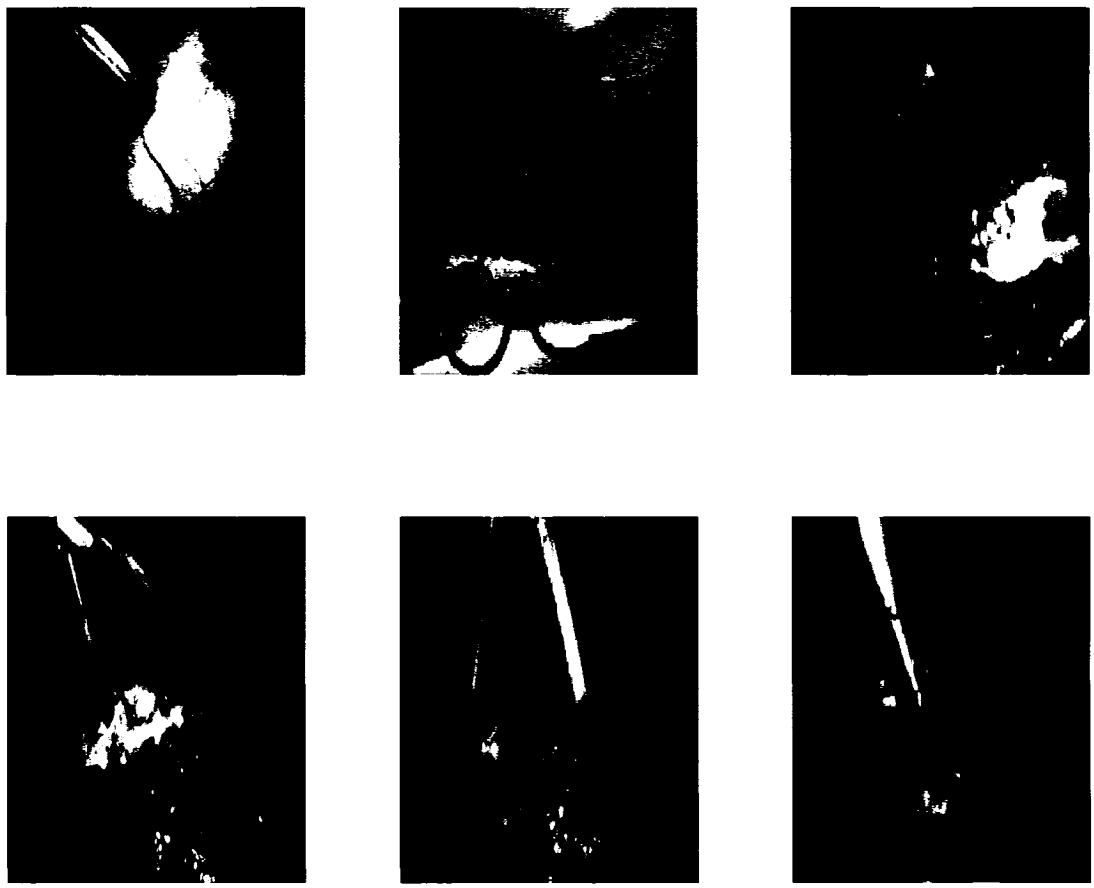


Figure 3.8: Final segmented images for the Bayesian line-object scheme. These results were all obtained with the parameters listed in table 3.8 and, with the sole exception of image 2, reliably segment only instrument content.

3.3.2 Discussion

Overall, the Bayesian line-object segmentation scheme achieved the original objectives very effectively: for all but the second test image, instrument regions were returned without any background content via a scheme that was completely constant for each case. Certainly the results leave room for improvement, and areas for future work will be discussed shortly. Still, considering the complexities of endoscope images, from complicated backgrounds to dramatic changes in the appearance of instruments, the method has proved itself to be quite robust.

Highlights

There are a few aspects of the method that have contributed to its robustness and are worth more focused attention[10].

Bayesian formulation

One of the principle advantages of the method is the Bayesian formulation. The overall utility of Bayesian concepts in computer vision problems was demonstrated earlier via the discussions of fundamental regularization concepts that underpin much computer vision work. The utility stems from the overall flexibility offered by the probabilistic terms, which allow myriad constraints to be introduced into a problem in a theoretically rigorous but mathematically flexible way.

This flexibility is essential to facilitate the use of line-object concepts for this segmentation method. The likelihood term nicely accounts for changes in the quality of an instrument's line-like structure due to varying endoscope distance, but cannot be relied upon alone for segmentation; the prior terms are essential to add additional constraint to help distinguish instrument line-objects from real and apparent background ones. The first test image is particularly illustrative of this flexibility. Recall from the Methodology discussion the difficulties presented by the numerous veins. Moreover, the instrument is short and wide and atypical of traditional line-object shape. The likelihood term effectively handles the latter problem, and is easily combined under the Bayesian scheme via straightforward multiplication with prior terms that remove the background line structures.

The flexibility of the Bayesian approach also lends itself well to developing effective prior terms: as long as the idea of *a priori* knowledge is respected, the terms can take any form. This makes it easy to build and integrate adaptive prior terms that are able to embed prior expectations without being limited by a lack of specific knowledge of the problem. In particular, although instruments are expected to be amongst the highest scale line-structures in the image, camera motion makes it difficult to know before beginning the segmentation process just what the definition of "highest scale" will be. A fixed

10 Throughout this discussion, recall from table 2.1 in the Methodology chapter how the various test images were selected to highlight the unique difficulties of endoscope images.

value to account for the extreme cases of scale might be considered, but it will run the risk of suppressing instruments when the camera is far away, as in test image 3 or 4.

Finally, the flexibility of the Bayesian implementation will make it easy to modify or tweak the developed probabilistic functions in future work. For instance, the two current prior terms could be combined into a single prior still based on grouping clusters of skeletons, but with the grouping now conditioned on the underlying eigenvalue response. In other words, grouping would only occur where the skeletons had very similar eigenvalue responses. Such conditional priors are quite possible in the Bayesian framework, and this might be a more natural constraint than the current stand-alone eigenvalue prior.

Line-object concepts

Advantages of line-object concepts for instrument segmentation were already highlighted from a structural perspective. Another significant benefit of the use of line-object concepts is the automatic interpretation of the resulting segmentation. The interpretation is inevitable, and occurs without the need to do any specific initialization. This is a considerable benefit for a surgical system, as it contributes towards a system that will integrate as seamlessly as possible into the workflow of standard procedures.

The automatic interpretation results from the fact that, in principle, line-object segmentation schemes are designed only to respond to target objects. An additional benefit of this property is that it makes segmentation independent of the number and position of target regions in the image. This is particularly advantageous for endoscope image segmentation. Test images 3, 4 and 5 (see figure 3.8, page 103) illustrate the dramatic changes in instrument appearance that are possible due to relative camera movements. In the extreme case of image 5, only the prongs are visible, and these are no longer connected within the image. The line-object method responds to both regions and correctly returns a full instrument segmentation.

Limitations

Despite the promising results and considerable benefits gained from the probabilistic line-object scheme, there are some limitations of the method as it currently stands. The

principle difficulties include a short-coming of the edge-based skeleton extraction approach, and general under-segmentation[11].

Numerical implementation of the edge-based skeleton extraction

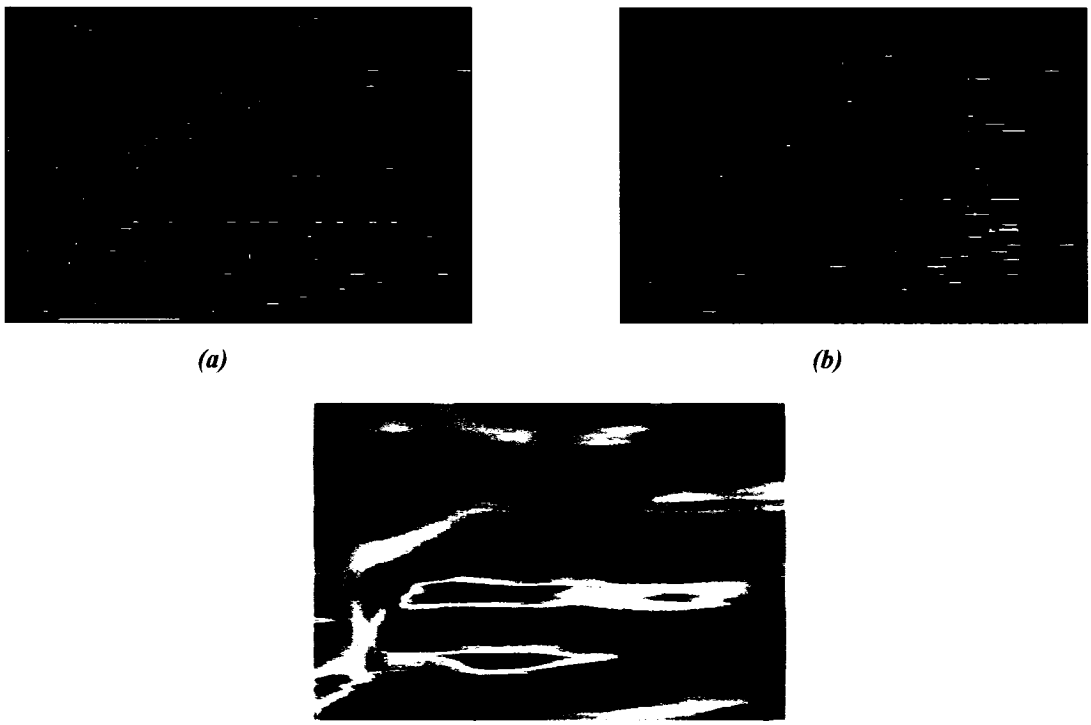
The poor results obtained with image 2 are not a problem of the Bayesian selection scheme so much as they are an artifact of the edge-based line-object extraction. As described earlier, any two non-co-linear lines will generate a skeleton in the image. Some of these skeletons will not correspond to any particular structure, and the maximum separation and parallelism parameters were implemented to remove these pseudo-skeletons. Any remaining pseudo-skeletons should be suppressed by the prior terms: the chance that they will not only fall in a region with high eigenvalue responses, but are also nearby numerous other skeletons, such that they will be grouped into a large aggregate line-object, are remote. However, image 2 demonstrates that all these events can occur simultaneously.

In figure 3.9a, the cropped edge content for this image is shown. There is a patch of background just between the two instrument prongs that generates a series of small edges longer than the minimum length criterion, nearly parallel with the real instrument edges, and close to the instrument itself. These edges generate a series of pseudo-skeletons, as can be seen in 3.9b. The proximity of the skeletons to the real instrument cause them to be grouped with the set of instrument skeletons, producing an aggregate object with a high apparent scale that responds well to the likelihood term since some of the real instrument skeletons are long. Further, the prior term based on the number of skeletons in the aggregate object takes a high value. Finally, by sheer coincidence the background texture in this region also happens to generate a decent line-response (red response in 3.9c), so that the eigenvalue response prior also gets set high. The resulting super-object includes all the instrument and background segmentation, as seen in figure 3.8.

The results for this image would almost certainly be better if the initial edge content

11 It is worthwhile to clarify the interpretation of under- and over- segmentation in the context of this work. One interpretation of under-segmentation implies that the target region in an image is not fully identified. In this context, over-segmentation implies more than the target region is segmented. An alternative interpretation implies that an over-segmented image is one with numerous distinct, isolated regions of segmented target region, regardless how much of the target object is ultimately identified. Throughout the following discussion, the former interpretation is intended.

was processed in a more sophisticated way to extract the initial skeletons. Barring that advance, the implementation of the grouping scheme that assembles the aggregate line objects might be improved – something similar in spirit to the conditioned prior term described in the last section, for example. Regardless the solution, it is important to understand that the scenario that has produced this result is unlikely. Moreover, since it will likely be avoided with improved implementations of the underlying segmentation theory, it should not be considered as too severe an indictment of the overall segmentation method.



(a) *(b)*

(c)

Figure 3.9: Limitations of the edge-based skeleton extraction scheme. Highly parallel background edges between the instrument prongs that are nearby instrument edges (a) generate a series of pseudo-skeletons (b). These skeletons overlap a background region with generally strong eigenvalue responses (c), and are grouped with the real instrument skeletons, leading to the poor segmentation of test image 2 in figure 3.8.

Under-segmentation

A more common problem than over-segmentation arising from the skeleton extraction is

that of under-segmentation. This problem is particularly noticeable in the results for test images 3, 4 and 6. The cause of this under-segmentation is primarily a design choice that has been brought-up previously: since the calibration task will rely on tracking instrument objects, it is better to under-segment the image and only return instrument regions than it is to segment both instrument and background. To this end, the parameters of the method have generally been set to err on under-segmentation.

Still, the calibration task will benefit if more instrument region can be returned, and it is worthwhile to consider how the segmentations can be improved. A promising possibilities involves the contour evolution schemes that led to the development of the line-object method in the first place. Armed now with a strong understanding of the positions and general shape of the instruments in the image, one could return to the idea of model constraints on GAC evolution like that proposed by Leventon et al. [2000] and described earlier. Though a PCA based analysis will still be impractical, a shape based constraint added to the GAC evolution approach in the spirit of (1.15) (see the Literature Review) could be developed and would probably be quite useful.

The basis of such as constraint could easily be rooted in the eigenvalue response information for an image, much of which is under-utilized in the current segmentation scheme. For example, consider the eigenvalue responses for the instruments in test images 4 and 6, which are shown in figure 3.10. Both of these instruments are currently under segmented, but their signed eigenvalue responses (blue and red regions in the figure 3.10 below have opposite signs) provide a powerful indication of where the rest of the instruments are likely to be. These responses were not fully exploited earlier due to precision issues, as well as the clear difficulties that can be anticipated if trying to distinguish the instrument eigenvalue responses from the background ones. However, now that portions of the instrument regions have been segmented, and their shapes are understood, this information could be used to strong advantage. At a minimum, a GAC scheme could be implemented to inflate the initial segmentation according to the underlying eigenvalue response itself, or according to original intensity content but with the eigenvalue response data a constraint to resolve ambiguous gradient information. The initial shape would also act as a constraint – penalties on any large deviations from the

proportions of the original segmentation would prevent the evolving contour from leaking into surrounding areas of high response.

Figure 3.10: Eigenvalue responses underlying endoscope images. Left: response to image 3; right: response to image 6. The scale is 12 in both cases. Dark blue and red regions represent areas of strong response (opposite sign) while greens and yellows are areas of low response. This information could be used to improve the initial segmentations, perhaps by integrating it into a GAC evolution scheme.

3.4 Tracking: Preliminary results and discussion

Remember that the proposed tracking scheme is identical to the Bayesian optical flow approach developed by Mansouri (1.27) except that eigenvalue responses like those in figure 3.10 have been substituted directly for intensity information during calculation of the level set propagation speed F. Since this is the only change, the results of the tracking algorithms will provide an easy to investigate whether the eigenvalue information holds as much potential for the segmentation and tracking schemes as anticipated.

3.4.1 Comparison with original intensity-based Bayesian scheme

As described in the Methodology chapter, the principle difficulty with Mansouri's original tracking approach is low propagation speed values in areas of reduced contrast: it becomes difficult to classify a pixel in the next image frame as region or background when the surrounding intensity values are washed-out. This is particularly problematic for endoscope images, since blood, steam and reflections often create areas of low

contrast along instrument boundaries, where the demand for good speed estimates is highest if the target is to be tracked. These problems are evident in the first row of figure 3.11, which highlights the intensity based tracking results over the first 12 frames of the video sequence containing the instrument of test image 2. By the sixth frame, the track has already started to loose the lower-right tip of the instrument, and by the 12^{th} frame it has become clear that this portion of the boundary is not being evolved at all.

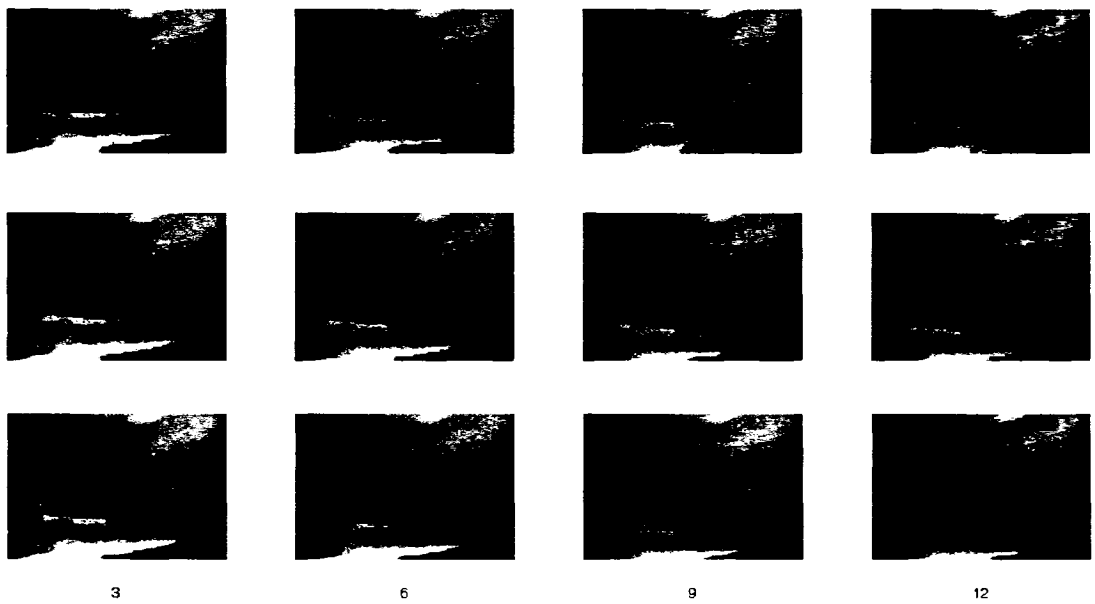


Figure 3.11: Preliminary tracking results. First row: Results obtained with original Bayesian scheme using an intensity-based speed functional. Second row: same method, but using eigenvalue response speed function. Third row: Results with eigenvalue response polarized on interval [-1 +1]. In general, the eigenvalue tracking is comparable, if not better than the intensity-based approach.

The potential advantages of using the eigenvalue line-response results in place of pure intensity information is that, while still intensity based, these responses encode both the contrast characteristics of the image regions in general (the eigenvalue response will be negative or positive depending if the object is brighter or darker than its surroundings), as well as a high-level description of scale. While the strength of eigenvalues will decrease under poor contrast, the nature of these contrast and structural markers should be quite stable. Therefore, even in areas of reduced contrast where response magnitudes are low,

there should exist differences in the signed response that represent well the general structure of the object and produce larger speed estimates than pure intensity information.

These expectations are confirmed by the second and third rows of figure 3.11. In the second row, the tracking results following a direct substitution of eigenvalue responses are shown. Immediately the results improve, as the end of the lower instrument prong is followed more closely. The eigenvalue response shown in figure 3.9c (page 107) is illustrative of those used to achieve these tracking results, and supports the observations about stable sign and structure information. Blue areas in the eigenvalue response have the same sign, with darker values indicating greater magnitude. Clearly, the basic structural and contrast characteristics of the instrument are still available via the eigenvalue response, even though the magnitude of the response is low.

In the third row, the results are obtained using only the sign of the eigenvalue response, and the instrument has been followed even closer. The results are not particularly stable, since the loss of the response magnitude information leaves no way to limit the contour evolution in areas of high uncertainty: even slight fluctuations of the response about 0 will lead to violent shifts in the classification of pixels between frames. Still, the results are very illustrative of the power of the additional markers, since the instrument tracking is at least as good, if not better, than the tracking results obtained with the eigenvalue magnitude information.

The qualitative results just described are confirmed by the quantitative results presented in table 3.9. In this table, the tracking results at numerous intermediate video frames have been analyzed against ground-truth segmentations according to the same measures applied earlier. The two distinct measures of background segmentation have been combined into one though; with the tracking initialized to the ground truth in frame 1 for these results, the distinction between isolated and adjacent background means little.

Table 3.9: Quantitative results for Bayesian-based tracking scheme. "%Background" is the sum of the "%Adjacent" and "%Isolated" measures defined in earlier results.

Basis for F	*Measure*	*Sequence frame*						
		5	*10*	*14*	*15*	*16*	*20*	*25*
Intensity	*%Instrument*	90	87	86	85	80	83	79
	%Correct	89	74	67	60	48	44	33
	%Background	11	26	33	40	52	56	67
Eigenresponse	*%Instrument*	93	90	89	85	78	83	81
	%Correct	90	78	71	61	48	46	35
	%Background	10	22	29	39	52	54	65
Sign of eigenresponse	*%Instrument*	92	91	90	90	84	87	78
	%Correct	90	82	74	65	52	46	34
	%Background	10	18	26	35	48	54	66

3.4.2 Discussion of eigenvalue-based tracking

Perhaps the most interesting element of the tracking results obtained thus far is that nothing more sophisticated than a direct substitution of eigenvalue responses for intensity information has been implemented. This suggests that there is a lot of room for improvement, and some of these avenues are discussed below.

Parameter selection

No effort has been made to optimize any of the tracking parameters yet. For the level set update, the relevant parameters are N, the number of iterations, Δt, the time step, and λ_L, the multiplier which limits the development of sharp corners in the evolving boundary. Values for these results were 250, 0.001, and 10, respectively, but better values can surely be found. The only constraint on these values is that the numerical domain of dependence of the updating contour must be contained within the mathematical domain of dependence. This is a statement of the CFL condition, and demands $\Delta x \leq F\delta t$, where Δx is the update step size, and F is the propagation speed. More information on this constraint is available in the text by Sethian [1999].

The propagation speed also has an important parameter: the search window size δ. This value controls the size of the neighbourhood around each pixel used for intensity or

eigenvalue comparisons when making the region and background classifications that underlie the speed estimate. As long as inter-frame displacements are small, small values for this parameter are desirable, since the speed estimate is less likely to be affected by noisy pixels. In these results, a value of 5 was used. This worked well for the most part, but explains why there is a fairly sharp increase in the amount of segmented background for all cases in table 3.9 between frames 15 and 16. As can be seen in figure 3.12, there is a large movement of the instrument between frames 15 and 16 compared to between frames 14 and 15. None of the methods follow this displacement well since it is larger than 5 pixels.

A larger search window would improve the chances of tracking this movement, but increase the risk of noise corrupting the quality of speed estimates in general for other frames. Therefore, even though a better value for this parameter is needed, the choice is not obvious.

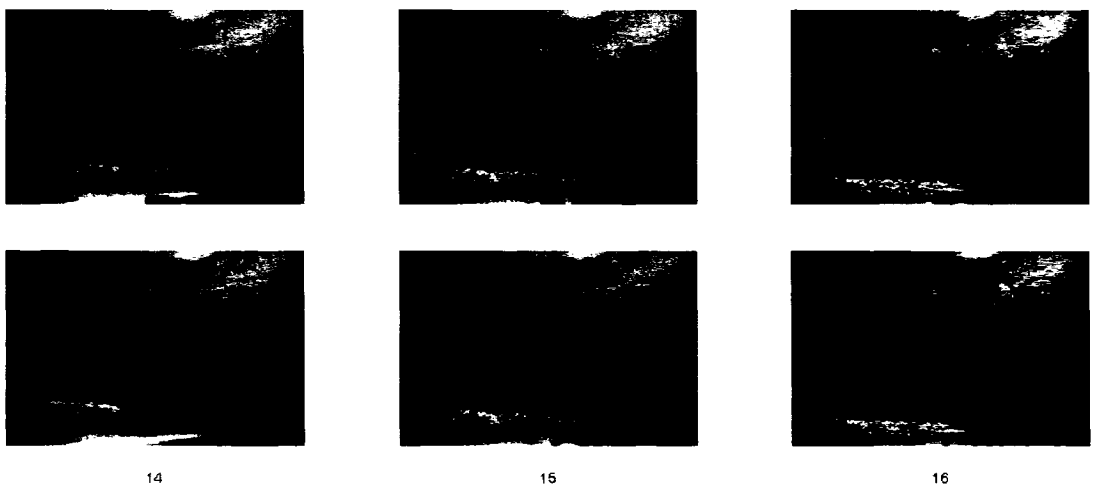


Figure 3.12: Advantages of modified tracking scheme compared to original. Both the intensity-based (top row) and eigenvalue-response methods (bottom) rely on larger search windows to handle inter-frame displacements. The eigenvalue response handles the large displacement between frames 15 and 16 slightly better than the original scheme.

Refined integration of eigenvalue responses

The current integration of the eigenvalue responses could probably be done in a much better way. For one, the overall shape encoded by the sign of the eigenvalue responses could be used to constrain the tracking when the eigenvalue magnitude is low.

Moreover, the shape estimates returned by the segmentation scheme could be useful constraints. Right now, the tracking initializes to a manual segmentation of the first image, but clearly it could be integrated with the results of the segmentation step. This integration may need to wait until a way to refine the initial segmentations has been established, but ultimately the natural connection between the segmentation and tracking steps should make it easy to extract a constraint from one for the other.

3.4.3 Next steps

To conclude the discussion of the tracking algorithm, it is worthwhile to recall the original motivations for pursuing this contour-based scheme. In the first place, if instrument shapes are ultimately used as the object to track for the calibration task, then tracking methods like this will be critically important to extract motion values. On the other hand, if point-like features are used, then these tracking results could be used to follow regions where the point-features are expected to be, or to supplement other motion inter-frame feature motion estimates like that used by the Kalman filter.

The results presented here, while only preliminary, suggest strongly that a method like this will not be accurate enough to use to extract the principle motion estimates, and tracking of point-like features such as corners is likely to be the ultimate solution required. In this context, and as stated already, this tracking scheme might help supplement a point-based correspondence tracking method; however, at this time energy would probably be better focused on either refining the segmentation step, or determining what image objects can be reliably extracted for tracking. In the former case, the results of this tracking investigation are very encouraging in that they show the power of the eigenvalue response. With knowledge of instrument shape and location now available, the eigenvalue response can probably be used very effectively for segmentation refinement.

In the latter case, if appropriate image objects can be identified, then a more applicable tracking scheme can be investigated. According to need based on the results of that tracking, this Bayesian framework could be re-investigated later.

CHAPTER 4 - CONCLUSION

The segmentation and tracking methods developed over the course of this thesis show strong potential. The segmentation algorithm, in particular, works overall well. There are a few shortcomings that must still be addressed before the segmentation algorithm can be used confidently, and much work remains to turn the tracking concepts into a viable methodology. These limitations and directions for future work will be addressed shortly; first, it is worth reviewing the strongest design aspects of both the segmentation and tracking methods.

The principle advantage of both methods is integration of robust and stable line-object concepts into the flexibility of a Bayesian probabilistic framework. This affords the methods, especially the current segmentation algorithm, the following strengths:

(1)Flexibility. As described in the Literature Review, Bayesian regularization affords more flexibility in how prior knowledge and assumptions can be used. Compared to Tikhonov regularization, the mathematical framework is simple, easy to implement, and less restrictive, therefore allowing specific prior constraints to be easily adapted to the needs of a problem. This property can be appreciated by considering the vastly different forms of the segmentation and tracking algorithms, even though both are based on Bayesian theory.

Further contributing to the flexibility of the methods is the use of line-objects. The idea of 'long-and-thin' structure is the only constraint imposed by line-object theory; appropriate definitions for this constraint can be developed according to the particulars of a situation. This flexibility is a crucial element of the likelihood term's ability to adapt to the large changes in instrument appearance caused by motions relative to the endoscope.

(2)Robustness and stability. A further benefit of line-object concepts is that the 'long-and-thin' characteristic encodes shape at a high perceptual level, which contributes to stability under sub-optimal imaging conditions. After blurring or other treatment to eliminate noise, the general, high-level shape of instruments should persist. Further, as shown in the discussion of the tracking results, even under poor contrast, the general nature of instrument shape is available via the eigenvalue response.

The eigenvalue-based line-object response is therefore another important element that contributes to the stability of the method. Not only does it embed important information such as magnitude, contrast and direction, but it is a much more robust indicator of line-object direction than the second derivative alone. Compared to line-object detection methods that search for strong gradients in the direction dictated by the 2nd derivative, embedding the 2nd derivative information into a local eigen-analysis using the Hessian is less sensitive to local imperfections in intensity *etc.*

(3)Automatic. A further contribution of line-object concepts to the efficacy of the segmentation scheme is that detectors rooted in this theory provide an automatic interpretation of the segmentation. When the final segmentation is achieved, target objects are the image regions with a high response, while background has a low response. The automatic interpretation is a crucial design element for the overall goal of a calibration system that minimizes required interactions from the surgical team.

(4)Precision. Even with these numerous advantages, traditional line-object detectors rely only on the line-object response to locate the boundaries of target objects in an image. Under poor contrast, or as Gaussian blurring increases at higher scales, the precision of the detection must necessarily decrease. Therefore, an advantage of the current segmentation framework is the use of line-objects to supplement information gained from a Canny-edge detection scheme. Edge detection to locate the boundaries breaks the dependency of precision on scale, so that the boundaries of large-scale line-objects could be detected with the same precision as low scale objects.

Unfortunately, despite the various advantages afforded by line-objects and Bayesian methods, the current segmentation and tracking algorithms are not without their limitations. These limitations include:

(1)Skeleton extraction. Although the edge-based line-object detection scheme offers the notable advantage of precision, the actual extraction of these skeletons is not as intuitive as with pure line-object detectors. This short-coming was evidenced by the need to introduce the parallelism and maximum scale parameters to limit the number of possible corresponding edges, even after the edges had been preprocessed by length and curvature to eliminate those that clearly did not represent good instrument content.

(2)Parameter selection. The segmentation scheme is also limited in its dependence on numerous parameters. The need for more parameters than previous line-object segmentation methods is expected, since there is an additional Bayesian detection step; still, the introduction of any parameters always limits the applicability of a method and makes it more cumbersome. Every effort to optimize these parameters has been made: the curvature, minimum length, and likelihood parameters should all be robust for a wide variety of endoscope images. Unfortunately, the segmentation does remain sensitive to the maximum scale, parallelism, and probability threshold values.

The tracking algorithm also depends on numerous parameters, although their influence on results has not been well-considered to date.

(3)Loss of track. A more significant problem than parameters for the current tracking method is the difficulty it has following the target in regions of poor contrast. Certainly the use of the eigenvalue response has helped, but overall, tests thus far suggest that the loss of track is not a problem that will be solved by parameter tuning alone, and a more fundamental investigation of the method will be demanded.

All told, the advantages and limitations just outlined have several implications on future directions for research into the segmentation and tracking methods. To build efficiently upon the results so far the most promising avenues for development include:

- Retaining the probabilistic implementations. The flexibility and adaptability of Bayesian schemes will prove an invaluable tool for the improvement of current constraints and, if necessary, the integration of new ones.
- Preprocessing images. The segmentation in particular is undoubtedly affected by the difficult characteristics of endoscope image backgrounds. Specular reflections, for example, will give rise to image edges and can generate decent eigenvalue responses.

If images can be preprocessed to eliminate some of these difficulties then the segmentation results should see improvement. Research into the removal of specular reflections has been conducted by another researcher in the LIV4D lab, but the results remain to be integrated with the work of this thesis.

- Refining the segmentation. As discussed in the Results & Discussion chapter, one option would be to investigate alternative forms for the probabilistic terms, such as replacing the current stand-alone eigenvalue-response and skeleton-cluster priors with a single prior term conditioned on the eigenvalue response. Another option involves treating the current segmentation results as initial estimates of instrument shape and position that can constrain a subsequent contour evolution step rooted in, for instance, GACs.
- Improving the skeleton extraction. An improved way to identify skeletons from the curvature-processed edge content could eliminate the need for the parallelism and maximum scale parameters.
- Better use of the eigenvalue response. This response encodes a great deal of information that is not fully used in the current methods. Not only might it be better applied to probabilistic priors as just mentioned, but some information could be reintroduced into the skeleton extraction step. An intelligent integration should be able to retain the precision of the edge-based detection and regain the intuition of other line-object detectors.
- An improved numerical implementation. The current implementation is by no means optimized. Not only will a better implementation improve the processing of the algorithms, which in turn will make it easier to meet overall speed requirements for the calibration system, but there may be better numerical implementations for some of the theoretical aspects of the algorithms that will improve results.
- Object investigation for tracking. The object type that will ultimately be tracked for the calibration task needs to be determined. From the preliminary investigation in this thesis, the overall object shape does not appear to be stable enough for the purposes of calibration, although more tests or better tracking methods may reveal otherwise.

Intuitively though, interior instrument contours and points (such as corners) will likely be more appropriate. Like general instrument shape, corners exist at a high perceptual level, and are generally quite stable under strong blurring. Accordingly, they may integrate naturally into the multi-scale nature of the segmentation method: at a high scale, begin by detecting corners. These are more likely to represent real object features than noise. Then track these corners through to smaller scales to refine their location. This detection theory is rooted in the work of Marr [1982], and has been developed in detail by Mokhatarian [1998].

Identifying appropriate objects for tracking is recommended before pursuing the current tracking algorithm any further. With the object type established, other tracking methods that are more likely to be directly applicable to the calibration task—correspondence methods like Kalman filtering—can be evaluated. If there appears to be a need for additional constraint for these methods, then the investigation can return to contour trackers like this one to see if the current problems can be overcome, and if they will serve as effective input.

While many areas for development and improvement clearly exist, the current algorithms, especially the segmentation ones, show a great deal of promise. Current results are overall quite good, and offer many avenues for future development with strong potential.

REFERENCES

[1] L. Alvarez and L. Mazorra, "Signal and image restoration using shock filters and anisotropic diffusion," *SIAM Journal on Numerical Analysis*, **31**(2), pp.590-605, 1994.

[2] F. Attneave, "Some informational aspects of visual perception," *Psychological Review*, **61**, pp.183-193, 1954.

[3] A. Blake and A. Zisserman, *Visual Reconstruction*, MIT Press, Cambridge, 1987.

[4] M. Bertero, T. A. Poggio and V. Torre, "Ill-posed problems in early vision," *Proc. IEEE*, **76**(8), pp. 869-889, 1988.

[5] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, **8**, pp. 679--698, 1986.

[6] V. Caselles, R. Kimmel and G. Sapiro, "Geodesic active contours," *Int'l Journal of Computer Vision*, **22**(1), pp.61-79, 1997.

[7] J. Chen and A. Amini, "Quantifying 3-D vascular structures in MRA images using hybrid PDE and geometric deformable models," *IEEE Trans Medical Imaging*, **23** (10), 1251-1262, 2004.

[8] D. Comaniciu, V. Ramesh and P. Meer, "Real-time tracking of non-rigid objects using mean shift," *IEEE Proc. Conf. Computer Vision and Pattern Recognition (CVPR'00)*, **2**, pp. 142-149, 2000.

[9] L. da F. Costa and R.M. Cesar Jr., *Shape Analysis and Classification: Theory and Practice*, CRC PressLLC, Boca Raton, 2001.

[10] P. Kovesi, "Phase Congruency Detects Corners and Edges," *The Australian Pattern Recognition Society Conference: DICTA 2003,* Sydney, pp. 309-318, 2003.

[11] M. R. Cox, J. L. McCall, J. Toouli, R. T. Padbury, T. G. Wilson, D. A. Wattchow, and M. Langcake. Prospective randomized comparison of open versus laparoscopic appendectomy in men. *World Journal of Surgery*, **20**(3) :263–266, 1996.

[12] J. Chen and A. A. Amini, "Quantifying 3-D vascular structures in MRA images using hybrid PDE and geometric deformable models," *IEEE Trans. Medical Imaging*, **23**(10), pp. 1251-1262, 2004.

[13] Y. Deng and S.S. Manjunath, "Unsupervised segmentation of color-texture regions in images and video," *IEEE Trans. Pattern Analysis and Machine Intelligence*, **23** (8), pp.800-810, 2001.

[14] R.O. Duda, P.E. Hart and D.G. Stork, *Pattern Classification (2nd Edition)*, John Wiley & Sons, Inc., New York, 2001.

[15] R. T. Frankot and R. Chellappa, "Integrability in shape from shading algorithms," *IEEE Trans. Pattern Analysis and Machine Intelligence,* **10**(4), pp. 439-451, 1988.

[16] S. Geman and D. Geman, "Stochastic relaxation, Gibb's distributions, and the

Bayesian restoration of images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, **6**(6), pp. 721-741, 1984.

[17] General Electric. Instatrack. http://www.gehealthcare.com/usen/xr/surgery/products/-nav.html, 2004.

[18] U. Guller, S. Hervey, H. Purves, L. H. Muhlbaier, E. D. Peterson, S. Eubanks, and R. Pietrobon. "Laparoscopic versus open appendectomy : outcomes comparison based on a large administrative database," *Annals of Surgery,* **239**(1) :43–52, 2004.

[19] J. Hadamard, "Lectures on the cauchy problem in linear partial differential equations," Yale University Press, New Haven, 1923.

[20] C. Harris and M. Stephens, "A combined corner and edge detector," *Proc. 4th Alvey Vision Conference*, Manchester, pp. 147-151, 1988.

[21] E. C. Hildreth. "The computation of the velocity field," *Proc. Royal Society of London B*, **221**, 1984.

[22] B.K.P. Horn and B.G. Shunck, "Determining optical flow," *Artificial Intelligence*, **17**, pp.185-203, 1981.

[23] M. Kass, A. Witkin and D. Tersopoulos, "Snakes: Active contour models," *Int'l Journal of Computer Vision*, **1**, pp.321-331, 1987.

[24] T. Koller, G. Gerig, G. Szekely and D. Dettwiler, "Multiscale detection of curvilinear structures in 2-D and 3-D image data," *IEEE Proc. 5th Int'l. Conf. Computer Vision*, 864-869, 1995.

[25] Y. Kotani, K. Abumi, M. Ito, and A. Minami, "Improved accuracy of computer-assisted cervical pedicle screw insertion," *J. Neurosurgery*, **99**, pp.257–263, 2003.

[26] M. Leventon, W. Grimson, O. Faugeras, "Statistical shape influence in geodesic active contours," *IEEE Proc. Conf. Computer Vision and Pattern Recognition (CVPR'00)*, **1**, pp. 316-323, 2000.

[27] K. H. Long, M. P. Bannon, S. P. Zietlow, E. R. Helgeson, W. S. Harmsen, C. D. Smith, D. M. Ilstrup, Y. Baerga-Varela, and M. G. Sarr. "A prospective randomized comparison of laparoscopic appendectomy with open appendectomy : Clinical and economic analyses," *Surgery*, **129**(4) :390–400, 2001.

[28] C. Lorenz, I.-C. Carlsen, T. Buzug, C. Fassnacht and J. Weese, "Multi-scale line segmentation with automatic estimation of width, contrast and tangential direction in 2D and 3D medical images," *Proc. 1st Int'l Joint Conf. CVRMed and MRCAS (CVRMed/MRCAS '97)*, 233-242, 1997.

[29] M. J. Mack, J. J. Regan, W. P. Bobechko, and T. E. Acuff, "Application of thoracoscopy for diseases of the spine," *Ann. Thoracoscopic Surgery*, **56**(3), pp. 736–738, 1993.

[30] D. Marr, *Vision*, W.H. Freeman and Company, 1982.

[31] A.-R. Mansouri, "Region tracking via level set PDEs without motion computation,"

IEEE Trans. Pattern Anlaysis and Machine Intelligence, **24**(7), pp. 947-961, 2002.

[32] J. Marroquin, S. Mitter and T. Poggio, Probabilistic solutions of il-posed problems in computational vision," *Journal of the American Statistical Association*, **82**(397), pp.76-89, 1987.

[33] Mokhtarian, F. and R. Suomela, "Robust image corner detection through curvature scale space," *IEEE Trans. Pattern Analysis and Machine Intelligence,* **20**(12), pp. 1376-1381, 1998.

[34] F. Mourgues. *Guidage par réalité augmentée : application à la chirurgie cardiaque robotisée*. PhD thesis, Université de Nice-Sophia Antipolis, 2003.

[35] W.L. Nowinski, Guo Liang Yang, and Tseng Tsai Yeo. Computer-aided stereotactic functional neurosurgery enhanced by the use of the multiple brain atlas database. *IEEE Transactions on Medical Imaging*, **19**(1), pp. 62–69, 2000.

[36] T.G. Obenchain, "Laparoscopic lumbar discectomy : case report," *Journal of Laparo.-Endoscopic Surgery*, **1**(3) :145–159, 1991.

[37] ORTHOsoft. Navitrack. http ://www.orthosoft.ca, 2005.

[38] S. Osher and J. Sethian, "Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations," *Journal of Computational Physics,* **79**, pp.12-49, 1988.

[39] T. Poggio and V. Torre, "Ill-posed problems and regularization analysis in early vision," *Proc. AARPA Image Understanding Workshop*, pp. 257-263, 1984.

[40] J.A. Sethian, *Level Set Methods and Fast-Marching Methods: Evolving Interfaces in Computation Geometry, Fluid Mechanics, Computer Vision, and Material Science*, Cambridge University Press, 1999.

[41] R. Shahidi, M.R. Bax, Jr. Maurer, C.R., J.A. Johnson, E.P. Wilkinson, Bai Wang, J.B. West, M.J. Citardi, K.H. Manwaring, and R. Khadem. "Implementation, calibration and accuracy testing of an image-enhanced endoscopy system," *IEEE Trans. Medical Imaging,* **21**(12), pp.1524–1535, 2002.

[42] L.G. Shapiro and G.C. Stockman, *Computer Vision*, Prentice Hall, New Jersey, 2001.

[43] A.N. Tikhonov and V.Y. Arsenin, *Solution of Ill-Posed Problems,* John Wiley, New York, 1977.

[44] D. Rosenthal, R. Rosenthal, and A. De Simone, "Removal of a protruded thoracic disc using microsurgical endoscopy. a new technique," *Spine*, **19**(9) :1087–1091, 1994.

[45] H. C. Sagi, R. Manos, R. Benz, N. R. Ordway, and P. J. Connolly, "Electromagnetic field based image-guided spine surgery part one : results of a cadaveric study evaluating lumbar pedicle screw placement," *Spine*, **28**, pp. 2013–2018, 2003a.

[46] H. C. Sagi, R. Manos, S. C. Park, R. Von Jako, N. R. Ordway, and P. J. Connolly. Electromagnetic field-based image-guided spine surgery part two : results of a cadaveric study evaluating thoracic pedicle screw placement. *Spine*, **28**, pp. E351–4, 2003b.

[47] Siemens. Surgical navigation. http://energized.esiemenshealthcare.com/surgicalimaging/ srgnav.htm, 2005.

[48] J.S. Suri, K. Liu, S. Singh, S.N. Laxminarayan, X. Zeng, and L. Reden, "Shape recovery algorithms using level sets in 2-D/3-D medical imagery: A state-of-the-art review," *IEEE Trans. Information Technology in Biomedicine*, **6**(1), pp.8-28, 2002.

[49] E.A. Wan and R. van der Merwe, "The unscented kalman filter for nonlinear estimation," *Proc. IEEE Symposium 2000 (AS-SPCC)*, Lake-Louise, pp. 153-158, 2000.

[50] L. Windisch, F. Cheriet and G. Grimard, "Bayesian differentiation of multi-scale line-structures for model-free instrument segmentation in thoracoscopic images," *Image Anlaysis and Recognition, Conference Procedings ICIAR 2005*, Springer-Verlag LNCS 3656, pp. 938-948, 2005.

[51] A.P. Witkin, "Scale-Space Filtering," *Proc. 8th Int'l Joint Conf. Artificial Intelligence*, pp.1019-1022, 1983.

[52] C. Xu and J. L. Prince, Gradient vector flow: A new external force for Snakes," *IEEE Proc. Conf. Computer Vision and Pattern Recognition (CVPR'97)*, pp. 66-71, 1997.

[53] A. Yilmaz and M. Shah, "Contour based object tracking using level sets," *Proc. Asian Conf. on Computer Vision (ACCV'04)*, pp. 1-7, 2004.

APPENDICES

APPENDIX A: CONFERENCE PAPER

BAYESIAN DIFFERENTIATION OF MULTI-SCALE LINE-STRUCTURES FOR MODEL-FREE INSTRUMENT SEGMENTATION IN THORACOSCOPIC IMAGES

Image Anlaysis and Recognition, Conference Procedings ICIAR 2005, LNCS 3656, pp. 938-948, 2005.

Bayesian Differentiation of Multi-scale Line-Structures for Model-free Instrument Segmentation in Thoracoscopic Images

Luke Windisch[1], Farida Cheriet[1,2], Guy Grimard[2]

[1] Ecole Polytechnique de Montreal. C.P. 6079. 53851 succ. Centre-ville.
Montreal. Canada H3C 3A7
{ first.last }@polytml.ca

[2] Department of Orthopaedics. Sainte Justine Hospital
Montreal. Canada
guy_grimard@ssss.gouv.qc.ca

Abstract. A reliable method to segment instruments in endoscope images is required as part of an enhanced reality system for minimally invasive surgery of the spine. Numerous characteristics of these images make typical intensity or model constraints for segmentation impractical. Rather, line-structure concepts are used to exploit the high length-to-diameter ratio expected of surgical instruments. A Bayesian selection scheme is proposed, and is shown to reliably differentiate these target objects from other line-like background structures.

Introduction

The risks associated with exposing the interior of a patient's body during spinal surgery can be mitigated by minimally invasive techniques, under which a surgeon inserts instruments through a number of small incisions and views the surgical site via video from an endoscope, a monocular camera with integrated light source (fig. 1). Compared to conventional surgeries that fully expose the surgical site, patient risk and costly in-hospital recovery periods are reduced.

Unfortunately, adoption of these procedures has been slowed by the considerable difficulties faced by surgeons, who lose depth perception since the surgical site is viewed indirectly via the monocular endoscope. Moreover, contextual depth cues are few, as the proximity of the endoscope to the objects being imaged creates a narrow field-of-view (FOV), on order of 5-cm or less. Ultimately, the long training needed to gain comfort and expertise working with such a difficult view of the surgical site has limited the spread of this approach.

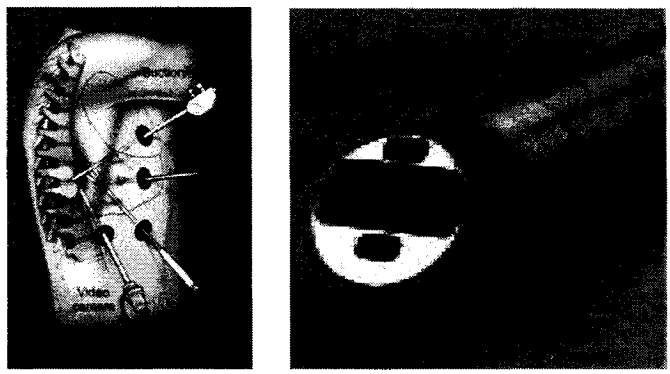


Fig. 1. Left—A typical set-up for a minimally invasive surgery (source: www.spineuniverse.com): Right—An endoscope tip, showing the integrated lens and light source (source: www.intuitivesurgical.com).

An enhanced reality surgical system to integrate 3D preoperative patient models with the endoscope video stream would help solve the problems of depth and context loss. Central to the success of this system is automatic recalibration of the extrinsic and intrinsic endoscope parameters, which may vary throughout the surgery due to camera motion and manual focusing. The extrinsic parameters of the endoscope are updated in real time by using an optical tracking system. However, updating the intrinsic parameters requires tracking features that can be easily and automatically detected from the image sequence. It is difficult to determine whether changes in features such as intensity arise due to camera or instrument movements, thus we only consider corners and others instrument features that represent precise 3D points on instruments. Relative camera and instrument movements reported by the optical tracking system can then be used to give real-world context to the image-based movements of the identified features so that the intrinsic camera parameters can be updated. Actually identifying these instrument features is complicated by several difficulties, the first of which are the characteristics of endoscope images themselves (see fig. 2):

i. Rapid intensity fall-off with radial distance from the image center;
ii. Strong specular reflections from metallic instruments and moist tissues, which change unpredictably under endoscope and instrument movements;
iii. Large changes in the appearance of objects due to motion and focusing of the endoscope;
iv. Blood, exposed tissue, and steam at the surgical site that blur and occlude boundaries; and
v. Cut or exposed tissues, which cause strong variations in background textures and colours.

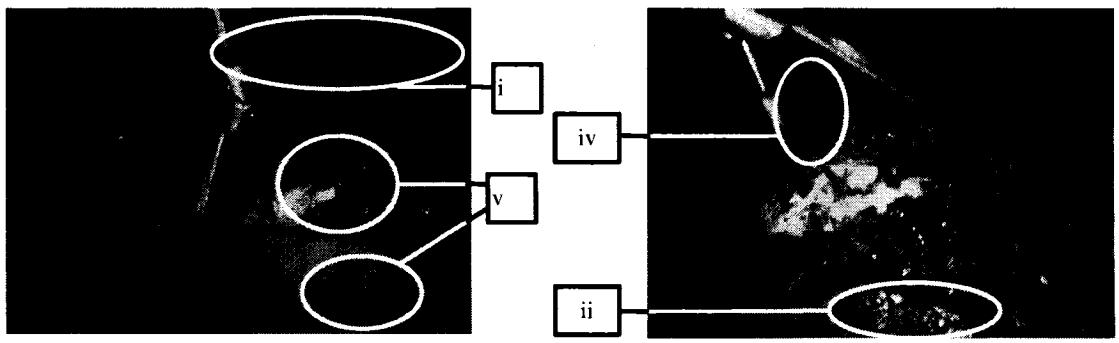


Fig. 2. Two images of the same instrument. Numbered areas correspond to the list of problems above. Note that due to endoscope movement, only the prongs of the instrument are visible in the second image.

A second complication arises from smooth instrument surfaces, which offer only a sparse set of features that are difficult to isolate from background features – a stringent detection criterion to suppress background may also too strongly prune the instrument features. Zhang and Payandeh solved this problem by affixing artificial markers to the instruments [1]; however, such features are not immune to noise, occlusion, poor lighting, or wash-out from specular reflections. More importantly, artificial markers complicate the sterilization of instruments, and we would like to move towards a tracking method that operates with instruments "as is." Accordingly, the preferred approach is to use as large a set of natural instrument features as possible. To do so, background features must be distinguished from instrument features, and this demands a viable scheme to segment instruments from the image.

Segmentation algorithms are typically constrained by metrics that are intensity-based [2, 3], region-based [4, 5], or object-model based [6]. Unfortunately, many of the image characteristics outlined above make such constraints unreliable or impractical for our context. Instead, motivated by recent work by Chen and Amini [7], we have approached segmentation using the concept of line-structures, which exhibit a significant length-to-width ratio, a property common to instruments for minimally invasive surgery.

In Section 2 we present a novel multi-scale, line-structure-based scheme to effectively distinguish between line-responses arising from instruments and from background objects. Results are presented in section 3. The discussion of section 4 summarizes the advantages of the segmentation scheme.

Multi-scale Line-Structure Segmentation

A line-structure has an extended length compared to its width (a high 'length-to-width ratio'), and the intensity profile of the image of such an object will show only slight variation along, relative to across, its longitudinal direction. Numerous line-detector operators have been developed to exploit this property [7, 8, 9]. Often, the scales and orientations of line-structures in an image are unknown, requiring the detection operator to be applied at several scales and orientations to identify the best line object at a point $\mathbf{x} = (x_1, x_2)$.

2.1 Line-Structure Detection with the Hessian Matrix

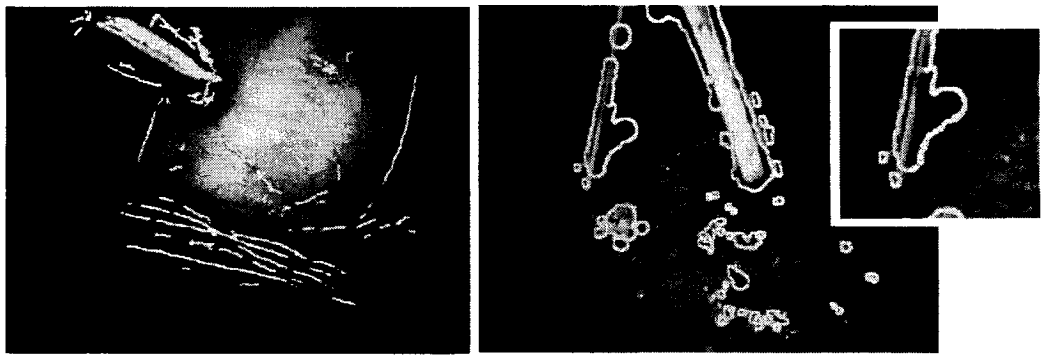
A popular detector, proposed by Lorenz et al. [9], is based on a Hessian matrix analysis. The characteristics of line-object intensity profiles imply that the second derivative across a line-structure should be large compared to along it. Accordingly, an eigen-analysis of the Hessian matrix $H_{\mathbf{x}}$ can be used to identify line-structures by looking for points with $|\lambda_1| >> |\lambda_2|$ and having a high line response given by:

$$R(\sigma, \mathbf{x})_{2D} = \sigma^{\gamma} \cdot \lambda_1(\sigma, \mathbf{x}) \quad \textbf{(1)}$$

where the scale, σ, is the standard deviation of a smoothing Gaussian kernel, $\gamma > 0$, and σ^{γ} compensates for the decrease in the response due to increased smoothing at higher scales. Detection proceeds in three steps: First, an image is blurred with a Gaussian kernel at several scales. Then, each image point **x** is assigned the extremum value of (1) across all scales, as well as the scale at which the extremum occurred. Finally non-maximum suppression of the set of responses (1) generates skeletons centered along the image's line-objects (note that a second function must be used with (1) to distinguish lines from step edges).

Problems with Current Detection Methods

Detection schemes like this have been shown to work very well for problems like vessel-tree extraction from 2D images and 3D volumes [7, 9], where the objects of interest have good contrast against relatively homogeneous backgrounds. Unfortunately, segmentation is complicated by the complex backgrounds of endoscope images. In certain cases, background objects such as veins are themselves line-objects. Current detection methods implicitly assume that any local extrema arising from a function such as (1) correspond to the objects to be segmented; the possibility of non-target line-structures is not considered, and line-responses arising from these objects cannot be differentiated. This problem is illustrated in fig. 3(a), where the skeletons extracted by the previously described detection scheme based on (1) have been superimposed on the original endoscope image. These skeletons all encode scale information and thus represent the final segmentation of the image. Clearly, the presence of blood vessels is problematic – our objective is to segment only instruments.



(a) (b)

Fig. 3. Current line-object detection methods applied to endoscope images. In (a), skeletons corresponding to the instrument and background objects (veins) cannot be differentiated. In (b), responses to background texture have led to isolated patches of segmented background, and have also resulted in inaccurate representations of instrument shape and size, as seen in the over-segmentation of the of the lower-left instrument prong (enlarged in image inset).

While the results in fig. 3(a) are clearly sub-optimal, the background skeletons at least correspond to real line-objects. In 3(b), the image background does not contain any line-objects, but the highly variable texture still gives rise to extrema in the response given by (1). This introduces two difficulties: First, for the same reasons just mentioned, these extrema cannot be distinguished by current methods, leading to isolated regions of incorrectly segmented background.

Second, skeletons corresponding to actual instruments can be quickly led astray of true object shapes, especially when instrument contrast is weak. Consider the lower left prong of the instrument, which has been enlarged in the image inset. At the extreme lower tip of the prong the instrument exhibits good contrast and the segmentation corresponds well with the actual shape. However, a problem quickly develops as the segmentation follows the prong upwards: the contrast towards the right instrument edge falls, while a significant edge in the background tissue develops. The response (1) to the intensity profile between this tissue and the left instrument edge is stronger than to the profile between the left and right instrument edges, and the non-maximum suppression scheme keeps only the response at the higher scale. Thus the skeleton is led astray of the true instrument shape and encodes too-high scale information, resulting in over-segmentation.

2.2 Modified Scale-Specific Line-Structure Detection

A first step towards successfully applying line-object concepts to endoscope images involves making the initial detection process more robust to the problems of weak instrument contrast in the presence of background textures. To increase robustness, we modify the line-structure detection process by applying non-maximum suppression to the responses (1) obtained at a *specific* scale, rather than to the set of across-scale extrema of (1). Numerous scales are still processed, but we now have a set of skeletons at each scale.

Limiting the extraction to individual scales reflects the observation that most line

objects exhibit a high degree of parallelism between their occluding boundaries, so that the object scale in an image should vary little along its length. By restricting the extraction in this way, textures in the background at higher scales are less likely to lead the skeleton astray of the true object shape.

Two problems remain: combining the skeletons from different scales, and identifying which skeletons correspond to instruments. While previous detection schemes used extracted skeletons as the final segmentation result, the skeletons in our scheme are simply a comprehensive set of *candidate* instruments. We now apply highly general assumptions to develop a Bayesian framework to distinguish which candidate skeletons to keep.

2.3 Bayesian Identification of Instruments

After the extraction step, each skeleton will have a scale and length. This information can be used to identify instrument skeletons by considering that instrumentation used in minimally invasive procedures must be narrow enough (small scale) to fit through the small incisions on a patient's body while long enough to reach the surgical site. We embed these characteristics into a Bayesian expression for the probability that a skeleton corresponds to an instrument, given the skeleton's observed scale σ and length l:

$$P(Inst\,\sigma, l) = \frac{P(\sigma, l\ Inst)P(Inst)}{P(\sigma, l)}. \qquad (2)$$

The left term in the numerator of the right-hand side of (2) represents the likelihood of a skeleton having the observed length and scale, assuming the skeleton corresponds to an instrument. The right term in the numerator reflects any *a piori* information regarding instruments in endoscope images. The denominator can be ignored, leaving only the terms in the numerator to be defined more rigorously.

Likelihood Term

The long-but-thin characteristic of instruments implies that corresponding skeletons should have a high length-to-width ratio, and that the first term in the numerator of the right-hand side of (2) should assign higher probabilities to skeletons accordingly. However, skeletons with high ratios are more likely at smaller scales – as an instrument gets near the camera, its image scale will necessarily increase while its visible length will simultaneously decrease. Therefore, any scheme that assigns probabilities to skeletons by looking for high ratios should allow for lower ratios at higher scales. The following non-linear weighting is applied to each skeleton, which assigns a weight on [0 1) and is plotted in fig. 4:

$$P(\sigma, l\ Inst) = \frac{1}{2} + \frac{1}{\pi}\operatorname{atan}\left\{\alpha_\sigma \left(l - \beta_\sigma\right)\right\} - P(\sigma = 0, l = 0\ Inst). \qquad (3)$$

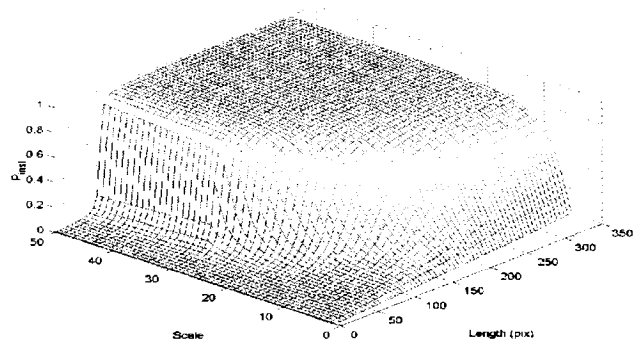


Fig. 4. The likelihood given by (3) plotted against length and scale. For the 320x240 images presented later. the parameters in (4) were: l_{min}=60. σ_{max}=40. l_{max}=240 and σ_{min}=2.

The shape of this functional is controlled as follows:

- β, a function of scale, positions the half-maximum point of (3) at the minimum length a skeleton must achieve before it can be considered a potential instrument. β varies inversely with scale:

$$\beta(\sigma) = \begin{cases} l_{\min} & \text{if} \quad \sigma > \sigma_h \\ \exp\left\{-\dfrac{\sigma - \sigma_l}{\eta}\right\} \cdot l_{\max} & \text{if} \quad \sigma_h \geq \sigma \geq \sigma_l \\ l_{\max} & \text{if} \quad \sigma < \sigma_l \end{cases} \tag{4}$$

  η controls the sharpness of the transition of β from high to low scales, and is specified according to: i) σ_h, the (high) scale at which it is reasonable to assume an instrument may give rise to a skeleton only l_{min} pixels long; and, ii) σ_l, the smallest scale at which skeletons corresponding to instruments are expected, which may be up to l_{max} long. To be as general as possible, these parameters are set once to cover the extreme cases of instrument length and scale that might arise in any endoscope image due to camera motion and focusing. For the results presented later with 320x240 images, l_{min}=60, σ_h=40, l_{max}=240 and σ_l=2 were used. These values set the half-maximum of (3) at a line ratio of only ~1.5 at higher scales, and at 100 or more for smaller scales.

- α determines the slope of the transition through the half-maximum. For most endoscope images, textures and tissues that generate longer skeletons are more common at lower scales. By extension, skeletons that correspond to background textures and exceed the l_{min} threshold in (4) are unlikely at higher scales, so we can allow a sharper transition to the high probability asymptote of (3) as the object scale increases:

$$\alpha(\sigma) = \frac{\sigma}{\beta_\sigma}. \tag{5}$$

- The last term on the right hand side of (3) simply enforces $P(\sigma, l \mid Inst) = 0$ for a zero-length chain.

Prior Term

The likelihood term just described cannot be guaranteed to suppress all non-instrument skeletons: at lower scales especially, long background objects may still exist with a length over the threshold assigned in (4). To increase the probability of isolating only instruments we observe that, regardless of the scale at which instruments appear in an image due to varying distances from the endoscope, they are typically physically larger than other highly line-like background objects such as veins. This characteristic is embedded in the following *a priori* term:

$$P(Inst) = \frac{\sigma_{skel}}{\sigma_{\max}}, \tag{6}$$

where σ_{skel} is the scale of the extracted skeleton, and σ_{max} is the maximum scale of all line objects detected.

2.3 Final Identification

It is important to note that we do not use (2) as a *maximum a posteriori* detector of instruments, since we have no knowledge about the number of instruments that will be visible in an image. Instead, we keep those skeletons in the image that have an *a posteriori* probability greater than 0.5. This value gives us reasonable confidence that the skeletons we retain correspond to instruments: while long line-objects such as veins may receive a high weighting from the likelihood term in (3) they are unlikely to be selected as an instrument unless they also have a large scale relative to other objects.

Results and Discussion

All segmentation results were obtained running non-optimized MATLAB code on a 512-MB P4 IBM PC. Parameters for detection using (2) are given in section 2, and were the same for all images. Segmentation results for the two endoscope images previously considered in fig. 3 are shown in fig. 5(a) and (b).

Segmentation of fig. 5(a) is complicated not only by having to differentiate background line-objects from instruments as previously discussed, but also by the instrument's proximity to the endoscope, which makes it appear short and wide. This problem is addressed by the likelihood term in (3), which is sensitive to the fact that shorter objects are common at higher scales. Removing the background structures is handled by the *a priori* term in (6), which takes a greater value for the instrument due to its higher scale.

Image 5(b) is of the same instrument shown in fig. 3(b). This image illustrates again the utility of the likelihood term in (3), which helps suppress the numerous small, isolated background textures. Further, the scale-specific skeleton extraction described in section 2.2 has helped ensure the final segmentation more closely reflects the true instrument shape.

In 5(c), endoscope and instrument movements have created a dramatically different view of the instrument from 5(b). Changes in appearance like this are common in our application, and would be difficult to handle with current model-based segmentation techniques; a comprehensive set of *a priori* models to guide segmentation is intractable,

even using PCA approaches. Our scheme robustly identifies the target regions without any prior knowledge of the *specific* instrument; only extended length relative to width is assumed.

Finally, in fig. 6 we have compared our method to segmentation results obtained using the Gradient Vector Flow (GVF) approach for Snakes [10]. The images in (a) and (b) serve to underscore some of our introductory comments about difficulties applying existing segmentation metrics to endoscope images (see discussion of fig. 2). In (a), the GVF segmenting contour was initialized to the interior of the instrument region, and eventually stabilized on the boundaries of the strong reflection on the instrument surface. In (b), the contour was initialized to a region outside the instrument, and the evolution was stopped after 150 iterations because the result was unacceptable – the contour had leaked over the real instrument boundary (a part of which has been superimposed in the image for clarity). The problem here is the strong dependence of the contour evolution on intensity/edge content in the image; the real instrument edge is weak, and evolution was only constrained by the apparent edge caused by the reflection. The general shape measures we have embedded in our Bayesian line-object scheme are more stable indicators of instruments in endoscope images, and results of our method applied to the same image are shown in (c). As in (a), the instrument is still under-segmented, but our method is more robust in general. Specifically, as 6(a) and (b) indicate, results with other methods can depend strongly on specifics of the implementation such as the initial location of the segmenting contour. They may also leave unresolved the interpretation of the segmented regions. In contrast, all the results shown using our method were obtained with an identical implementation, and unlike in 6(a), the interpretation of the final segmentation in all cases is automatically provided.

(a) (b) (c)

Fig. 5. Results obtained with our proposed segmentation scheme using parameters given in section 2. Results in (a) and (b) can be directly compared with those in fig. 3. The result in (c), while sub-optimal, illustrates that our method is highly robust to changes in instrument shape – the instrument is the same as that in (b).

(a) (b) (c)

Fig. 6. Segmentation results obtained with a Snake-based contour approach are shown in (a) and (b), where the segmenting contour was initialized inside and outside the instrument, respectively. Segmentation of the same image obtained with our method is shown in (c). The segmentation in (b) has mistakenly included part of the instrument in the background. Moreover, unlike our method, neither of the results in (a) or (b) provide an interpretation of the final segmentation.

At this time, a persistent problem with our method tends to be under-segmentation. This is particularly notable in images 5(c) and 6(c). Work is currently under way to automatically identify and remove the strong reflections on the instrument surfaces and should improve results: with these reflections diminished, the intensity profiles will be closer to ideal, and the response in (1) will be more stable.

Under-segmentation problems also result in part from a deliberate choice we have made. If a we are to rely on natural instrument features for the overall tracking and calibration objective, it is better to under-segment an image than over-segment; we must have confidence that any features identified in the segmented regions correspond to an instrument rather than background. Accordingly, we have set up the Bayesian detection scheme to be strict, and consequently, some skeletons corresponding to instruments are suppressed. For instance, in fig. 5(c), the prior probability term in (6) has caused the loss of a skeleton corresponding to the bright reflection on the left instrument prong since the skeleton encodes a very small scale compared to the upper portion of the instrument.

To conclude these discussions, it is important to reiterate why we do not circumvent the segmentation issue altogether and meet our calibration objective by simply tracking a set of artificial markers attached to the instruments. Natural instrument features such as corners are consistent with an instrument's structure at a much higher-level compared to artificial markers, and tracking these natural, high-level features is expected to create a more robust method overall. Additionally, this approach does not impose any extra procedures on the surgical teams that will ultimately use the application. An intensive validation using a large number of images is currently in progress to evaluate the precision of our approach compared to existing methods.

Conclusions

Endoscope images provide a particularly challenging set of obstacles for segmentation. Camera motion, inconsistent intensity, variable colour information, and uncertain contrast all hamper traditional segmentation approaches: while certain algorithms may perform well with one image, these constantly changing characteristics can render them

ineffective on others. These problems demand segmentation approaches based on more stable aspects of the images. One constant, regardless of camera position, lighting, or background appearance, is the general shape of the target instruments themselves. Their long and narrow structures point towards the use of line-structure concepts; however, previous line-structure segmentation schemes do not transfer well to these images. In particular, changing contrast characteristics mean that responses similar to (1), which depend strongly on intensity, are unreliable. In our method, we have dramatically reduced the dependence of line-structure segmentation on the actual response magnitude. Instead, we have incorporated line-structure ideas into a Bayesian framework that allows us to extract many line-objects while still remaining confident we will isolate the ones that correspond to instruments. This flexibility affords greater independence from, and adaptability to, the changing properties typical of endoscope images.

At the cost of under-segmentation, we have shown the results of the method to reliably isolate only instrument regions in the image. In other contexts where over-segmentation is not as critical a concern, the Bayesian selection criteria can be loosened accordingly. This speaks to a particular advantage of our approach: integration of line-structures into Bayesian concepts, which are highly adaptable to myriad problems, should help extend the use of line-structure concepts to more challenging images than those with good contrast or homogeneous backgrounds.

Acknowledgments

Financial support for this project was provided by the Natural Sciences and Engineering Research Council of Canada and the MENTOR-CIHR program for research into spinal deformities.

References

1. X. Zhang and S. Payandeh, Application of Visual Tracking for Robot Assisted Laparoscopic Surgery. *J. Robotic Systems*, 19(7), 315-328, 2002.
2. V. Caselles, R. Kimmel and G. Sapiro, Geodesic Active Contours, *Int. J. Computer Vision*, 22(1), 61-79, 1997.
3. M. Leventon, O. Faugeras, E. Grimson, and W. Wells, Level Set Based Segmentation with Intensity and Curvature Priors, *Mathematical Methods in Biomedical Image Analysis 2000*.
4. T. Chan and L. Vese, Active Contours Without Edges, *IEEE Trans Image Pro*, 10(2), 266-277, 2001.
5. X. Xie and M. Mirmehdi, RAGS: Region-Aided Geometric Snake, *IEEE Trans Image Pro*, 13(5), 640-652, 2004.
6. M. Leventon, W. Grimson, O. Faugeras, Statistical Shape Influence in Geodesic Active Contours, *Comp. Vision and Patt. Recon. (CVPR)*, 2000.
7. J. Chen and A. Amini, Quantifying 3-D Vascular Structures in MRA Images Using Hybrid PDE and Geometric Deformable Models, *IEEE Trans Medical Imaging*, 23(10), 1251-1262, 2004.
8. T. Koller, G. Gerig, G. Szekely and D. Dettwiler, Multiscale Detection of Curvilinear Structures in 2-D and 3-D Image Data, *IEEE Proc. 5th Int'l. Conf. Comp. Vision*, 864-869, 1995.
9. C. Lorenz, I.-C. Carlsen, T. Buzug, C. Fassnacht and J. Weese, Multi-scale Line Segmentation with Automatic Estimation of Width, Contrast and Tangential Direction in 2D and 3D Medical Images, *Proc. 1st Int'l Joint Conf. CVRMed and MRCAS (CVRMed/MRCAS '97)*, 233-242, 1997.
10. C. Xu and J. L. Prince, ``Snakes, Shapes, and Gradient Vector Flow,'' *IEEE Transactions on Image Processing*, 7(3), 359-369, 1998.

APPENDIX B: SEGMENTATION CODE

extractSkeletons.m

bestEdges.m

breakEdges_crv.m

edgeLink.m

im_edge2param.m

buildDistMap.m

normProj.m

selectSkeletons.m

buildEnvelope.m

eigenCalc.m

lengthResp.m

```
% EXTRACTSKELETONS - Find the line-object skeletons in an image
%
% This function extracts the Canny edge content of an image, then breaks
% these edges at points of high curvature. After edge breaking, short
% edges are removed. The remaining edges are considered in pairs. Pairs
% of edges that are near parallel and close to one another are considered
% to represent possible instrument edges. These are then processed further
% to extract the set of equidistant pixels between them, producing a set of
% candidate "skeletons" along the centre of line-objects in an image.
% These skeletons are then processed by selectSkeletons.m.
%
% Relevant parameters are specified in the "Parameters" section below
%
% See also:  BUILDDISTMAP, NORMPROJECT, IM_EDGE2PARAM, SELECTSKELETONS,
%          BESTEDGES

%
% Luke Windisch
% Institut de genie biomedical
% Ecole Polytechnique de Montreal
% LIV4D Lab (contact: Farida Cheriet, PhD)
%
% July 2005

clear all;
close all;

% These paths work if you keep the directory structure of the folders
% containing these files intact.
addpath('../Images');
addpath('../Global');

%========== Image reading ==========%
imind = 2; % Image to be segmented
imname = ['img' num2str(imind) '.jpg'];
rgbim = imread(imname);

if size(rgbim, 3) == 3, % Convert image to gray scale
   im = double( rgb2gray(rgbim) );
else
   im = double( rgbim );
end

%========== Parameters ==========%
% Thresholds for initial edge extraction
edgemin = 10;              % Minimum length of extracted edges
cannylh = [0.04 0.2062];   % canny thresholds - use [] for default
cannysd = 1.725;           % Canny standard deviation
breaktol = .35;            % Max curvature before breaking a line

% Thresholds for pairing edges
maxscale = 20;    % Maximum seperatation for valid line pairs
angThresh = pi/15; % Minimum degree of parallelism between two edges
```

```
% Thresholds for extracting skeletons
minextent = 0;  % Skeleton length must have a length at least this
          % proportion of shortest edge in pair: [0 1]
          % This parameter could almost certainly improve results, at
          % the risk of reduced generality
diffthresh = 1; % Two edges may diverge up to this amount from original
          % detected scale and still give a skeleton - DOESN'T WORK!

%========== Global definitions ==========%
global M; % Image size parameters
global N;
[M, N] = size(im);

global distGrid; % Base grids for cartesian distances
[X, Y] = meshgrid( [-(N-1):(N-1)], [-(M-1):(M-1)] );
distGrid = sqrt( X.^2 + Y.^2 );

global XY_mag; % Cartesian angle and magnitude grids
global XY_ang;
[X, Y] = meshgrid( [0:N-1], [0:-1:(-M+1)] );
XY_mag = sqrt( X.^2 + Y.^2 );
XY_ang = atan2( Y, X );

%========== Define/calc inputs for skeleton extraction ==========%
% Exract edges from the image
[ edgeim, linelist, lineim, curveim ] = ...
   bestEdges( im, cannylh, cannysd, edgemin, breaktol );

% Initialize structures for skeleton extraction
% The use and meanining of these structures will be apparent in the
% skeleton extraction step below
projdist = zeros( M, N, length( linelist ) ); % Projection around each edge
projinfo = zeros( length( linelist ), 5 );    % Data to define each proj
skelim = zeros( M, N );                  % Final set of skeletons
skellist = {};                           % [r,c] form of skeletons
skelregion = {};                         % Regns surrounding skeletons
q = 1;                                   % Index for skel list

%========== Extract skeletons ==========%
for i = 1 : length( linelist ), % for each extracted edge

   % Output current processing status
   if rem( i, round( .2 * length( linelist ) ) ) == 0,
      disp(['Processing ' int2str(i) ' of ' ...
         int2str( length(linelist) ) ' edges.']);
   end

   % 1 -> Project a distance envelope around the edge in normal direction
   [ distevlp, valdist ] = buildDistMap( linelist{i}, maxscale );
   [ projtan, xychain, valproj ] = normProject( linelist{i} );
   valunion = ( valdist > 0 ) & ( valproj > 0 ); % keep dist map that is
                                 % within normal proj
   % Store information for this line
   projdist(:,:,i) = -100*( ones(M,N) - valunion ) + valunion.*distevlp;
```

```
projinfo(i,:) = [ projtan, xychain(1,1), xychain(1,2),...
      xychain( length(xychain),1 ), xychain( length(xychain),2 ) ];

% 2 -> Determine if any pairings with already processed edges
% There is almost certainly a more efficient way to implement this
% step...it is by far the slowest part of the code
id = [];
if i > 1,

   % Non-intuitive projective geometry to determine distance between
   % endpoint i and other lines. '0' distance point is intersection
   % of a line extended from origin in direction of edge i and the
   % normal through endpoint of edge i. Distance of a line is from
   % '0' point along normal through endpoint i, to point where line
   % intersects this normal.
   d_top = sqrt( projinfo(:,2).^2 + projinfo(:,3).^2 ) .* ...
      sin( projtan - atan2( projinfo(:,3), projinfo(:,2) ) );
   d_bot = sqrt( projinfo(:,3).^2 + projinfo(:,4).^2 ) .* ...
      sin( projtan - atan2( projinfo(:,4), projinfo(:,3) ) );

   delta_top = d_top( 1:(i-1) ) - d_top(i);
   delta_bot = d_bot( 1:(i-1) ) - d_bot(i);

   % Parallelism test
   interAng = abs( cos( projtan - projinfo( 1:(i-1), 1 ) ) );

   % Project envelopes around each edge onto normal of edge i
   projOff = maxscale * interAng;
   proj_top_rt = projOff + delta_top;
   proj_top_lf = -projOff + delta_top;
   proj_bot_rt = projOff + delta_bot;
   proj_bot_lf = -projOff + delta_bot;

   % Pair those edges that are near parallel and close to one another
   ang_id = interAng > cos(angThresh);

   top_id = ( proj_top_lf <= maxscale ) & ( proj_top_rt >= -maxscale );
   bot_id = ( proj_bot_lf <= maxscale ) & ( proj_bot_rt >= -maxscale );
   id = find( ang_id .* ( top_id + bot_id ) );
end

% 3 -> Extract skeletons (if any) from edge pairs
for j = 1:length(id), % for each edge pair

   evlp1 = projdist( :,:,i ) >= 0;    % Keep good parts of projected
   evlp2 = projdist( :,:,id(j) ) >= 0; % dist map around both edges
   diffMap = abs( projdist( :,:,i ) - projdist(:,:,id(j)) );
   overlap = ( diffMap <= diffthresh );
   overlap = overlap .* evlp1 .* evlp2 .* ( diffMap + eps );

   % Skeletons may represent a range of scales if diffthresh > 1
   skel = ( diffthresh - overlap ) .* sign( overlap );
   skel = bwmorph( skel, 'skel', Inf );
   skel = bwmorph( skel, 'spur');
```

```
      % Suppress trivial skeletons that have a length less than minextent
      % percent of shortest of two edges in pair (minextent = 0 to keep
      % all skeletons
      if minextent
         skelthresh = minextent * ...
            min( size( linelist{i},1 ), size( linelist{id(j)},1 ) );
      else
         skelthresh = 1;
      end
      [ skelpts, skelmap ] = im_edge2param( skel, skelthresh );

      % Populate structures of skeleton information for future processing
      for k = 1: length( skelpts ), % Record intersection scales

         skelind = sub2ind( [M N], skelpts{k}(:,1), skelpts{k}(:,2) );
         skelim( skelind ) = q;  % Image of skeleton assigned unique id

         % Structure: skeleton pts, scale at each skeleton pt, length of
         % skeleton, recorded at each skeleton point
         skellist{q} = [ skelpts{k} round( distevlp( skelind ) ) ];
         skelregion{q} = find( evlp1 .* evlp2 );
         q = q+1;

      end % end for k

   end % end for j
end % end for i

%% Use this code to save the relevant results of the skeleton extraction
%% step for subsequent processing
%save( ['img' num2str(imind) '_june19_OPT_' num2str( 180*angThresh/pi ) ...
%    '_' num2str( maxscale )], 'skelregion', 'skelim', 'skellist' );

skel_ind = find( skelim );
final_im = sign( lineim );
final_im( skel_ind ) = 2;
figure, imagesc( final_im ), colormap( copper );

rim = ( ones(M,N) - sign( skelim ) ).*double(rgbim(:,:,1));
gim = ( ones(M,N) - sign( skelim ) ).*double(rgbim(:,:,2)) + sign( skelim )*255;
bim = ( ones(M,N) - sign( skelim ) ).*double(rgbim(:,:,3));
colorim = cat(3, rim, gim, bim);
figure, imagesc( colorim./255 );
```

```
% BESTEDGES - Extract edge content and return those likely to be instrument edges
%
% This function extracts Canny edge content from an image.  The Canny edges
% are then broken at points of high curvature
%
% Arguments:  im       - Image for edge extraction
%         cannylh  - [ low high ] canny thresholds.  Pass as an empty
%                 array for automatic parameter selection
%         cannysd  - Canny standard deviation
%         edgemin  - Minimum edge length before keeping (pixels)
%         breaktol - When curvature along an edge exceeds this value,
%                 break the edge at the high curvature point
%
% Returns:    edgeim   - Original Canny edge content
%         postlist - Cell array of final set of extracted edges in the
%                 form:
%               { [r1 c1   [r1 c1   etc }
%                 r2 c2    ...
%                 ...
%                 rN cN]   ....] }
%         postim   - Image with pixels labeled with edge number
%                 corresponding to postlist
%
% See also:  EXTRACTSKELETONS, EDGELINK, BREAKEDGES_CRV
%
% Luke Windisch
% Institut de genie biomedical
% Ecole Polytechnique de Montreal
% LIV4D Lab (contact: Farida Cheriet, PhD)
%
% July 2005

function [ edgeim, postlist, postim, curveim ] = ...
   bestEdges( im, cannylh, cannysd, edgemin, breaktol )

global M;
global N;

edgeim = edge( im, 'canny', cannylh, cannysd );

% Parameterize edges and suppress those that will
% be removed later anyway (i.e. length < edgemin )
[ prelist, preim ] = edgelink( edgeim, edgemin );

% Break edges at points where curvature exceeds 'breaktol'
[ straightim, curveim ] = ...
   breakEdges_crv( sign(preim), prelist, breaktol );

% Remove those broken edges that no longer meet min length requirement
[ postlist, postim ] = edgelink( straightim, edgemin );
```

```
% BREAKEDGES_CRV - Break edges at points of high curvature
%
% This function breaks image edge content at points where the curvature
% exceeds a threshold.
%
% Arguments:  im       - Image for edge extraction
%          edgelist - Cell array of extracted edges in the form:
%                { [r1 c1   [r1 c1   etc }
%                  r2 c2    ...
%                  ...
%                  rN cN]   ....] }
%          maxcurve - When curvature along an edge exceeds this value,
%                  break the edge at the high curvature point
%
% Returns:    straightim - Edge image after breaking at high curv. points
%          Kim       - Image with high pixels at points where curvature
%                    exceeded maxcurve
%
% See also:  BESTEDGES
%
% Luke Windisch
% Institut de genie biomedical
% Ecole Polytechnique de Montreal
% LIV4D Lab (contact: Farida Cheriet, PhD)
%
% July 2005

function [ straightim, Kim ] = breakEdges_crv( edgeim, edgelist, maxcurve )

global M;
global N;

Kim = zeros(M,N);

for i = 1:length(edgelist), % For each parameterized edge

   xsig = edgelist{i}(:,2) - 1;    % Convert edges: x = m-1
   ysig = -(edgelist{i}(:,1) - 1); %              y = -(n-1)

   dx = gradient( xsig );  % Computer derivatives along edges
   dxx = gradient( dx );
   dy = gradient( ysig );
   dyy = gradient( dy );

   % Find curvature and populate a curvature image
   K = (dx.*dyy - dy.*dxx) ./ (( dx.^2 + dy.^2 ).^(3/2)+eps);

   k = sub2ind( [M N], edgelist{i}(:,1), edgelist{i}(:,2) );

   Kim(k) = K;
end

% Remove points that exceed curvature threshold
k = find( abs(Kim) > maxcurve );
```

```
straightim = edgeim;

for j = 1:length(k),
   [m,n] = ind2sub( [M N], k(j) );
   a = 4;
   straightim( max(1,m-a):min(M,m+a), max(1,n-a):min(N,n+a) ) = 0;
end
```

```
% EDGELINK - Link edge points in an image into lists
%
% Usage: [edgelist edgeim] = edgelink(im, minlength, location)
%
% Arguments: im        - Binary edge image, it is assumed that edges
%                        have been thinned.
%            minlength - Minimum edge length of interest
%            location  - Optional complex valued image holding subpixel
%                        locations of edge points. For any pixel the
%                        real part holds the subpixel row coordinate of
%                        that edge point and the imaginary part holds
%                        the column coordinate. See NONMAXSUP. If
%                        this argument is supplied the edgelists will
%                        be formed from the subpixel coordinates,
%                        otherwise the the integer pixel coordinates of
%                        points in 'im' are used.
%
% Returns:  edgelist - a cell array of edge lists in row,column coords in
%                      the form
%                     { [r1 c1   [r1 c1   etc }
%                       r2 c2    ...
%                       ...
%                       rN cN]   ....]
%
%           edgeim   - Image with pixels labeled with edge number.
%
%
% This function links edge points together into chains.  Where an edge
% diverges at a junction the function simply tracks one of the branches.
% The other branch is eventually processed as another edge.  These `broken
% branches' can be remerged by MERGESEG after a call to LINESEG.
%
% See also:  BESTEEDGES

% Acknowledgement:
% This code is inspired by David Lowe's Link.c function from the Vista image
% processing library developed at the University of British Columbia
%   http://www.cs.ubc.ca/nest/lci/vista/vista.html
%
% Peter Kovesi
% School of Computer Science & Software Engineering
% The University of Western Australia
% pk at csse uwa edu au
% http://www.csse.uwa.edu.au/~pk
%
% February  2001 - Original version
% September 2004 - Revised to allow subpixel edge data to be used

function [edgelist, edgeim] = edgelink(im, minlength, location)

  global EDGEIM;     % Some global variables to avoid passing (and
                     % copying) of arguments, this improves speed.
  global ROWS;
  global COLS;
```

```
   elist = {};

   EDGEIM = im ~= 0;                % make sure image is binary.

%  EDGEIM = bwmorph(EDGEIM,'thin',Inf);  % make sure edges are thinned.
%  show(EDGEIM,1)

   EDGEIM = double(EDGEIM);
   [ROWS, COLS] = size(EDGEIM);
   edgeNo = 1;


   % Perform raster scan through image looking for edge points.  When a
   % point is found trackedge is called to find the rest of the edge
   % points.  As it finds the points the edge image pixels are labeled
   % with the -ve of their edge No

   for r = 1:ROWS
         for c = 1:COLS
            if EDGEIM(r,c) == 1
                  edgepoints = trackedge(r,c, edgeNo, minlength);
                  if ~isempty(edgepoints)
                     edgelist{edgeNo} = edgepoints;
                     edgeNo = edgeNo + 1;
                  end
            end
         end
   end

   edgeim = -EDGEIM;  % Finally negate image to make edge encodings +ve.

   % If subpixel edge locations are supplied upgrade the integer precision
   % edgelists that were constructed with data from 'location'.
   if nargin == 3
         for I = 1:length(edgelist)
            ind = sub2ind(size(im),edgelist{I}(:,1),edgelist{I}(:,2));
            edgelist{I}(:,1) = real(location(ind))';
            edgelist{I}(:,2) = imag(location(ind))';
         end
   end

%  show(edgeim,2), colormap(flag);

%--------------------------------------------------------------------
% TRACKEDGE
%
% Function to track all the edge points associated with a start point.  From
% a given starting point it tracks in one direction, storing the coords of
% the edge points in an array and labelling the pixels in the edge image
% with the -ve of their edge number.  When no more connected points are
% found the function returns to the start point and tracks in the opposite
% direction.  Finally a check for the overall number of edge points found is
% made and the edge is ignored if it is too short.
```

```
%
% Note that when a junction is encountered along an edge the function
% simply tracks one of the branches.  The other branch is eventually
% processed as another edge.  These `broken branches' can be remerged by
% MERGESEG.
%
% Usage:   edgepoints = trackedge(rstart, cstart, edgeNo, minlength)
%
% Arguments:   rstart, cstart   - row and column No of starting point
%          edgeNo          - the current edge number
%          minlength       - minimum length of edge to accept
%
% Returns:    edgepoints      - Nx2 array of row and col values for
%                       each edge point.
%                       An empty array is returned if the edge
%                       is less than minlength long.


function edgepoints = trackedge(rstart, cstart, edgeNo, minlength)

   global EDGEIM;
   global ROWS;
   global COLS;

   edgepoints = [rstart cstart];      % Start a new list for this edge.
   EDGEIM(rstart,cstart) = -edgeNo;   % Edge points in the image are
                                      % encoded by -ve of their edgeNo.

   [thereIsAPoint, r, c] = nextpoint(rstart,cstart); % Find next connected
                                      % edge point.

   while thereIsAPoint
         edgepoints = [edgepoints          % Add point to point list
                       r   c   ];
         EDGEIM(r,c) = -edgeNo;            % Update edge image
         [thereIsAPoint, r, c] = nextpoint(r,c);
   end

   edgepoints = flipud(edgepoints);  % Reverse order of points

   % Now track from original point in the opposite direction

   [thereIsAPoint, r, c] = nextpoint(rstart,cstart);

   while thereIsAPoint
         edgepoints = [edgepoints
                       r   c   ];
         EDGEIM(r,c) = -edgeNo;
         [thereIsAPoint, r, c] = nextpoint(r,c);
   end

   % Reject short edges
   Npoints = size(edgepoints,1);
   if Npoints < minlength
```

```
        for i = 1:Npoints   % Clear pixels in the edge image
          EDGEIM(edgepoints(i,1), edgepoints(i,2)) = 0;
        end
        edgepoints = [];    % Return empty array
    end


%---------------------------------------------------------------------
%
% NEXTPOINT
%
% Function finds a point that is 8 connected to an existing edge point
%

function [thereIsAPoint, nextr, nextc] = nextpoint(rp,cp);

    global EDGEIM;
    global ROWS;
    global COLS;

    % row and column offsets for the eight neighbours of a point
    % starting with those that are 4-connected.
    roff = [1  0 -1  0  1  1 -1 -1];
    coff = [0  1  0 -1  1 -1 -1  1];

    r = rp+roff;
    c = cp+coff;

    % Search through neighbours and return first connected edge point
    for i = 1:8
        if r(i) >= 1 & r(i) <= ROWS & c(i) >= 1 & c(i) <= COLS
          if EDGEIM(r(i),c(i)) == 1
                nextr = r(i);
                nextc = c(i);
                thereIsAPoint = 1;
                return;         % break out and return with the data
          end
        end
    end

    % If we get here there was no connecting next point.
    nextr = 0;
    nextc = 0;
    thereIsAPoint = 0;
```

```
% IM_EDGE2PARAM - Link edge points in an image into lists
%
% Usage: [edgelist edgeim] = im_edge2param(im, minlength)
%
% Arguments:  im         - Binary edge image, it is assumed that edges
%                          have been thinned.
%             minlength  - Minimum edge length of interest
%
% Returns:  edgelist - a cell array of edge lists in row,column coords in
%                      the form
%                     { [r1 c1   [r1 c1   etc }
%                       r2 c2    ...
%                       ...
%                       rN cN]   ....]
%
%           edgeim   - Image with pixels labeled with edge number.
%
%
% This function links edge points together into chains.  Where an edge
% diverges at a junction the function simply tracks one of the branches.
% The other branch is eventually processed as another edge.  These `broken
% branches' can be remerged by MERGESEG after a call to LINESEG.
%
% Acknowledgement:
% This code is inspired by David Lowe's Link.c function from the Vista image
% processing library developed at the University of British Columbia
%   http://www.cs.ubc.ca/nest/lci/vista/vista.html
%
% Peter Kovesi
% School of Computer Science & Software Engineering
% The University of Western Australia
% pk @ csse uwa edu au
% http://www.csse.uwa.edu.au/~pk
%
% February 2001

function [edgelist, edgeim] = im_edge2param(im, minlength)

   global EDGEIM;      % Some global variables to avoid passing (and
                       % copying) of arguments, this improves speed.
   global M;
   global N;

   edgelist = {};

   EDGEIM = im ~= 0;                  % make sure image is binary.
   EDGEIM = bwmorph(EDGEIM,'thin',Inf);  % make sure edges are thinned.

%  show(EDGEIM,1)
   EDGEIM = double(EDGEIM);
   %[M, N] = size(EDGEIM);
   edgeNo = 1;
```

```
  % Perform raster scan through image looking for edge points.  When a
  % point is found trackedge is called to find the rest of the edge
  % points.  As it finds the points the edge image pixels are labeled
  % with the -ve of their edge No

  for r = 1:M
        for c = 1:N
          if EDGEIM(r,c) == 1
                edgepoints = pk_trackEdge(r,c, edgeNo, minlength);
                if ~isempty(edgepoints)
                  edgelist{edgeNo} = edgepoints;
                  edgeNo = edgeNo + 1;
                end
          end
        end
  end

  edgeim = -EDGEIM;  % Finally negate image to make edge encodings +ve.
%  show(edgeim,2), colormap(flag);

%-----------------------------------------------------------------------
% PK_TRACKEDGE
%
% Function to track all the edge points associated with a start point.  From
% a given starting point it tracks in one direction, storing the coords of
% the edge points in an array and labelling the pixels in the edge image
% with the -ve of their edge number.  When no more connected points are
% found the function returns to the start point and tracks in the opposite
% direction.  Finally a check for the overall number of edge points found is
% made and the edge is ignored if it is too short.
%
% Note that when a junction is encountered along an edge the function
% simply tracks one of the branches.  The other branch is eventually
% processed as another edge.  These `broken branches' can be remerged by
% MERGESEG.
%
% Usage:   edgepoints = pk_trackEdge(rstart, cstart, edgeNo, minlength)
%
% Arguments:  rstart, cstart   - row and column No of starting point
%           edgeNo          - the current edge number
%           minlength       - minimum length of edge to accept
%
% Returns:    edgepoints      - Nx2 array of row and col values for
%                               each edge point.
%                               An empty array is returned if the edge
%                               is less than minlength long.


function edgepoints = pk_trackEdge(rstart, cstart, edgeNo, minlength)

  global EDGEIM;
  global M;
  global N;
```

```
   edgepoints = [rstart cstart];      % Start a new list for this edge.
   EDGEIM(rstart,cstart) = -edgeNo;   % Edge points in the image are
                                      % encoded by -ve of their edgeNo.

   [thereIsAPoint, r, c] = pk_nextPoint(rstart,cstart); % Find next connected
                                      % edge point.

   while thereIsAPoint
         edgepoints = [edgepoints          % Add point to point list
                       r   c  ];
         EDGEIM(r,c) = -edgeNo;            % Update edge image
         [thereIsAPoint, r, c] = pk_nextPoint(r,c);
   end

   edgepoints = flipud(edgepoints);  % Reverse order of points

   % Now track from original point in the opposite direction

   [thereIsAPoint, r, c] = pk_nextPoint(rstart,cstart);

   while thereIsAPoint
         edgepoints = [edgepoints
                       r   c  ];
         EDGEIM(r,c) = -edgeNo;
         [thereIsAPoint, r, c] = pk_nextPoint(r,c);
   end

   % Reject short edges
   Npoints = size(edgepoints,1);
   if Npoints < minlength
         for i = 1:Npoints   % Clear pixels in the edge image
            EDGEIM(edgepoints(i,1), edgepoints(i,2)) = 0;
         end
         edgepoints = [];    % Return empty array
   end


%--------------------------------------------------------------------
%
% PK_NEXTPOINT
%
% Function finds a point that is 8 connected to an existing edge point
%

function [thereIsAPoint, nextr, nextc] = pk_nextPoint(rp,cp);

   global EDGEIM;
   global M;
   global N;

   % row and column offsets for the eight neighbours of a point
   % starting with those that are 4-connected.
   roff = [1  0 -1  0  1  1 -1 -1];
   coff = [0  1  0 -1  1 -1 -1  1];
```

```
r = rp+roff;
c = cp+coff;

% Search through neighbours and return first connected edge point
for i = 1:8
        if r(i) >= 1 & r(i) <= M & c(i) >= 1 & c(i) <= N
           if EDGEIM(r(i),c(i)) == 1
                   nextr = r(i);
                   nextc = c(i);
                   thereIsAPoint = 1;
                   return;          % break out and return with the data
           end
        end
end

% If we get here there was no connecting next point.
nextr = 0;
nextc = 0;
thereIsAPoint = 0;
```

```
% BUILDDISTMAP - Grid of distances to a line in an image.
%
% Usage: [ distMap, validMap ] = buildDistMap( edgeList, max_dist )
%
% This function builds a matrix of cartesion distances from each point
% along an edge.
%
% Arguments:  edgelist   - Array of edge points in the format
%                     [r1 c1
%                      r2 c2
%                      ...
%                      rN cN]
%           max_dist   - The maximum distance of interest from an edge
%
% Returns:    distMap    - A matrix of cartesian distances from edge pixels
%           validMap   - A binary matrix indicating which pixels are
%                        within max_dist of the edge
%
% See also:  EXTRACTSKELETONS, NORMPROJECT
%
% Luke Windisch
% Institut de genie biomedical
% Ecole Polytechnique de Montreal
% LIV4D Lab (contact: Farida Cheriet, PhD)
%
% July 2005

function [ distMap, validMap ] = buildDistMap( edgeList, max_dist )

global M;
global N;
global distGrid;

distMap = 1000*ones( M,N );
validMap = ones( M,N );

for i = 1:length( edgeList ), % Center current pixel on distance map

   fore_n = edgeList(i,2) - 1; % Column range
   back_n = N - edgeList(i,2);

   fore_m = edgeList(i,1) - 1; % Row range
   back_m = M - edgeList(i,1);

   distMatrix = distGrid( (M-fore_m : M+back_m), (N-fore_n : N+back_n) );
   distMap = min( distMap, distMatrix );

end

if ~isempty( max_dist ),
   validMap = ( distMap <= max_dist );
   validMap = sign( validMap - .5 );
end
```

```
% NORMPROJ - Pixels withing a the zone bounded by the normals through the
%         endpoints of an edge
%
% Usage: [ ang_t, xycoord, validMap ] = normProject( lineList )
%
% This function returns a matrix of 0 and 1s. It indicates the pixels
% bounded by the normal vectors through the two endpoints of an edge.
%
% Arguments:  lineList   - Array of edge points in the format
%                    [r1 c1
%                     r2 c2
%                     ...
%                     rN cN]
%
% Returns:    ang_t     - The approximate direction of the edge (radians)
%           xycoord   - Equivalent cartesian coordinates of lineList
%           validMap  - A matrix where pixels bounded by the endpoint
%                     normals are high
%
% See also:  EXTRACTSKELETONS, BUILDDISTMAP
%
% Luke Windisch
% Institut de genie biomedical
% Ecole Polytechnique de Montreal
% LIV4D Lab (contact: Farida Cheriet, PhD)
%
% July 2005

function [ ang_t, xycoord, validMap ] = normProject( lineList )

global M;
global N;
global XY_mag;
global XY_ang;

y = -( lineList(:,1) - 1 ); % Convert line coordinates to Cartesian
x = ( lineList(:,2) - 1 );

x1 = x(1); % Isolate endpoints
y1 = y(1);
x2 = x( length(x) );
y2 = y( length(y) );

if y1 ~= y2, % Ensure first point is the (cartesian) top of line
   if y1 < y2, % Have to reverse the chain order
      [x, y, x1, y1, x2, y2] = reverseOrder(x, y, x1, y1, x2, y2);
   end
else
   if x2 < x1, % Have to reverse the chain order
      [x, y, x1, y1, x2, y2] = reverseOrder(x, y, x1, y1, x2, y2);
   end
end

% Define tangent direction at top and bottom of line
```

```
ang_t = atan2( ( y2 - y1 ), ( x2- x1 ) );

% Define line endpoint distances in Cartesian frame
rho1 = sqrt( x1^2 + y1^2 ) * cos( ang_t - atan2( y1, x1 ) );
rho2 = sqrt( x2^2 + y2^2 ) * cos( ang_t - atan2( y2, x2 ) );

% Define distances to all points in Cartesian frame
rho1_XY = XY_mag .* cos( ang_t - XY_ang );
rho2_XY = XY_mag .* cos( ang_t - XY_ang );

% Keep points within normal projections of line endpoints
projLine = ( rho1_XY >= rho1 ) .* ( rho2_XY <= rho2 );
validMap = sign( projLine - .5 );

% Remap the line into image coordinates
xycoord = [ x y ];


%================================
function [x, y, x1, y1, x2, y2] = reverseOrder(x, y, x1, y1, x2, y2)

x = flipud( x );
y = flipud( y );
temp1 = x1; temp2 = y1;
x1 = x2;    y1 = y2;
x2 = temp1; y2 = temp2;
```

```
% EXTRACTSKELETONS - Find the line-object skeletons in an image
%
% This script takes a set of extracted line-object skeletons from the script
% extractSkeletons.m as input, and processes them to determine which
% correspond to instruments. The first step is to collect skeletons into
% clusters that are assumed to represent the same instrument. The scale
% and length of these clusters are used to assign a likelihood measure.
% The number of skeletons in each cluster is used to produce an a priori
% probability. Also, the average eigenresponse along a skeleton is used as
% an additional a priori probability weighting. The likelihood and two
% prior terms are combined in a Bayesian manner to give a final probability
% that the cluster of skeletons corresponds to an instrument.
%
% Relevant parameters are specified in the "Parameters" section below
%
% See also:  BUILDENVELOPE, EIGENCALC, LENGTHRESP, EXTRACTSKELETONS
%
% Luke Windisch
% Institut de genie biomedical
% Ecole Polytechnique de Montreal
% LIV4D Lab (contact: Farida Cheriet, PhD)
%
% July 2005

close all;

%========== Parameters ==========%
global M;
global N;

global pset; % Parameters to control beta term of line-length weight
pset = [40 40 80 2]; % [min_len max_scl ctrl_len ctrl_scl]

maxext = 80;       % Maximum filter extent for Gaussian blurring
gamma = (3/2);     % Exponent in eigenvalue response to offset blurring
probthresh = 0.121; % Keep skeletons that have a final probability > thresh
initcnst = 1000;   % Arbitrary intialization ...set to high value (~1000)
parthresh = 15;    % Only group skeletons near parallel

%========== Input/Output Structures ==========%
if ~ exist('im','var'), % If used the save command in extractSkeletons.m

   addpath('../Images');
   addpath('../Global');

   imind = 3; % Image to be segmented
   imname = ['img' num2str(imind) '.jpg'];
   rgbim = imread(imname);

   if size(rgbim, 3) == 3, % Convert image to gray scale
      im = double( rgb2gray(rgbim) );
   else
      im = double( rgbim );
   end
```

```
   [M, N] = size(im);
   % Must use load to read in results of extractSkeletons results (if they
   % were saved).

end

%=========== Group skeletons ===========%
% Begin with large scales and find those skeletons at lower scales that are
% within its envelope. The skeletons that are grouped are limited to those
% that are near parallel, as per the parthresh parameter

firstskel = 1;
for i = 1:length( skellist ),

   skelscl = round( mean( skellist{i}(:,3) ) ); % Scales of skels
   skellen = length( skellist{i}(:,1) );       % Lengths of skels

   if (skellen > 2) & (skellen >= 0.25*skelscl),

      % Determine approximate orientation of current skeleton
      skelang = atan( -( skellist{i}( length(skellist{i}(:,1)),1 ) ...
         - skellist{i}(1,1) ) / ( skellist{i}( length(skellist{i}(:,1)),2 ) ...
         - skellist{i}(1,2) + eps) );

      % Build an envelope or 'footprint' around the skeleton with an
      % extent equal to the skeleton's scale
      skelevlp = buildEnvelope( zeros(M,N), skellist{i}(:,1:2), skelscl, 'cnst');

      % Build a summary of each skeleton group
      skelgrp = [ i; skelang; skellen; skelscl ];
      if firstskel,
         SKELGRPS{1} = skelgrp;
         SKELEVLPS = skelevlp;
         skelcnt = [ 1 ];
         firstskel = 0;
      else
         globalid = initcnst;
         globalmin = initcnst;
         for p = 1:length(SKELGRPS),

            % Check that the skeleton is near parallel
            angdiff = abs( skelang - SKELGRPS{p}(2,:) );
            [mindiff, minid] = min(angdiff);

            if (mindiff < parthresh*pi/180) & (mindiff < globalmin),
               globalmin = mindiff;
               globalid = p;
            end
         end

         if globalid >= initcnst,
            newgrp = length( SKELGRPS ) + 1;
            SKELGRPS{ newgrp } = skelgrp;
```

```
            tempSKELEVLPS = SKELEVLPS;
            SKELEVLPS = zeros(M,N,newgrp);
            SKELEVLPS(:,:,1:newgrp-1) = tempSKELEVLPS;
            SKELEVLPS(:,:,newgrp) = skelevlp;
            skelcnt = [ skelcnt; 1 ];
          else
            oldevlp = SKELEVLPS(:,:,globalid);
            [ regionim, nregion ] = bwlabel( sign(skelevlp + oldevlp), 8 );

            if nregion > 1,
              newgrp = length( SKELGRPS ) + 1;
              SKELGRPS{ newgrp } = skelgrp;
              tempSKELEVLPS = SKELEVLPS;
              SKELEVLPS = zeros(M,N,newgrp);
              SKELEVLPS(:,:,1:newgrp-1) = tempSKELEVLPS;
              SKELEVLPS(:,:,newgrp) = skelevlp;
              skelcnt = [ skelcnt; 1 ];
            else
              grparray = SKELGRPS{globalid};
              grparray = [grparray skelgrp];
              SKELGRPS{globalid} = grparray;
              newevlpim = sign( SKELEVLPS(:,:,globalid) + skelevlp );
              SKELEVLPS(:,:,globalid) = newevlpim;
              skelcnt( globalid ) = skelcnt( globalid ) + 1;
            end
          end
        end
      end
    end

%========== Assign probabilistic measures ==========%
for j = 1:length(skelcnt),

  % 1 -> Measure scale, length and number of skeletons for group
  grpscl = max( SKELGRPS{j}(4,:) ); % Maximum scale of current group

  % Extract the overall length of the current group
  [ grplen, lenid ] = max( SKELGRPS{j}(3,:) );
  rotang = SKELGRPS{j}(2,lenid) * 180/pi;
  rotim = imrotate( SKELEVLPS(:,:,j), -rotang );
  [rgn_m, rgn_n] = find(rotim);
  grplen = max( grplen, abs( max(rgn_n) - min(rgn_n) ) );

  grpskel = SKELGRPS{j}(1,lenid);
  grpcnt = skelcnt(j); % The number of skeletons in the group

  % 2 -> Measure eigenresponse and likelihood for group
  % Begin by blurring the image
  f = fspecial( 'gaussian', min( maxext, 4*grpscl ), grpscl );
  fim = imfilter( im, f, 'replicate' );
  % Calculate the eigenvalue line-object response
  [ D1, D2, V1, V2 ] = ...
    eigenCalc( fim, skellist{grpskel}(:,1), skellist{grpskel}(:,2) );
```

```
   id1 = abs( D1 ) > abs( D2 ); % Find principle eigenvalues
   id2 = ones( size(id1) ) - id1;

   eigval = abs( D1 ).*id1 + abs( D2 ).*id2;
   [ dum1, dumb2, eigskel ] = find( grpscl^gamma * eigval ); % Blur compensation
   grpeign = mean( eigskel ); % Mean eigenresponse along chain

   grplkhd = lengthResp( grplen, grpscl ); % Likelihood term

   % 3 -> Form final vector of values for this group
   grpvals(j,:) = [ j grpskel grplen grpscl grplkhd grpcnt grpeign ];

end

%========== Assess final probabilities ==========%
grpprob = [ grpvals(:,1) grpvals(:,5) grpvals(:,6)./max( grpvals(:,6) ) ...
   grpvals(:,7)./max( grpvals(:,7) ) ];
probset = [ grpprob grpprob(:,2).*grpprob(:,3).*grpprob(:,4) ];


% Plotting
figure;
for q = 1:length(probthresh),
   keepid = find( probset(:,5) >= probthresh(q) );

   segim = sign( sum( SKELEVLPS(:,:,keepid), 3 ) );
   segedge = edge( segim, 'sobel' );

   rim = ( ones(M,N) - segedge ).*double(rgbim(:,:,1));
   gim = ( ones(M,N) - segedge ).*double(rgbim(:,:,2)) + segedge*255;
   bim = ( ones(M,N) - segedge ).*double(rgbim(:,:,3));
   colorim = cat(3, rim, gim, bim);
   subplot(2,3,q),imagesc( colorim./255 );
end
```

```
% BUILDEVLP - Build a footprint around an edge at a given scale
%
% Usage: evlp = buildEnvelope( evlp, lineobj, footsize, evlp_type)
% Arguments:  evlp      - Current footprint image (possibly containing
%                         footprints around other edges)
%             lineobj   - Edge content to build envelope around.  Can be an
%                         image or in the form [r1 c1
%                                               ...
%                                               rN cN]
%             footsize  - Extent of the envelope (i.e. scale)
%             evlp_type - 'cnst' for an envelope of ones, and 'gaus' for an
%                         envelope with a Gaussian structure
%
% Returns:    evlp    - Updated footprint image
%
% See also:  SELECTSKELETONS
%
% Luke Windisch
% Ecole Polytechnique de Montreal
% July 2005

function evlp = buildEnvelope( evlp, lineobj, footsize, evlp_type)

global M;
global N;

if size( lineobj, 2 ) ~= 2, % lineobj may be passed as an image
   [r, c] = find( lineobj );
   lineobj = [r c];
   val = footsize( sub2ind( [M N], r, c ) );
else
   val = ones(length( lineobj(:,1) )) * footsize;
end

for k = 1:length( lineobj(:,1) ), % Build a footprint around line structure

   lft = max( 1, round(lineobj(k,2))-val(k) );
   rgt = min( N, round(lineobj(k,2))+val(k) );
   top = max( 1, round(lineobj(k,1))-val(k) );
   bot = min( M, round(lineobj(k,1))+val(k) );

   [X, Y] = meshgrid([lft:rgt], [top:bot]);
   circ = ( ( X-lineobj(k,2) ).^2 + ( Y-lineobj(k,1) ).^2 ) < val(k).^2;

   switch evlp_type,
      case 'cnst'
         evlp(top:bot, lft:rgt) = max( evlp(top:bot, lft:rgt), circ );

      case 'gaus'
         gaus = circ .* fspecial('gaussian', [bot-top+1, rgt-lft+1], 0.4*footsize);
         evlp(top:bot, lft:rgt) = max( evlp(top:bot, lft:rgt), gaus./max(max(gaus)) );

   end
end % for k
```

```
% EIGENCALC - Calculate the eigenvalue line-object response
%
% Usage: [D1, D2 V1, V2, lineInten, lineRatio] = eigenCalc( fim, mpts, npts )
%
% Arguments:  fim  - Image blurred at Gaussian scale of interest
%         mpts - row coordinates of skeleton points of interest
%         npts - col coordinates of skeleton points of interest
%
% Returns:   D1, D2   - Eigenvalues
%         V1, V2   - Eigenvectors
%         lineInten - Image of pure eigenvalue magnitudes
%         lineRatio - Matrix of ratio: max(D1,D2)/min(D1,D2) at each
%                 pixel
%
% This function calculate the eigenvalue-based line object response as
% developed by:
% C. Lorenz, I.-C. Carlsen, T. Buzug, C. Fassnacht and J. Weese,
% "Multi-scale Line Segmentation with Automatic Estimation of Width,
%  Contrast and Tangential Direction in 2D and 3D Medical Images," Proc. 1st
% Int'l Joint Conf. CVRMed and MRCAS (CVRMed/MRCAS '97), 233-242, 1997.
%
%   pset = [1-beta_lmin 2-beta_smax 3-beta_lmax 4-beta_smin 5-slope_max]
%
% See also:  SELECTSKELETONS
%
% Luke Windisch
% Ecole Polytechnique de Montreal
% July 2005

function [ D1, D2 V1, V2, lineInten, lineRatio ] = eigenCalc( fim, mpts, npts )

global M;
global N;

[ FX FY ] = gradient( fim ); % Calculate first and second numerical derivatives
[ FXFX FYFX ] = gradient( FX );
[ FXFY FYFY ] = gradient( FY );

[M, N] = size( FXFX );

D1 = zeros(M,N);   D2 = zeros(M,N);
V1x = zeros(M,N);  V1y = zeros(M,N);
V2x = zeros(M,N);  V2y = zeros(M,N);

lineInten = zeros(M,N);
lineRatio = zeros(M,N);

for i = 1:length(mpts),

   m = mpts(i);    n = npts(i);

   H = [ FXFX(m,n) FXFY(m,n); FYFX(m,n) FYFY(m,n) ];
   [ V, D ] = eig( H, 'nobalance' );
   D1(m,n) = D(1,1);  D2(m,n) = D(2,2);
```

```
    V1x(m,n) = V(1,1); V1y(m,n) = V(2,1);
    V2x(m,n) = V(1,2); V2y(m,n) = V(2,2);

    if abs( D(1,1) ) > abs( D(2,2) ),
       lineInten(m,n) = sign( D(1,1) );
       lineRatio(m,n) = abs( D(1,1) ) / ( abs( D(2,2) ) + eps );
    else
       lineInten(m,n) = sign( D(2,2) );
       lineRatio(m,n) = abs( D(2,2) ) / ( abs( D(1,1) ) + eps );
    end

end

V1 = cat( 3, V1x, V1y );
V2 = cat( 3, V2x, V2y );
```

```
% LENGTHRESP - Determine the length-based weight of a line structure
%
% Usage: p = lengthResp( line_len, scale )
%
% Arguments:  linelen - Length of the line structure currently considered
%             scale   - Scale of the line structure currently considered
%
% Returns:    p - The weight for the current line structure based on length
%                 and scale
%
% This function weights line structures according to their length and scale.
% Lower length structures are given lower weighting.  However, the bias in
% favour of long structures is reduced as the scale increases.  The length
% at which the 1/2-max weight occurs at each scale is controlled by the
% beta parameter, and the transition through the 1/2-max point is controlled
% by alpha.  Beta is the most important parameter and is specified according
% to pset as:
%
%   pset = [1-beta_lmin 2-beta_smax 3-beta_lmax 4-beta_smin 5-slope_max]
%
%   w/  beta_lmin := Length below which object is not given 1/2-max weight,
%                    regardless of scale
%       beta_lmax := Length beyond which object is given 1/2-max weight, regardless
%                    of scale
%       beta_smin := Scale below which the beta_lmax constraint is applied
%       beta_smax := Scale above which the beta_lmin constraint is applied
%       slope_max := Maximum slope through 1/2-max point (NOT USED)
%
% See also:  SELECTSKELETONS
%
% Luke Windisch
% Ecole Polytechnique, Montreal
% July 2005

function p = lengthResp( line_len, scale )

global M;
global N;
global pset;

% Determine the beta parameter
beta_sd = -( pset(3) - pset(1) ) / log( pset(4)/pset(2) ); % 1/2-max curve
betai = -beta_sd * log( scale/pset(2) ) + pset(1);

beta = min( max(betai, pset(1)), max(M, N) ); % Enforce the pset constraints

% Determine the alpha parameter
alpha = min(1, scale./beta);
% alpha = (pset(2)/pset(1)) * min(1, scale/beta);

% Determine final weight
p1 = .5 + (1/pi)*atan( alpha*( line_len - beta ) );
p_off = .5 + (1/pi)*atan( -alpha*beta ); % Force response --> 0 for length --> 0
p = p1 - p_off;
```

APPENDIX C: TRACKING CODE

LSTrack.m

LSTrack_script.m

imagein.m

getEigenResponse.m

estimateF.m

evolveLS.m

```
% LSTrack - Bayesian version of optical flow tracking
%
% Luke Windisch
% Ecole Polytechnique
% July 2005
% ------------------------
% Implementation of the level-set based Bayesian optical flow formulation
% developed by A.R. Mansouri in:
% A.-R. Mansouri, "Region tracking via level set PDEs without motion
%  computation," IEEE Trans. Pattern Anlaysis and Machine Intelligence,
% 24(7), pp. 947-961, 2002.
%
% Integrates the use of an eigenvalue based line-object response through
% the imagein.m function.
%
% Example parameter values can be found in the scrip LSTrack_script.m
%
% SEE ALSO: LSTRACK_SCRIPT, IMAGEIN, ESTIMATEF, EVOLVELS
%

function BOUNDseq = LSTrack( version )

%========== Parameters ================%
global Irange;  % Intensity range
global boundcol; % Boundary colour
global seqid;   % Sequence to process
global frameA;  % Index of first frame for current set
global frameZ;  % Index of last frame of current set
global frame1;  % Index of first possible frame of this sequence

global deltaT;  % Time step for LS update
global lambdaL;  % Multiplier to weight curvature prior
global srchwin;  % Search window size
global numiter;  % Number of level set evolutions per frame

global scale;   % Scale of eigenresponse
global polarize; % 1 if speed is [-1 +1] only

global M;
global N;
global P;

zerodir = 1; % 1 if zero level set is region, -1 otherwise

%========== Initializations ==========%
addpath(['../Images/IntenSequence' num2str(seqid)]);
addpath(['../Images/EigenSequence' num2str(seqid)]);
addpath(['../Global']);

[ im0, LS ] = imagein( frameA, version );
[ M, N, P ] = size( im0 );

LS = sign( sign(LS) + zerodir*0.5 ); % reset level set
                        % if zerodir = +1, then 0 --> +1
```

```
%========== Structures to hold the tracked video sequence ==========%
BOUNDseq = zeros( size(im0,1), size(im0,2), frameZ-frameA+1 );
BOUNDseq( :,:,1 ) = LS;

%========== Region tracking ==========%
for index = frameA+1 : frameZ,

   [ im1, dummy ] = imagein( index, version ); % get next image

   disp(['Processing speed estimate between frames ' num2str(index-1) ...
        ' and ' num2str(index) ]);
   F = 255.*estimateF( im0, im1, LS );

   disp('Evolving level set');
   LS = evolveLS( LS, F ); % Level set evolution of region

   LS = sign( sign(LS) + zerodir*0.5 ); % reset level set
                                % if zerodir = +1, then 0 --> +1

   disp(['LS update complete between frames ' int2str(index - 1) ...
      ' and ' int2str(index)]);

   im0 = im1; % Advance the sequence and write to structures
   BOUNDseq( :,:,index ) = LS;

end
```

```
% testScript.m
% Runs a series of MATLAB scripts in pseudo-batch mode
% Included to give example parameter values
%
% Luke Windisch
% Ecole Polytechnique de Montreal
% July 2005

clear all;
close all;

global Irange;  % Intensity range
global boundcol; % Boundary colour
global seqid;   % Sequence to process
global frameA;  % Index of first frame for current set
global frameZ;  % Index of last frame of current set
global frame1;  % Index of first possible frame for sequence

global deltaT;  % Time step for LS update
global lambdaL; % Multiplier to weight curvature prior
global srchwin; % Search window size
global numiter; % Number of level set evolutions per frame

global scale;   % Scale of eigenresponse
global polarize;

Irange = 255;  boundcol = 2;
seqid = 1;  frameA = 1; frameZ = 30; frame1 = 1;
scale = 12;
deltaT = 0.001; lambdaL = 10;  srchwin = 3;   numiter = 250;
version = 'eigen'; polarize = 0;

BOUNDseq = LSTrack( version );
```

```
% imagein: part of Level Set PDE region tracking implementation
%
% Luke Windisch
% Ecole Polytechnique
% July 2005
% ------------------------
% Reads in either a jpg image or an eigenvalue based line-object response.
% Also reads a level set image if the tracking sequence is not initiated to
% the first frame (i.e. read in an existing level set image if you are
% resuming tracking mid-way through a sequence).
%
% INPUTS:  index   - current index into current set
%          version - 'inten' if processing a pure image. 'eigen' if
%                    processing an eigenvalue line response
%
% OUTPUTS: im - current image (corresponds to index)
%          LS - level set/segmentation if index is first for current set
%
% SEE ALSO: GETEIGENRESPONSE, LSTRACK

function [ im, LS ] = imagein( index, version )

global Irange;  % Intensity range
global seqid;   % Sequence to process
global frameA;  % Index of first frame for current set
global frame1;  % Index of first possible frame of this sequence
global scale;   % Scale of eigenresponse

global M;
global N;

%==========Definitions==========%
switch version
   case 'inten'
      imname = ['endoSeq' num2str(seqid) '_' num2str(index) '.jpg'];
      rgbim = imread(imname);

      if size(rgbim, 3) == 3, % Convert image to gray scale
         im = double( rgb2gray(rgbim) );
      else
         im = double( rgbim );
      end

   case 'eigen'
      % To speed testing, this step assumes the eigen/line response has
      % been preprocessed.  To determine response in real time, call
      % getEigenResponse.m
      imname = ['eigenSeq' num2str(seqid) '_' num2str(index) '_scale' ...
               num2str(scale)];
      tempim = load(imname);
      im = tempim.eigresp;

end
```

```
im = im ./ max(max(abs(im))); % Normalize image

%==========Read segmentation for first frame =========%
LS = 0;

if index == frameA % First frame must have an associated segmentation
if index ~= frame1 % To start from intermediate frame, LS must exist from
                   % a previous run
   load( ['LS num2str(index)'] );

else % Otherwise use the predefined segmentation for the first frame
   LS = imread(['endoSeq' num2str(seqid) '_' int2str(index) 'b.jpg']);

   if size(LS, 3) == 3,
      LS = double( rgb2gray(LS) );
   else
      LS = double( LS );
   end

   LS = LS./max(max(LS)) - 0.5; %Set background to -0.5 and region to +0.5
   LS = LS(:, :, 1); % Only take first layer if LS is an rgb structure

end
end
```

```
% getEigenResponse: calculate eigenvalue line-object response
%
% Luke Windisch
% Ecole Polytechnique
% July 2005
% ------------------------
% Calculates the eigenvalue-based line-object response based on:
% C. Lorenz, I.-C. Carlsen, T. Buzug, C. Fassnacht and J. Weese,
% "Multi-scale Line Segmentation with Automatic Estimation of Width,
%  Contrast and Tangential Direction in 2D and 3D Medical Images," Proc. 1st
% Int'l Joint Conf. CVRMed and MRCAS (CVRMed/MRCAS '97), 233-242, 1997.
%
% INPUTS:  index    - current index into current set
%          scale    - scale at which the response should be calculated
%          saveresp - Save the eigenresponse if set to '1.' Suggested to
%                     speed future processing
%
% OUTPUTS: eigresp - eigenresponse at indicated scale
%
% SEE ALSO: EIGENCALC, IMAGEIN

function eigresp = getEigenResponse( index, scale, saveresp )

global seqid;    % Number of sequence to consider

maxext = 80;    % Extent of blurring filter
gamma = (3/2);  % Compensates for blurring at different scales

global M;
global N;

%========= Read the image ==========%
if ~ exist('im','var'),

   imname = ['endoSeq' num2str(seqid) '_' num2str(index) '.jpg'];
   rgbim = imread(imname);

   if size(rgbim, 3) == 3, % Convert image to gray scale
      im = double( rgb2gray(rgbim) );
   else
      im = double( rgbim );
   end

   [M,N] = size(im);
end

%========= Determine eigen/line response ==========%
f = fspecial( 'gaussian', min( maxext, 4*scale ), scale );
fim = imfilter( im, f, 'replicate' );

[ D1, D2, V1, V2, lineInten, lineRatio ] = eigenCalc( fim, [], [] );

id1 = abs( D1 ) > abs( D2 );  % see idealHessianResp_acrossScale.m
id2 = ones( size(id1) ) - id1; % (in Methodology - Segmentation Figs)
```

```
eigresp = scale^gamma * ( D1.*id1 + D2.*id2 ); % Final eigenresponse

if saveresp,
   save ( ['../Images/EigenSequence' num2str(seqid) '/eigenSeq' ...
         num2str(seqid) '_' num2str(index) '_scale' num2str(scale)], ...
         'eigresp' );
end
```

```
% estimateF - Determine the level set propagation speed
%
% Luke Windisch
% Ecole Polytechnique
% July 2005
% ------------------------
% Uses pixel-wise comparisons with the segmented previous frame to
% determine a Bayesian probability that each pixel in the new frame is a
% region or background pixel. The probability is represented by the
% strength of the propagation speed F that is returned. This speed will be
% used to evolve the level set region between frames
%
% INPUTS:  im0    - previous image (with known region area)
%          im1    - current image (with unknown region area)
%          im0rgn - target object region in im0
%
% OUTPUTS: F - Bayesian-based evolution speed for level set
%
% SEE ALSO: LSTRACK, EVOLVELS

function F = estimateF( im0, im1, im0rgn )

%========= Initializations ===========%
global M;
global N;
global srchwin;
global polarize;

if size( im0, 3 ) > 1, % Normalize colour images
   im0nrm = sqrt(im0(:, :, 1).^2 + im0(:, :, 2).^2 + im0(:, :, 3).^2);
   im1nrm = sqrt(im1(:, :, 1).^2 + im1(:, :, 2).^2 + im1(:, :, 3).^2);
else
   im0nrm = im0;
   im1nrm = im1;
end

V = round( max(max( 1, max(max( im0nrm )) )) );

im0rgn = padarray( sign(im0rgn+1), [srchwin srchwin], 'replicate', 'both' );
LSbck = ones( size(im0rgn) ) - im0rgn;

LSrgn_c = conv2(im0rgn, ones(2*srchwin+1), 'valid');
LSbck_c = conv2(LSbck, ones(2*srchwin+1), 'valid');

pure_rgn = LSrgn_c == ( 2*srchwin+1 )^2;
pure_bck = LSbck_c == ( 2*srchwin+1 )^2;

im0nrm_p = padarray( im0nrm, [srchwin srchwin], 'replicate', 'both' );
im0rgn_c = conv2( im0rgn .* im0nrm_p, ones(2 * srchwin + 1), 'valid');
im0bck_c = conv2( LSbck .* im0nrm_p, ones(2 * srchwin + 1), 'valid');

Vrgn = ( im1nrm - im0rgn_c ./ max(LSrgn_c, 1) ) .^ 2;
Vbck = ( im1nrm - im0bck_c ./ max(LSbck_c, 1) ) .^ 2;
```

```
Vrgn = V^2 * pure_bck + sign(-pure_bck + 1) .* Vrgn;
Vbck = V^2 * pure_rgn + sign(-pure_rgn + 1) .* Vbck;

F = Vrgn - Vbck;

if polarize,
   F = sign(F);
end
```

```
% evolveLS - Level set evolution
%
% Luke Windisch
% Ecole Polytechnique
% July 2005
% ------------------------
% Uses the speed funtion estimate returned by estimateF.m to evolve a level
% function using the numerical scheme described in:
% J.A. Sethian, 'Level Set Methods and Fast-Marching Methods: Evolving
%  Interfaces in Computation Geometry, Fluid Mechanics, Computer Vision,
%  and Material Science,' Cambridge University Press, 1999.
%
% INPUTS:  Un1    - Initial position of level set function
%          F      - Level set propagation speed
%
% OUTPUTS: Un1    - Final position of level set function after iterative
%                   evolution process
%
% SEE ALSO: LSTRACK, ESTIMATEF


function Un1 = evolveLS( Un1, F );

%========== Parameters ==========%
global M;
global N;

global deltaT;  % Time step for LS update
global lambdaL;  % Multiplier to weight curvature prior
global numiter;  % Number of level set evolutions per frame

%=======Update the level set ==========%
for i = 1:numiter,

   Un = Un1; % current level <-- most recently updated level set

   %========== Numerical differences ==========%
   U_Px0y = [Un(:, 2 : N) Un(:, N)];     % u(i+1,j)
   U_Nx0y = [Un(:, 1) Un(:, 1 : N - 1)];  % u(i-1,j)
   U_0xNy = [Un(2 : M, :); Un(M, :)];    % u(i,j-1)
   U_0xPy = [Un(1, :); Un(1: M - 1, :)];  % u(i,j+1)

   U_PxPy = [U_Px0y(1, :); U_Px0y(1 : M - 1, :)];  % u(i+1,j+1)
   U_PxNy = [U_Px0y(2: M, :); U_Px0y(M, :)];      % u(i+1,j-1)
   U_NxPy = [U_Nx0y(1, :); U_Nx0y(1 : M - 1, :)];  % u(i-1,j+1)
   U_NxNy = [U_Nx0y(2: M, :); U_Nx0y(M, :)];      % u(i-1,j-1)

   Ux = (U_Px0y - U_Nx0y) ./ 2; % Spatial derivatives
   Uy = (U_0xPy - U_0xNy) ./ 2;
   Uxx = U_Px0y - 2*Un + U_Nx0y;
   Uyy = U_0xPy - 2*Un + U_0xNy;
   Uxy = (U_PxPy - U_PxNy - U_NxPy + U_NxNy) ./ 4;

   %========= Numerical flux function (based on Engquist Osher ) =======%
```

```
    Dx_pos = U_Px0y - Un; % u(i+1,j)-U(i,j)
    Dx_neg = Un - U_Nx0y; % u(i,j)-U(i-1,j)
    Dy_pos = U_0xPy - Un; % u(i,j)-U(i,j-1)
    Dy_neg = Un - U_0xNy; % u(i,j+1)-U(i,j)

    D0x = Ux; % [U(i+1,j)-U(i-1,j)]/2
    D0y = Uy; % [U(i,j+1)-U(i,j-1)]/2

    flux_frwrd = sqrt((max(Dx_neg, 0).^2 + min(Dx_pos, 0).^2 + ...
       max(Dy_neg, 0).^2 + min(Dy_pos, 0).^2));
    flux_bkwrd = sqrt((max(Dx_pos, 0).^2 + min(Dx_neg, 0).^2 + ...
       max(Dy_pos, 0).^2 + min(Dy_neg, 0).^2));

    %============ Actual update of level set values ==========%
          % Central diff approx. to curavture
    K = (Uxx.*(Uy.^2) - 2*Uy.*Ux.*Uxy + Uyy.*(Ux.^2)) ./ ...
       ( Ux.^2 + Uy.^2 + eps ).^(3/2);

    Un1 = Un + deltaT * (-(max(F, 0) .* flux_frwrd + ...
       min(F, 0) .* flux_bkwrd) + lambdaL * K .* sqrt(D0x.^2 + D0y.^2));

end
```