MINES PARIS
ParisTech

ED n°84 : Sciences et technologies de l'information et de la communication

N° attribué par la bibliothèque
|__|__|__|__|__|__|__|__|__|__|

THÈSE

pour obtenir le grade de
Docteur de l'École des Mines de Paris
Spécialité "Informatique temps réel, robotique et automatique"

présentée et soutenue publiquement par
Tom VERCAUTEREN

le 25 janvier 2008

Image Registration and Mosaicing for Dynamic In Vivo Fibered Confocal Microscopy

Directeur de thèse : Nicholas Ayache,
Equipe-Projet Asclepios, INRIA Sophia Antipolis

Jury

Olivier Faugeras, INRIA Sophia Antipolis	Président
Polina Golland, MIT	Rapporteur
Nassir Navab, TUM	Rapporteur
Nicholas Ayache, INRIA Sophia Antipolis	Directeur
Xavier Pennec, INRIA Sophia Antipolis	Co-directeur
Aymeric Perchant, Mauna Kea Technologies	Examinateur
Valentin Becker, TUM	Invité
Sacha Loiseau, Mauna Kea Technologies	Invité

Résumé

Recalage et Mosaïques d'Images pour la Microscopie Confocale Fibrée Dynamique In Vivo

La microscopie confocale classique permet d'obtenir des images à haute résolution de cellules en culture ou dans un tissu biologique excisé. Cette technologie peut être adaptée aux applications in vivo grâce à l'utilisation de fibres optiques et d'optiques miniaturisées. A terme, la microscopie confocale fibrée devrait permettre aux médecins et biologistes de réaliser des biopsies optiques; c'est à dire un examen histologique, en temps réel, des tissus biologiques à l'intérieur d'un organisme vivant et directement au contact de la zone d'intérêt.

Le but premier de cette thèse est de dépasser les limites matérielles de ces instruments d'imagerie en développant des outils de recalage d'images spécifiques et innovants. En particulier, le propos de ce manuscrit est cadré par l'objectif de proposer, au travers d'outils de création de mosaïques d'images, des biopsies optiques à grand champ aux médecins. Cette application est considérée, dans cette thèse, comme un système, ou un circuit, qui prendrait en entrée un flot de données brutes et délivrerait en sortie des mosaïques d'images à grand champ. Nous détaillons les éléments critiques de ce système, en particulier la reconstruction d'images en temps réel, le recalage linéaire d'images et le recalage non linéaire, avant de présenter la structure du système complet.

Les données brutes produites par la microscopie confocale fibrée sont difficiles à interpréter parce qu'elle sont modulées par la structure en nid d'abeille du réseau de fibres optiques et parce qu'elle sont entachées d'artefacts géométriques. Dans ce contexte, nous montrons qu'une reconstruction en temps réel des images peut être utilisée en pré-traitement afin de produire des séquences vidéos directement interprétables. Comme la microscopie confocale fibrée est une imagerie qui se fait au contact des tissus, le mouvement relatif du tissu par rapport à la sonde optique implique qu'il est parfois difficile d'obtenir de manière robuste certaines mesures quantitatives d'intérêt. Nous avons donc attaqué le problème du recalage linéaire, efficace et robuste de paires d'images. Nous montrons que des outils récents provenant du domaine du contrôle robotique par la vision peuvent surpasser les solutions standards utilisées en analyse d'images biomédicales. L'adéquation de ces outils au problème du recalage linéaire d'images nous a amenés à revisiter le problème du recalage non-linéaire. En interprétant le recalage non-linéaire comme un problème d'optimisation sur un groupe de Lie, nous développons un algorithme rapide de recalage difféomorphe non-paramétrique d'images. En plus d'être difféomorphe, notre algorithme produit des résultats qui sont similaires à ceux de

l'algorithme des démons de Thirion mais qui sont plus lisses et plus proche de la vérité.

Finalement, nous obtenons une boîte à outils de reconstruction et de recalage d'images que nous utilisons pour proposer un algorithme robuste de création de mosaïques d'images qui permette de calculer un alignement globalement cohérent à partir de résultats locaux, de compenser les distorsions liées au mouvement et de retrouver les déformations non-rigides. Par ailleurs, notre algorithme de mosaïques d'images a récemment été incorporé dans un essai clinique multicentrique. Cet essai illustre l'intérêt clinique de nos outils dans le cadre spécifique de la surveillance de l'oesophage de Barrett.

Abstract

Classical confocal microscopy can be used to obtain high-resolution images of cells in tissue samples or cell cultures. Translation of this technology for *in vivo* applications can be achieved by using optical fibers and miniature optics. Ultimately, fibered confocal microscopy should enable clinicians and biologists to perform what can be referred to as an *optical biopsy*: a real-time histological examination of biological tissues in the living organism directly onto the region of interest.

The main goal of this thesis is to move beyond current hardware limitations of these imaging devices by developing specific innovative image registration schemes. In particular, this manuscript is framed by the goal of providing, through video sequence mosaicing tools, wide field-of-view optical biopsies to the clinicians. This targeted application is seen as a pipeline that takes raw data as input and provides wide field-of-view image mosaics as output. We detail the critical building blocks of this pipeline, namely real-time image reconstruction, linear image registration and non-rigid registration, before presenting our mosaicing framework.

The raw data that fibered confocal microscopy produces is difficult to use as it is modulated by a fiber optics bundle pattern and distorted by geometric artifacts. In this context, we show that real-time image reconstruction can be used as a preprocessing step to get readily interpretable video sequences. Since fibered confocal microscopy is a contact imaging modality, the combined movement of the imaged tissues and the flexible optical microprobe makes it sometimes difficult to get robust and accurate measurements of parameters of interest. To address this problem, we investigated efficient and robust registration of pairs of images. We show that registration tools recently developed in the field of vision-based robot control can outperform classical image registration solutions used in biomedical image analysis. The adequacy of these tools for linear image registration led us to revisit non-rigid registration. By casting the non-rigid registration problem as an optimization problem on a Lie group, we develop a fast non-parametric diffeomorphic image registration scheme. In addition to being diffeomorphic, our algorithm provides results that are similar to the ones from Thirion's demons algorithm but with transformations that are smoother and closer to the true ones.

Finally, we use these image reconstruction and registration building blocks to propose a robust mosaicing algorithm that is able to recover a globally consistent alignment of the input frames, to compensate for the motion distortions and to capture the non-rigid deformations. Moreover, our mosaicing algorithm has recently been incorporated within a multicenter clinical trial. This trial illustrates the clinical value of our tools in the particular application of Barrett's esophagus surveillance.

Remerciements

Je tiens tout d'abord à exprimer ma plus vive gratitude à Sacha Loiseau et à Nicholas Ayache; Sacha pour m'avoir accueilli et intégré au sein de Mauna Kea Technologies et Nicholas pour avoir dirigé cette thèse et m'avoir permis d'interagir avec tous les membres d'Epidaure puis d'Asclepios. Sans eux, cette thèse CIFRE n'aurait pas vue le jour.

Sans le dévouement de Xavier Pennec, Aymeric Perchant et Grégoire Malandain qui m'ont encadré tout au long de cette thèse, la qualité des travaux présentés ici aurait été bien moindre. Merci à eux.

Je remercie également très chaleureusement mon jury de thèse pour l'attention que ses membres ont portée à ce manuscrit. Merci tout particulièrement à Polina Golland et Nassir Navab pour avoir pris le temps, au travers de leurs rapports de thèse, d'étudier mon travail en détail. Leurs commentaires éclairés et leurs remarques constructives ont largement contribué à la clarté du manuscrit. Un grand merci à Olivier Faugeras de m'avoir fait l'honneur de présider ce jury. Je tiens aussi à remercier Valentin Becker d'avoir accepté de représenter le monde médical et d'apporter un point de vue applicatif au sein du jury. Enfin, un second merci à Aymeric, Nicholas, Sacha et Xavier pour m'avoir fait l'amitié de faire partie de mon jury de thèse.

Je suis très reconnaissant envers toutes les personnes qui ont contribuées à faire avancer mon projet de thèse d'un point de vue scientifique. Je tiens tout particulièrement à remercier Ezio Malis dont la présentation de l'ESM a fait germer quelques unes des idées présentées ici. Merci à François Lacombe pour ces remarques toujours pertinentes et ses remises en question, à Nicolas Savoire pour ses commentaires éclairés, et au reste de l'équipe R&D de MKT pour leurs interactions constructives (notamment Bertrand, Christophe et Jérôme) Merci à Olivier Clatz et Boon Thye Thomas Yeo pour avoir crut en mes travaux et m'avoir changé les idées en m'impliquant dans un nouveau projet de recherche durant la rédaction de ce manuscrit. Merci également à Xiaodong Wang, Alberto Lopez Toledo et Ta-Hsin Li pour m'avoir initié à la recherche, avant cette thèse, dans un domaine complètement différent de l'analyse d'images biomédicales.

D'un point de vue applicatif, l'aide de Anne Osdoit ainsi que des autres membres des équipe Med et Lab de MKT (notamment Alexandre, Arnaud, Caroline, Charlotte et Pierre-Alix) aura été pour moi inestimable. Merci à Luc Thiberville et Alexander Meining ainsi que Valentin Becker, Raphaël Boisgard, Frédéric Ducongé et Bertrand Tavitian pour leur enthousiasme vis à vis de mes travaux et leurs efforts pour la validation applicative de mes outils. Un grand merci à Claude Collin pour avoir mis en oeuvre les manipulations de validation techniques qui ont fait la force de mes travaux.

Sur un plan technique, je suis également redevable à plusieurs personnes. Tout d'abord, merci à Fabrice Laporte, Florence Portron et à toute l'équipe Dev (Anand, Guillaume, Miriame et JB) de MKT pour leur aide précieuse durant l'essai clinique et le développement du produit *mosaicing*. Merci à Nicolas Toussaint et Pierre Fillard pour avoir intégré certains de mes travaux dans MedINRIA.

Je pense également à ceux qui ont, malgré eux, partagé un peu de ma vie au travail. Merci à Fabien Oberrietter d'avoir essuyé les plâtres de ma première expérience d'encadrement. Merci à Olivier Commowick, Tomaso Mansi et Maxime Sermesant d'avoir été mes co-bureaux pendant quelques temps. Merci à Isabelle Strobant et Florence Guerber d'avoir tant facilité mes différents séjours à Sophia et Paris.

Ayant décidé de faire une thèse entre Sophia Antipolis et Paris, je n'aurais jamais pu m'en sortir sans l'aide de quelques amis. Merci à Adrien, Demian, Florina, François, Heike, Jean-Marc, Nicolas, Ricardo et Sylvain de m'avoir, entre autres, hébergé à un moment ou un autre. Merci également à Marius, et Tristan pour avoir rendu mes passages sur la côte d'azur plus agréables. Merci aussi à toute l'équipe de MKT et d'Asclepios.

Enfin, cette thèse marque pour moi un aboutissement sur le plan personnel. C'est pourquoi je remercie, plus que tout, ma famille et ma très chère Adèle.

Table of Contents

Chapter 1

Introduction

Table of Contents

Foreword This thesis stems from a CIFRE agreement[1] with Asclepios research group at INRIA Sophia Antipolis, `http://www-sop.inria.fr/asclepios`, and the company Mauna Kea Technologies, Paris, `http://www.maunakeatech.com`, which is specialized in the development of dynamic *in vivo* cellular imaging systems for biomedical applications.

In the later half of the twentieth century, imaging technologies like CT scanners and MRIs changed the way doctors treated patients and biologists envisioned their experiments, by offering them extraordinary new tools to look inside the living human or animal body. Unfortunately, the definition of these new imaging tools is limited to providing macroscopic, i.e., relatively large, views of organs and other body parts. This limitation of biomedical imaging has been partly alleviated by relying on more basic tools, like traditional microscopes, to analyze tissue at the cellular level.

For instance, many medical conditions, including cancer, must be diagnosed by removing a sample of tissue from the patient and sending it to a pathologist for examination under a microscope. This procedure is called a biopsy. Being an invasive procedure, it is not always possible for the clinician to take a biopsy sample in every suspicious area. The capacity of traditional biopsy for early cancer detection is therefore limited. More than a hundred million biopsies are performed each year. Each of them require a few days before the pathologist diagnosis gets reported back to the patient. Biopsies are thus also associated with a significant cost. Therefore, although over the past ten years, the progress in biomedical imaging has been almost unbelievable, a simple need remained unanswered. How can we image inside the living organism, without damaging it, at the microscopic level, as life happens?

[1] CIFRE (Convention Industrielle de Formation par la Recherche / Industrial Agreement for Training via Research) agreements aim at fostering innovative processes and technology transfer between public research organizations and industry by supporting a young researcher based in industry, to complete the PhD. They are administered by ANRT (Association Nationale de la Recherche Technique / National Association for Technical Research), `http://www.anrt.asso.fr`.

Recent technological developments have allowed for minimally invasive imaging systems to perform what can be referred to as *optical biopsy*. *Optical biopsy* is a histological examination of biological tissues *in vivo* and *in situ*, that is, directly in the living organism and in contact with the tissue of interest. Optical biopsy does not suffer from the same limitations as traditional biopsy. It could thus potentially improve early cancer detection.

Cellvizio® developed by Mauna Kea Technologies, Paris, is one such optical biopsy tool that has proved its unique capabilities to image cellular architecture *in vivo*. Cellvizio basically relies on putting a microscope objective at the end of an ultra-thin, three-meter-long flexible fiber optics microprobe. Technically it is based on the principle of confocal microscopy. Cellvizio can therefore be referred to as a *fibered confocal microscope*.

The main goal of this thesis is to move beyond current hardware limitations of fibered confocal microscopy (FCM) by developing specific innovative image registration[2] schemes. In particular, we provide wide field-of-view (FOV) optical biopsies to the clinicians through fully automatic stitching of images within large mosaics.

1.1 At Stake: Observing Life at Cellular Level. . . Better

There are many hurdles to overcome to get a genuine optical biopsy system. The raw data that Cellvizio produces is for example difficult to interpret for a common physician or biologist as it is modulated by a fiber optic bundle pattern and distorted by geometric artifacts. The very first objective of image processing in this context should therefore be to reconstruct, in real time, an easily interpretable smooth-motion video sequence. We show that a strong knowledge of the physics of acquisition can lead to the development of such specific image processing algorithms. By combining the hardware with image processing software, the complete Cellvizio system therefore already meets many of the requirements of a genuine optical biopsy system.

Such an imaging device supports a general trend of biomedical imaging. Thanks to technological progress, more detailed and more targeted sets of biologically relevant data can now be acquired. This data provides new evidence to build more personalized methods of diagnosis and decision-making. If the available data gets more accurate and more relevant, it also becomes more and more complex. The classical way of making use of it, mostly based on human interpretation, becomes therefore less and less feasible. Facing the need of a more quantitative and automated path, typical users will greatly benefit from advanced image analysis tools capable of extracting the pertinent information.

This thesis focuses on one particular class of such automated image analysis tools, namely image registration and mosaicing tools. The high resolution images provided by Cellvizio are mostly acquired on living organs with a hand-held optical microprobe. This implies that both natural movements of the imaged tissues and

[2]Image registration is the process of finding how to best align given views of an object.

gestures of the physician or biologist influence the acquired data. Image alignment or registration can help us automatically recover the motion of the imaged objects. It is thus an essential building block of many advanced image analysis problems. Some of the recent advances made in macroscopic medical image registration could be directly applied to microscopic biomedical image analysis. However, the specificity of our imaging modality often requires the development of specialized new image processing techniques. Furthermore, it should be noted that despite remarkable efforts and advances during the past twenty years, the central problem of image registration has not been solved in the general case. It is indeed often an ill-posed problem. Our research efforts have thus been aimed towards the design of suitable solutions for the registration of Cellvizio images.

The first important application of image registration for fibered confocal microscopy presented in this manuscript is given by region-of-interest tracking. In many fields such as gene expression monitoring, drug biodistribution or pharmacokinetics, it is important to quantify biological processes. However, because of the natural movement of the imaged tissue, it is often very difficult to get robust and accurate measurements of the parameters of interest. By focusing on a specific part of the acquired sequence, region-of-interest tracking can help stabilize the images which eases the measurements of the required parameters. When the optical microprobe can be repositioned at a specific location, image registration is also necessary to track the evolution of a pathology and compare series of images taken at distant time instants.

As interesting as dynamic sequences may be during the time of the medical procedure or biological experiment, there is a need for the expert to get an efficient and complete representation of the entire imaged region. Since fibered confocal microscopy is a contact imaging modality, there is also an inevitable hardware trade-off between resolution, field-of-view and invasiveness. In this thesis, we show that video mosaicing techniques can help us move beyond this trade-off by providing a complete wide field-of-view image without modifying the hardware. Video mosaicing techniques also offer a way to bridge the gap between scales. Fusing the information provided by different imaging modalities that have different imaging scales is one of the most challenging tasks in the field of biomedical image analysis. It is also one of the most promising ways of increasing our knowledge about biological processes. If registration between two macroscopic imaging modalities is already an ambitious project, it is nonetheless clear that in order to fuse information across scales, we need a way to extract macroscopic information from the microscopic video sequences. This problem is at the heart of the mosaicing methods we developed.

It is the objective of this thesis to present some of the specificities involved in the registration of Cellvizio images, to develop some advances in the field of biomedical image registration and to show how our cutting-edge image alignment tools can be used for the automatic construction of very large mosaics from video sequences. But, simply stating that the tools we propose can help the clinician or biologist go forward is clearly not convincing without further proof. A great deal of effort has thus been put both into a quantitative validation of our schemes and into a real-life

evaluation of our work. We have incorporated our mosaicing algorithm within a multicenter clinical trial to assess the clinical value of our tools on the particular application of Barrett's esophagus diagnosis. First qualitative results suggest that mosaicing decreases the time required to make a confident diagnosis. It could thus be a cost-efficient alternative representation.

1.2 Contributions and Manuscript Organization

This manuscript is framed by the goal of providing, through video sequences mosaicing tools, wide field-of-view optical biopsies to the clinicians. This targeted application is seen as a pipeline that takes some raw data as input and provides wide FOV image mosaics as output. Our contributions are detailed in five chapters with a bottom-up approach. We start in Chapters 2, 3 and 4 by describing some of the building blocks of the pipeline. We assemble these blocks within an image mosaicing algorithm in Chapter 5 and describe the full mosaicing pipeline with its application in Chapter 6. More specifically, each chapter is organized as follows.

In Chapter 2 we provide a rather detailed description of fibered confocal microscopy and the physics of acquisition involved. The picture we draw of Cellvizio is built on top of the ones we previously published in [Ayache 06, Vercauteren 08a, Vercauteren 07e]. The raw data from Cellvizio is considered as the input to our targeted "wide FOV optical biopsy through mosaicing" application. A first building block towards this goal is given by the *real-time image reconstruction* that allows the user to get readily interpretable video sequences. We show that with such a real-time image reconstruction scheme, FCM provides a good match for a dynamic optical biopsy system. The analysis we draw in this chapter will, however, also help us understand and point out the main limitations of the current approach. The work presented in this thesis has been driven by the aim of moving beyond these restrictions by making use of image registration and video sequence mosaicing tools. In addition to the insight into fibered confocal microscopy, our main contribution in this chapter resides in an extensive analysis of many different real-time image reconstruction schemes. This has led to a change of the scheme used in Mauna Kea Technologies proprietary software.

Our first contribution to the field of image registration appears in Chapter 3. We look at the problem of efficiently and robustly registering pairs of images through a linear transformation. This is, for example, a crucial point for the application of *video sequence stabilization and region-of-interest tracking* that we presented in [Perchant 07] and develop in this chapter. Linear image registration is also a major computational bottleneck in our mosaicing algorithm. We showed in [Vercauteren 07b] that some tools that have recently been developed in the field of vision-based robot control can outperform classical image registration solutions used in biomedical image analysis. This first main contribution will thus help us restrain the computational burden of linear image registration and can be used for many mono-modal registration problems.

For many applications, linear image registration will be sufficient. However, when fine and accurate alignment is required, one needs to take into account the (non-rigid) deformations of the imaged soft tissue. If we only consider linear transformations, our mosaicing algorithm would for example often provide blurred results. In Chapter 4, we develop our second main contribution to the field of image registration. The adequacy of the tools we present in Chapter 3 for linear image registration led us in [Vercauteren 07b] to revisit non-rigid registration. The insight we gained allows us to provide interesting theoretical justifications for the different variants of Thirion's demons algorithm. And more importantly, by casting the non-rigid registration problem into an optimization problem on a Lie Group, we develop in [Vercauteren 07d] and in this chapter a fast non-parametric diffeomorphic image registration scheme. Our algorithm proved to outperform many existing registration approaches even in 3D.

In Chapter 5, we present our main contribution to the problem of video sequence mosaicing. We use the image reconstruction and registration building blocks of the previous chapters to create a robust mosaicing algorithm that can automatically combine successive frames of the acquired video sequence, cancel motion artifacts, and reconstitute wide FOV images of the tissues. To image and explore a region of interest, the confocal microprobe is simply glided along the tissue, either hand-held for small animal research, or through the endoscope bending. This chapter is, *in extenso*, the work we published in [Vercauteren 06] which builds on [Vercauteren 05]. This work also features an important quantitative validation of our mosaicing framework by using rigorous controlled experiments.

In Chapter 6 we continue the evaluation of our mosaicing algorithm for real-life clinical applications that we started in [Becker 07, Perchant 06, Thiberville 07, Vercauteren 07a]. This chapter is devoted to the integration of our mosaicing algorithm into a multicenter clinical trial which is our main contribution in terms of application. Since a fully automatic solution is necessary in this setup, Chapter 6 is also focusing on an optimized engineering approach of the mosaicing pipeline. To feed the mosaicing algorithm with suitable video sequences, we designed an automatic scene-splitting algorithm that segments the acquired sequence into smooth-motion, relatively stable, scenes. The computational cost of the full pipeline is handled by a simple distributed computing model. Another contribution in this setting was to introduce feedback within the mosaicing pipeline. It is indeed difficult, without further aid, for the practitioner to know whether the sequence he is acquiring is a good candidate for the post-processing mosaicing. We have thus designed a real-time algorithm that creates a rough mosaic as long as the motion of the probe is sufficiently smooth. This provides direct visual feedback and teaches the clinicians to acquire smooth motion sequences.

Finally, Chapter 7 concludes this thesis by discussing our contributions and providing various perspectives on our research work.

1.3 List of Publications

This thesis is a monograph, which contains unpublished material. It is however largely based on the following publications:

Descriptions of Cellvizio

[**Ayache 06**] Nicholas Ayache, Tom Vercauteren, Grégoire Malandain, Fabien Oberrietter, Nicolas Savoire and Aymeric Perchant. *Processing and mosaicing of fibered confocal images*. In MICCAI Workshop on Microscopic Image Analysis with Applications in Biology (MIAAB'06), October 2006. Invited

[**Vercauteren 08a**] Tom Vercauteren, Nicholas Ayache, Nicolas Savoire, Grégoire Malandain and Aymeric Perchant. *Processing of* in vivo *fibered confocal microscopy video sequences*. In Jens Rittscher, Raghu Machiraju and Stephen T. C. Wong, editors, Microscopic Image Analysis for Life Science Applications. Artech House, 2008. Approximatively 30 pages. In press

[**Vercauteren 07e**] Tom Vercauteren, Aymeric Perchant and Nicholas Ayache. *In vivo microscopy for real-time structural and functional cellular imaging*. ERCIM News, volume 69, pages 34–35, April 2007

Image Registration Algorithms

[**Perchant 07**] Aymeric Perchant, Tom Vercauteren, Fabien Oberrietter, Nicolas Savoire and Nicholas Ayache. *Region tracking algorithms on laser scanning devices applied to cell traffic analysis*. In Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'07), pages 260–263, Arlington, USA, April 2007

[**Vercauteren 07b**] Tom Vercauteren, Xavier Pennec, Ezio Malis, Aymeric Perchant and Nicholas Ayache. *Insight into efficient image registration techniques and the demons algorithm*. In Nico Karssemeijer and Boudewijn P. F. Lelieveldt, editors, Proceedings of Information Processing in Medical Imaging (IPMI'07), volume 4584 of *Lecture Notes in Computer Science*, pages 495–506, Kerkrade, The Netherlands, July 2007. Springer-Verlag

[**Vercauteren 07c**] Tom Vercauteren, Xavier Pennec, Aymeric Perchant and Nicholas Ayache. *Diffeomorphic demons using ITK's finite difference solver hierarchy*. Insight Journal – ISC/NA-MIC Workshop on Open Science at MICCAI, November 2007. Available online with source code at `http://hdl.handle.net/1926/510`

[**Vercauteren 07d**] Tom Vercauteren, Xavier Pennec, Aymeric Perchant and Nicholas Ayache. *Non-parametric diffeomorphic image registration with the demons algorithm*. In Nicholas Ayache, Sébastien Ourselin and Anthony J. Maeder, editors, Proceedings of the 10th International Conference on Medical

Image Computing and Computer Assisted Intervention (MICCAI'07), volume 4792 of *Lecture Notes in Computer Science*, pages 319–326, Brisbane, Australia, October 2007. Springer-Verlag

Mosaicing Algorithm

[**Vercauteren 05**] Tom Vercauteren, Aymeric Perchant, Xavier Pennec and Nicholas Ayache. *Mosaicing of confocal microscopic* in vivo *soft tissue video sequences*. In James S. Duncan and Guido Gerig, editors, Proceedings of the 8th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'05), volume 3749 of *Lecture Notes in Computer Science*, pages 753–760. Springer-Verlag, 2005

[**Vercauteren 06**] Tom Vercauteren, Aymeric Perchant, Grégoire Malandain, Xavier Pennec and Nicholas Ayache. *Robust mosaicing with correction of motion distortions and tissue deformation for* in vivo *fibered microscopy*. Medical Image Analysis, volume 10, number 5, pages 673–692, 2006. Annual MedIA/MICCAI Best Paper Award 2006

[**Vercauteren 08b**] Tom Vercauteren, Alexander Meining, François Lacombe and Aymeric Perchant. *Real time autonomous video image registration for endomicroscopy: Fighting the compromises*. In Jose-Angel Conchello, Carol J. Cogswell and Tony Wilson, editors, Proc. SPIE BIOS - Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing XV, San Jose, CA, USA, January 2008. SPIE. Full-paper accepted on abstract

Clinical Applications of Mosaicing

[**Becker 07**] Valentin Becker, Tom Vercauteren, Claus Hann von Weyern, Christian Prinz, Roland M. Schmid and Alexander Meining. *High resolution miniprobe-based confocal microscopy in combination with video-mosaicing*. Gastrointestinal Endoscopy, volume 66, number 5, pages 1001–1007, November 2007

[**Perchant 06**] Aymeric Perchant, Tom Vercauteren, Charlotte Cavé and Sacha Loiseau. *Mosaicing of confocal microscopic in vivo soft tissue video sequences improves multimodality approach*. In Proceedings of the Radiological Society of North America (RSNA'06), Chicago, Illinois, December 2006. Abstract only

[**Thiberville 07**] Luc Thiberville, Sophie Moreno-Swirc, Tom Vercauteren, Eric Peltier, Charlotte Cavé and Geneviève Bourg Heckly. In vivo *imaging of the bronchial wall microstructure using fibered confocal fluorescence microscopy*. American Journal of Respiratory and Critical Care Medicine, volume 175, number 1, pages 22–31, January 2007. Chosen for the cover of the AJRCCM paper issue

[**Vercauteren 07a**] Tom Vercauteren, Anne Osdoit, Aymeric Perchant and Sacha Loiseau. *Mosaicing of confocal microscopic video sequences: Larger field of view and still higher resolution!* In Proceedings of the Digestive Disease Week (DDW'07), page AB352, Washington, DC, May 2007. Abstract only

Chapter 2

Fibered Confocal Microscopy

Table of Contents

2.1 Motivation: Aiming at a Genuine Optical Biopsy System

Classical confocal microscopy is an established optical imaging technique that can be used to obtain high-resolution images of cells on tissue samples or cell cultures. Translation of this technology for *in vivo* applications can be achieved by using optical fibers, miniature optics and robust laser scanning approaches. Fibered confocal microscopy, and especially Cellvizio, developed by Mauna Kea Technologies, Paris, allows clinicians and biologists to easily get a real-time view of cellular structures without removal of biological tissue. Ultimately, these tools should enable the practitioner to perform what can be referred to as an *optical biopsy*: a histological examination of biological tissues *in vivo* and *in situ*, i.e., directly in the living organism and in contact with the tissue of interest.

In this chapter, we build on our previous descriptions of fibered confocal microscopy published in [Ayache 06, Vercauteren 08a, Vercauteren 07e]. We detail the

characteristics of a genuine optical biopsy system. The meaning of genuine is obviously arguable. We will advocate using a system that optimizes simultaneously different imaging characteristics. These include spatio-temporal resolution, signal-to-noise-ratio but also user-friendliness and compatibility of the acquired sequences with post-processing algorithms. We will detail how Cellvizio answers those requirements. The goal of this chapter is also to understand the physics of acquisition in order to enhance the complete imaging system, both hardware and software. We will show that even if the raw data is almost unusable for user interpretation or for automated analysis if left untreated, specific software algorithms that perform the image reconstruction task in real time can be developed to provide users with high-quality, smooth-motion video sequences. This makes the image sequences readily interpretable by the professionals who rely on them for diagnosis and make them readily usable for further automated image processing and analysis.

From an algorithmic point of view, most of the material presented in this chapter relies on previous work done at Mauna Kea Technologies [Le Goualher 04, Perchant 05]. One of our contributions here is to provide a fair comparison of the different image reconstruction algorithms. This analysis led to a change of the real-time reconstruction scheme used in the proprietary softwares (ImageCell, Cellvizio-GI software and Cellvizio-Lung software) that Mauna Kea Technologies develops to control the hardware and perform image analysis tasks.

2.2 Applications of Optical Biopsy

Current *in vivo* imaging technologies such as MRI, CT, PET, cytometrics, bioluminescence, fluorescence tomography, high-resolution ultrasound or SPECT are only capable of producing images at resolutions between 30 µm and 3 mm. This range of resolution, while largely acceptable for a wide scope of applications, is insufficient for cellular level imaging. On the other side of the resolution range, we find several types of microscopy. The vast majority of microscopes, whether conventional or confocal, are limited for use with cell cultures or *ex vivo* tissue samples. The tiny fraction of microscopes that are dedicated to *in vivo* use, and that can function inside the living organism, are called intravital microscopes. These apparatuses are cumbersome, difficult to use and restricted to research applications on small animals. They also require a very delicate and specific preparation of the animal which includes installing a window on the animal through which the microscope can look into the body.

As we will see, a promising tool to fill this gap and perform optical biopsy is given by fibered confocal microscopy (FCM). Other approaches are obviously possible but none have reached the production stage yet [Jung 04, Oh 06]. We will now focus on a set of applications that are limited by classical biomedical imaging tools and can benefit from optical biopsy.

2.2.1 Observing Life at the Cellular Level as it Happens

In small animal imaging alone, the scope of use of optical biopsy is tremendous. Research in this domain can not only yield new therapeutic solutions or medication. It can also play a pivotal role in further understanding and treating a variety of cellular pathologies including cancer, Alzheimer's and Parkinson's diseases among others. We shall briefly review some of the applications of Cellvizio for small animal imaging[1].

Figure 2.1: The Leica FCM1000 microscope (Cellvizio technology for small animal imaging distributed by Leica microsystems) used to image peripheral nerves of a mouse.

In the field of mouse neuroscience for example, new models of green fluorescent protein (GFP) and yellow fluorescent protein (YFP) animals as well as new fluorophores, that can label subsets of neurons durably, have recently been combined with acute or chronic brain slice preparation for classical microscopic fluorescence imaging. This has led to new insights into the processes occurring in rodent brains. Furthermore, multi-photon microscopy has been used for dynamic *in vivo* studies. To date, multi-photon imaging has, however, been able to acquire only one or two images per second at depths up to 500 µm in the adult mouse brain. Deep cortical structures and sub-cortical areas remain out of reach. *In vivo* deep brain imaging is therefore limited to non invasive yet less effective techniques like MRI and PET whose spatio-temporal resolutions are unfit. In [Vincent 06], the authors have shown that Cellvizio has the spatial resolution to image various neural structures in the living animal, the consistency needed for quantitative evaluation of axonal degeneration and regeneration of peripheral nerves, and the sensitivity to detect calcium transients on a sub-second timescale.

Another example of applications of optical biopsy can be seen in the field of regenerative medicine. This emerging field considers the repair or replacement of tissues and organs by incorporating the use of cells, genes and other biological building blocks along with bio-engineered materials and technologies. Although regenerative medicine is by definition developed towards clinical applications, re-

[1]Mauna Kea Technologies Cellvizio product range dedicated to small animal imaging is now distributed by Leica Microsystems under the name Leica FCM1000 microscope, `http://www.leica-microsystems.com/FCM1000`

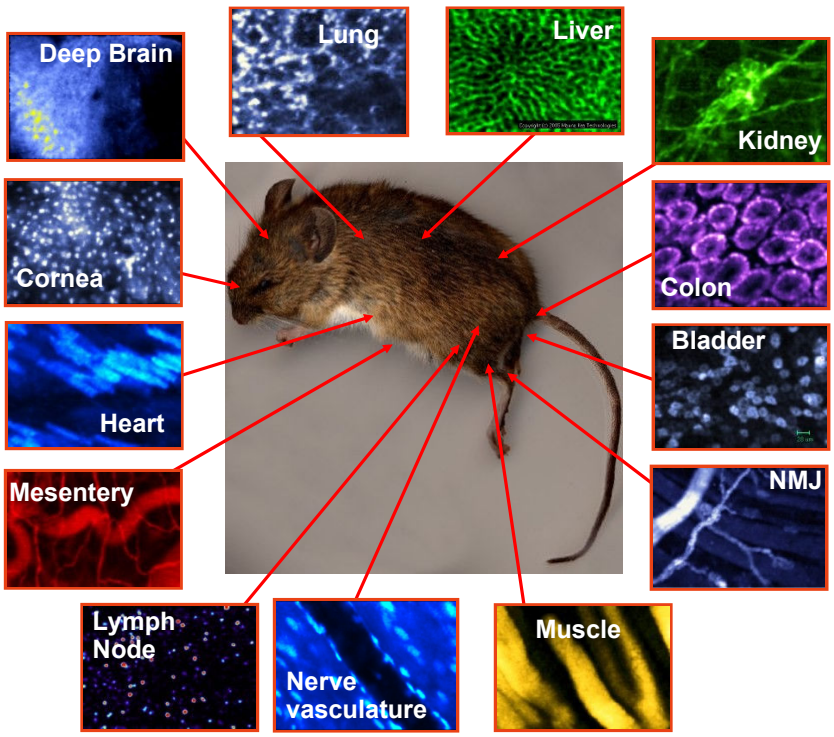


Figure 2.2: Different types of images acquired with Leica FCM1000 microscope (Cellvizio technology for small animal imaging distributed by Leica microsystems).

search is done as a first step on animal models to validate concepts and procedures. For *in vitro* evaluation of the quality of the future grafts, there is a clear need for a minimally invasive high resolution imaging system that can track labeled cells integrated in a scaffold. Such a system is also critical *in vivo* to monitor transplants over days and months during longitudinal studies. The ideal imaging technology (hardware and labeling) should be at the same time biocompatible, safe, non-toxic and non-invasive. It should imply no genetic modification or perturbation of the cells. It should be able to detect single cells at any anatomical location and quantify the number of cells. There should be minimal dilution of signal with cell division, minimal transfer of contrast agent to non-cells. Ideally, no additional contrast agent should be required. In this context, the current standard screening method is to take a tissue sample *in vivo* and have it analyzed under the microscope. Current histological methods can take as long as 24 hours. They also require the animal to be sacrificed at each measurement to get conclusive results, which obviously prevents longitudinal monitoring. In contrast, fibered confocal microscopy has been shown to meet many of the criteria listed above. Cellvizio therefore offers a compelling opportunity for cell tracking. It enables the operator to see the tissue of interest in real time, to make an almost immediate evaluation. Cellvizio also allows one to track specific cells *in vivo* over several days provided one manages to image the same region.

Let us finally mention that the use of small animal *in vivo* imaging systems has led to multiple advances in drug discovery and the study of diseases. Pharmaceutical companies are thus using them increasingly to assess the effectiveness of new drugs. However, classical biomedical imaging devices do not always allow for fine scale measurements. In the field of cancer therapy, scientists have been trying to overcome the limitations and toxicity of radiotherapy and chemotherapy by fighting neo-angiogenesis. Angiogenesis is the formation of new blood vessels and only happens during embryonic development, wound healing and tumor growth. Compared to traditional therapies, angiogenesis inhibition has the advantage of not targeting the tumor itself. Thus, it does not increase tumor aggressiveness by a genetic selection of cells. Furthermore, the toxicity is limited to new blood vessels which limits side effects. However, the interest for angiogenesis therapies has been hindered by the difficulties researchers meet in imaging and characterizing angiogenesis. Assessing the efficiency of these therapies is indeed not straightforward as opposite effects are visible. Since the aim is not to kill tumoral cells directly, the traditional definitions of appropriate dose and maximal supported dose become irrelevant. Therefore, to determine the optimal dose, one needs a thorough assessment of the therapy impact on the tumor vascularization. On tumors growing in window chambers or implanted subcutaneously, intravital microscopy yields valuable information. Nevertheless, it is limited by its access capabilities [McDonald 03]. In this context, Cellvizio offers a new way to image and characterize tumoral angiogenesis with the advantages of requiring very little preparation and allowing any anatomical location to be imaged over time [Laemmel 04].

2.2.2 Early Cancer Detection

Cancer is a group of diseases characterized by uncontrolled growth and spread of abnormal cells. It is the second leading cause of death worldwide. This simple definition of cancer makes it quite clear that cells play a key role in the different stages of cancer development. Some ninety percent of cancers are epithelial. These cancers are preceded by a curable, pre-cancerous and non-invasive stage that progresses without symptoms over a period of years before reaching a cancerous and invasive stage. In the very first step of epithelial cancer, anomalous cells first appear in the deepest layer of the epithelium, directly above the basal membrane. The basal membrane separates the epithelium from the deeper layers of the tissue. It provides a very strong and effective protection. It is located approximately at 100 µm deep from the tissue surface for malpighian epithelium, such as cervix epithelium, and 300 µm for glandular epithelium, i.e., tissue that contains secretion glands such as colon, pancreas and thyroid. Because there are no blood vessels in the epithelium, epithelial cells cannot spread to other parts of the body. Hence the importance of detecting an anomaly at a very early stage before the cancer becomes invasive, i.e., before the basal membrane is broken.

Since cancer is a disease that affects cells and starts below the surface of the tissue, its early diagnosis requires a subsurface visualization of the tissue at the

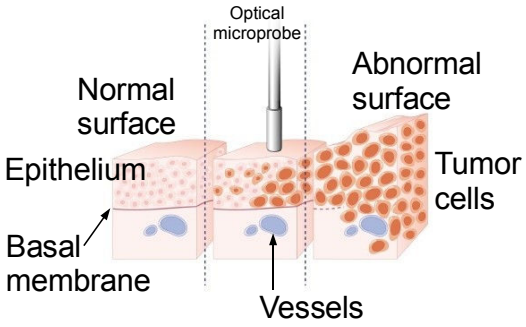


Figure 2.3: Cancer Progression.

cellular level. As conventional imaging techniques do not allow for a subsurface cellular visualization of the tissue during a clinical procedure, standard cancer detection protocols are not straightforward.

For epithelial cancers, i.e., most cancers affecting solid organs, the current medical diagnosis procedure is to take a tissue sample, or biopsy, and to have it examined under the microscope by a pathologist. Most of these biopsy procedures are performed via endoscopy. An endoscope allows for the visualization of the tissue surface at the macroscopic level. It can neither see below the surface nor provide a microscopic view of the tissue. Because of these limitations, biopsies have to be performed without a relevant visual guide.

Several systems are under study to help the endoscopist make an informed decision during the diagnostic endoscopic procedure. Fluorescence spectroscopy can be used to detect dysplasia and early carcinoma based on the analysis of fluorescence spectra [Bourg-Heckly 00]. Drawbacks of fluorescence spectroscopy lie in the lack of morphological information, i.e., no cell architecture is available from this modality, and the significant rate of false positives due to inflammatory processes. Chromoendoscopy, i.e., endoscopy with a topical contrast agent, combined with magnification endoscopy has become popular as a diagnostic enhancement tool in endoscopy [Sharma 03]. In particular *in vivo* prediction of histological characteristics by crypt or pit pattern analysis can be performed using high magnification chromoendoscopy [Kudo 01]. One drawback of this technique is that it cannot provide simultaneously a macroscopic view, for global localization, and a zoomed image.

Further innovations for better differentiation and characterization of suspicious lesions, such as autofluorescence endoscopy and narrow band imaging, are currently under investigation. However, for targeting both biopsies and endoscopic resection and improving patient care, the ideal situation is to characterize tissues completely *in vivo*, and thus to visualize cellular architecture. This implies the availability of microscopic imaging during endoscopic examination as provided by Cellvizio. The confocal nature of this technology makes it possible to observe subsurface cellular structure which is of particular interest for early detection of cancer. This

technology can serve as a guide during the biopsy and can potentially perform optical biopsies by making is a high resolution non invasive optical sectioning within a thick transparent or translucent tissue [Meining 07a].

2.2.3 Microscopic Imaging of Breathing Lung

With advances in imaging, both computed tomography (CT) and positron emission tomography (PET) are playing an increasing role in lung pathology diagnosis. Imaging with CT is, for example, very useful in the evaluation of lung cancer, but it lacks the sensitivity and specificity to firmly establish a diagnosis on its own. A biopsy is thus necessary to confirm a diagnosis such as cancer and to identify the specific type of cancer. Most of the time, biopsies will be performed during a bronchoscopy which involves the use of a bronchoscope to directly view the airways into the lungs.

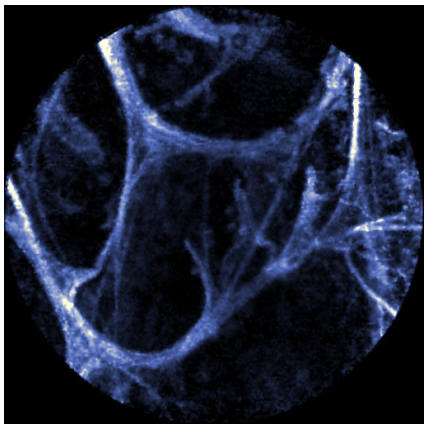


Figure 2.4: Healthy alveolar sac network acquired *in vivo* using Cellvizio in autofluorescence mode (no exogenous fluorophores), FOV: 600 × 600 µm. Courtesy of Pr. A. Ernst, Beth Israel Deaconess Hospital, Boston.

Visualization of the respiratory system with bronchoscopy has been limited so far to the central bronchi due to the narrowing diameter of the terminal bronchi and to the size of existing endoscopes. This major limitation has prevented clinicians from gaining a better understanding of peripheral lung cancers and diffuse interstitial diseases. To examine areas of the lungs that are not accessible during a bronchoscopy, physicians may perform a needle biopsy, i.e., a rather invasive biopsy done from the outside through the chest wall.

In [Thiberville 07], we performed FCM imaging and biopsies guided by white-light and autofluorescence bronchoscopy. We showed that Cellvizio makes it possible to produce clear microscopic images of the subepithelial *lamina reticularis* of the bronchial and bronchiolar wall. Because fibered confocal microscopy takes advantage of naturally occurring endogenous fluorescence, no additional preparation of the patient is needed. We showed that the main endogenous fluorescent signal originates from the elastin component of the bronchial wall.

Since the flexible optical microprobe of Cellvizio can be passed through the operating channel of any bronchoscope, its size allows it to go further than the bronchoscope by gently pushing it down the airways. These specifications enable, for the first time, imaging of the alveoli with microscopic resolution at near video rate. Dynamic processes can thus be observed for *in vivo* histology.

2.3 Principles of Fibered Confocal Microscopy

From all the possible applications we presented here, one can suspect that the ideal characteristics or specifications of a system dedicated to optical biopsies are numerous. The resolution should not exceed a few microns to make it possible to distinguish individual cells and possibly sub-cellular structures. And above all, the system should be easy to operate, in complete adequacy with the current clinical practice. It should not modify the clinician's procedure and practice. Following a short learning stage due to the manipulation of a new instrument and to the interpretation of images never obtained before in such conditions, no more adaptation should be needed in the clinical setting. In particular, the parts of the system meant to be in contact with the patient should be easily disinfected, using standard procedures. Such characteristics are crucial for the development of a system for use in routine practice and not just in clinical research.

2.3.1 Confocal Microscopy

Confocal microscopy enables microscopic imaging of untreated tissue without previous fixation and preparation of slices and thus meets some of the operational-ease requirements as well as the resolution requirements. As shown in Fig. 2.5, the technical principle is based on point-by-point imaging. In a laser scanning confocal microscope, a laser beam is focused by the objective lens into a small focal volume within the imaged sample. A mixture of emitted fluorescence light as well as reflected laser light from the illuminated spot is then collected by the objective lens. The detector aperture obstructs the light that is not coming from the focal point, thereby suppressing the effect of out-of-focus points. Depending on the imaging mode the detector either measures the fluorescence light or the reflected light. The measured signal represents only one pixel in the resulting image. To get a complete image and perform dynamic imaging, the imaged sample has to be scanned in the lateral plane for 2D imaging as well as the axial plane for 3D imaging.

Confocal microscopy can be adapted for *in vivo* and *in situ* imaging by, schematically, inserting a fiber optics link between the laser source and the objective lens.

Fluorescence Mode

The most commonly used mode of confocal microscopy relies on fluorescent light. Fluorescence is an optical phenomenon in which the molecular absorption of a photon triggers the emission of another photon with a longer wavelength. There

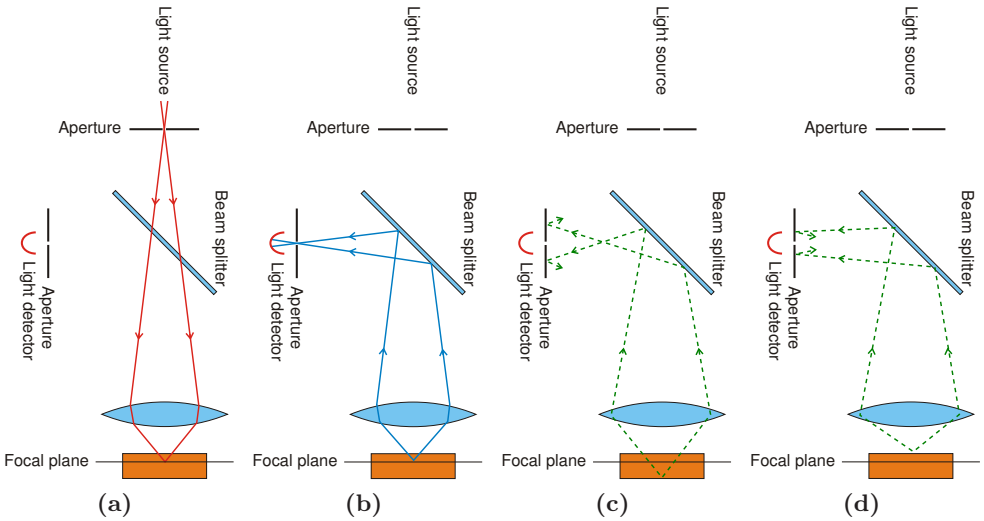


Figure 2.5: Schematic principle of confocal microscopy. (a) The laser beam is focused in the focal volume or imaged point. (b) The light that comes back from the imaged point is focused onto the detector. (c) and (d) The light that comes back from other areas than the focal volume is obstructed by the aperture or pinhole.

is a wide range of applications for fluorescence in the field of biomedical imaging. Biological molecules can be tagged with a fluorophore, and the fluorescence of the tag enables sensitive and quantitative detection of the molecule.

A variety of fluorescent dyes and labels may be used for *in vivo* imaging. Vital dyes can for example stain cellular membranes. Some dyes can be used to monitor physiological processes such as changes in calcium levels. Genetically engineered fluorescent protein tags such as green fluorescent protein (GFP), derived from the jellyfish *Aequoria Victoria*, are used to label proteins. These dyes allow specific organelles and cell components to be labeled, and changes in them to be followed over time. Let us also mention that some biological tissues present some endogenous fluorescence which allows for fluorescence imaging without further staining.

When a confocal microscope is used with a fluorescent tissue, the light that comes back from the imaged point is a mixture of emitted fluorescence light as well as reflected laser light. The emitted light is uncorrelated with the reflected light and has a different spectrum. By putting a chromatic filter in front of the detector, it is thus possible to consider only the information from the labeled tissue. Note that if several fluorophores that have non-overlapping emission spectra are used to stain different kinds of tissue, it is possible to get multi-channel imaging where each channel relates to a distinct type of information. In the remainder of this thesis, we will however focus on single-band imaging.

In [Wilson 90] it is shown that in the fluorescence mode, the intensity I that is measured by the detector is proportional to the concentration of the fluorophore, α_{fluo} convolved by the squared effective point spread function (PSF) h_{eff} of the

system:

$$I \propto |h_{eff}|^2 \star \alpha_{fluo}, \tag{2.1}$$

where h_{eff} depends on the PSF of the system at the two wavelength involved, the size of the pinhole and the optical aberrations.

Reflectance Mode

For most clinicians and biologists, confocal microscopy implies fluorescence and consequently fluorescent labeling. Fluorescent staining allows for imaging of specific tissues with a very good contrast and can even enable functional imaging. The major drawback is, however, that the fluorophores that have to be used require specific tissue preparation or, for small animal imaging, genetic bio-engineering of the animal model. Furthermore, the staining operation should often be non-toxic. In the clinical setting, the non-toxicity constraint results in the fact that the number of approved and useful fluorophores can be counted on one hand.

Despite this common belief, confocal systems can also be used to detect the light that has been reflected by the tissue. Reflectance confocal microscopy has no need for dye or tissue preparation and can readily provide images of morphological architecture. It thus seems perfectly fit for non-invasive clinical use. Another advantage of not using dyes is that the imaging system cannot suffer from photobleaching problems, i.e., photochemical destruction of the fluorophore.

The drawback of reflectance confocal microscopy is that it is mostly limited to morphological imaging. This modality also suffers from a poor signal-to-noise ratio and has a high background signal. The most striking problems come from the internal reflections of the laser beam. Since we can no longer use the spectral properties to separate the light reflected by the tissue from the internally reflected laser light, we have to resort to using higher quality optics and use more technically involved light separation schemes, such as using time of flight information.

If we look at the image formation equations, it is shown in [Wilson 90] that the intensity I that is measured by the detector depends on the transmission capacity γ_{refl} of the image tissue and the effective PSF h_{eff}:

$$I \propto |h_{eff} \star \gamma_{refl}|^2, \tag{2.2}$$

where h_{eff} and γ_{refl} are both complex functions. It can thus be seen that this system is not linear and that, strictly speaking, we should take care of this non-linearity when designing image-processing schemes for this modality.

In spite of the drawbacks that we mentioned, it appears that a reflectance confocal microscope adapted for dynamic *in vivo* imaging would meet many of the requirements of a genuine optical biopsy system for the clinic.

2.3.2 Distal Scanning Fibered Confocal Microscopy

First attempts for developing fibered confocal microscopes (FCM) have historically been made by teams coming from the world of microscopy. The well-known constraints of confocal microscopy were directly transposed to specific architectures

to develop new *endomicroscopes*. With a similar technological core, these designs were in majority based on distal scanning schemes[2]. In such a scheme, one single fiber transports the laser from the proximal light source to the distal end and back. A scanning mechanism is installed at the distal tip, in contact with the tissue, to perform a laser scanning and acquire an image. Different architectures have been investigated by several academic groups and commercial companies: distal fiber scanning, distal optical scanning and micro-electromechanical systems (MEMS) distal scanning. The great advantage of this distal technology is illustrated by its very good lateral resolution: almost that of non-fibered systems. Indeed, the systems developed by Optiscan [Hoffman 06], Olympus [Inoue 05, Murakami 03], Stanford University [Dickensheets 96, Wang 03] and others produce very crisp images through only one optical fiber. These systems have first been driven by technological solutions already existing in the microscopy field. Their ability to obtain high resolution images is one of their strengths. However, image quality is just one clinical requirement among many different needs absolutely necessary for performing *in vivo* microscopic imaging, such as miniaturization, ease of use and real-time imaging. Therefore, one can point out that distal scanning solutions are not able to meet all the demands of a clinical routine examination, even if they are considered as very good imaging solutions.

2.3.3 Proximal Scanning Fibered Confocal Microscopy

To circumvent the problems of distal scanning FCM, a number of teams have tried to use a proximal scanning architecture [Lane 00, Sokolov 03, Sung 02] or a mix of proximal and distal scanning [Sabharwal 99, Yang 05]. In this work, we use a second generation confocal endoscopy system called Cellvizio developed by Mauna Kea Technologies (MKT), Paris, France. MKT's adaptation of a confocal microscope for *in situ* and *in vivo* imaging can be viewed as replacing the microscope objective by a flexible optical microprobe. The length and diameter of the microprobe has been designed both to be compatible with the working channel of any flexible clinical endoscope and to minimize invasiveness when accessing remote organs of small animals. A fiber bundle composed of tens of thousands of fiber optics is used as the link between the proximal scanning unit and the microscope objective, remotely placed at the tip of the flexible optical microprobe. The schematic principle of fibered confocal microscopy is presented in Fig. 2.6 and typical images appear in Fig. 2.2.

Such a proximal scanning choice has many advantages on the application side as shown for example in [Flusberg 05, Helmchen 02]. Decoupling the scanning function from the imaging one allows for the optimization of both functions independently. The distal optics can thus be miniaturized down to a volume much smaller than what distal scanners can reach now, with very little compromise between optical quality and size. Since the scanner can be allocated an arbitrary volume, well known

[2]It is common to refer to laser scanning mechanisms as being either proximal or distal to the light source

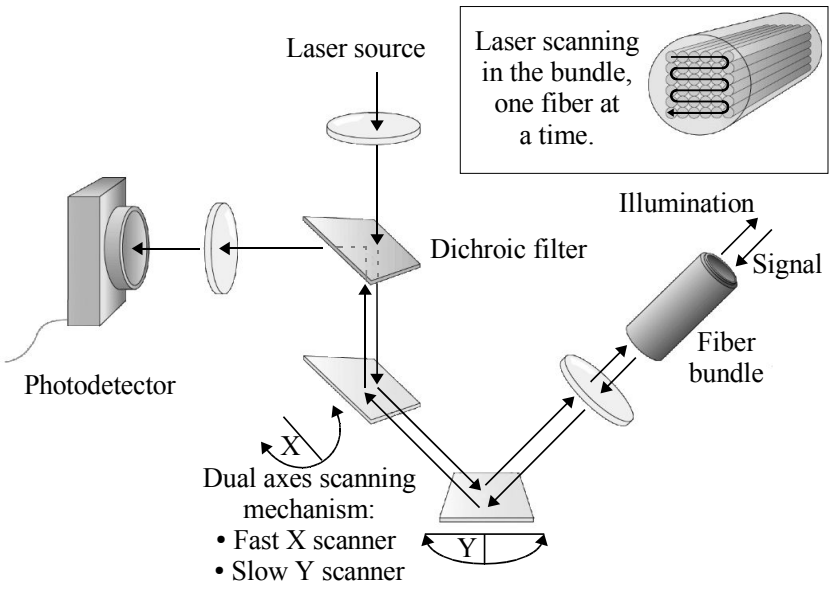
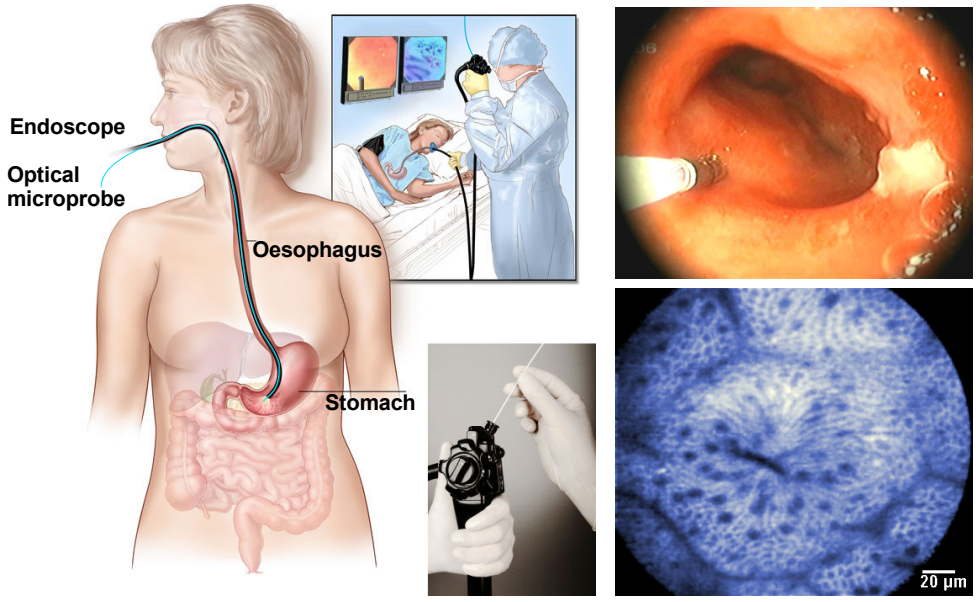


Figure 2.6: Schematic principle of fibered confocal microscopy.

and reliable rapid scanning solutions such as resonant or galvanometric mirrors can be used. A purely passive optical microprobe is also more compatible with cleaning or decontamination procedures that regular clinical or biological use requires. Last but not least, fiber bundles are already existing products, available on the market under different designs and specifications. There is no need for specific developments before using them in biomedical devices. Their association with simplified distal ends should enable their manufacturing at a relatively low cost, opening the way to the production of disposable items.

Cellvizio makes it possible to observe subsurface cellular structures with an optical section parallel to the tissue surface at a depth between 0 and 100 µm. The imaging depth cannot be controlled on a single optical microprobe but depends on the specific optical microprobe used. Therefore, the physician or biologist will be using different optical microprobes, with different technical specifications, for different applications. Confocal image data is collected at a rate of 12 frames per second. The smallest lateral and axial resolutions available are 1 µm and 3 µm. Fields of view ranging from 130×130 µm to 600×600 µm can be obtained thanks to a set of flexible optical microprobes whose diameters vary from 0.16 to 1.5 mm. Given their diameters, the microprobes can be inserted through the working channel of any endoscope. In the clinical setting, FCM is thus typically used in conjunction with an endoscope. Both macroscopic (endoscope image) and microscopic view (FCM image) are available at the same time as shown in Fig. 2.7. This dual view facilitates the selection of the area to be biopsied.

Of course, this approach has well known drawbacks. Even if it is simplified by the absence of a distal scanner, proper light focusing and collection at the distal end of the probe remains a critical issue, in particular when reaching very small



(a) (b)

Figure 2.7: Illustration of a typical Cellvizio-GI setting. (a) The miniaturized flexible optical microprobe is small enough to fit into the accessory channel of any flexible endoscope. The optical microprobe is made to be used like biopsy forceps, its flexibility allows most anatomical configuration of the endoscope. (b) Wide field-of-view imaging remains available when the optical microprobe is in contact with the tissue and microscopic imaging comes in.

optics dimensions. Because of the passivity of the fiber optic bundle, the system also suffers from some loss of resolution and is limited to 2D imaging. Nonetheless, the most widely reported problem of this approach is certainly related to the image itself. It normally shows a strongly visible honeycomb pattern and also often suffers from other artifacts and aliasing effects.

In the next section, we show that specific image processing schemes can be designed to cope with these artifacts in real time. The control and acquisition software that comes with the Cellvizio is thus an inherent part of the imaging device and helps it move towards a true optical biopsy system.

2.4 Real-time Fiber Pattern Rejection

The specific imaging modality we focus on raises specific image processing problems. The non-uniform honeycomb pattern and the geometric distortions that appear on the raw, unprocessed, data, makes it impractical for user interpretation or for automated analysis if left untreated. Algorithms that take on the image reconstruction task in real time have thus been developed to provide users with high-quality, smooth-motion video sequences. These sequences become effortlessly interpretable by the professionals who rely on them for diagnosis and are readily usable for fur-

ther automated image processing and analysis. Most of the available methods are only focused on the removal of the honeycomb pattern [Elter 06, Winter 06] and often only imply performing a simple low-pass filtering. The approach of Mauna Kea Technologies not only removes the honeycomb pattern but also recovers the true signal that comes back from the tissue and compensates for the geometric distortions [Le Goualher 04, Perchant 05].

2.4.1 Calibrated Raw Data Acquisition

Proximal implementation of the scanning function enables the use of very robust and reliable solutions for a fast and accurate scanning. The laser scanning unit uses two mirrors to scan the proximal surface of the flexible optical microprobe with the laser source. Horizontal line scanning is done using a 4 kHz oscillating mirror while a galvanometric mirror handles frame scanning at 12 Hz. A custom synchronization hardware controls the mirrors and digitizes, synchronously with the scanning, the signal coming back from the tissue using a mono-pixel photodetector. Cellvizio scanning reproducibility is better than one half of a fiber diameter. This performance enables the calibration of the continuous motion of the illuminating spot. The sine-wave shape of the geometric distortion due to the resonant mirror (fisheye-like effect) can thus be compensated. This correction permits a comprehensive rectangular mapping of the field-of-view. This is mandatory for any metric interpretation using the images, or any complex combination of individual frames. The proximal scanning is associated with optimized optics that guarantees a very good injection of the laser within the fibers.

When organized according to the scanning, the output of the FCM can be viewed as a raw image of the surface of the flexible optical microprobe. Scanning amplitude and signal sampling frequency have been adjusted to perform a spatial *oversampling* of the fiber bundle in the sense that the raw images have many more pixels than the number of fibers in optical microprobe. This is clearly visible on the raw image in Fig. 2.8 where one can see the individual fibers composing the bundle. Such an oversampling is needed to be able to distinguish the signal coming from each individual fiber. A typical fiber bundle is composed of $30,000$ fibers, with a fiber inter-core distance d_{ic} of 3.3 µm, and a fiber core diameter of 1.9 µm. Fiber arrangement is locally quasi hexagonal, but does not show any particular order at larger scales.

Critical elements in the raw image formation process lie in a correct spatial sampling of the flexible optical microprobe but also in the adjustment of the point spread function (PSF) of the system with this spatial sampling. We indeed want to avoid aliasing on the tissue side. When analyzing the system from the point of view of the sampling theory, the PSF corresponds to the low pass filter and the fiber optics bundle to the sampling grid. The Nyquist frequency is then given by $(M/d_{ic})/2$ where M is the magnification of the optical head and typically ranges from 1.0 to 2.5. The PSF of the system should therefore satisfy this frequency. As a rule of thumb the PSF width should approximately be the Nyquist period. The

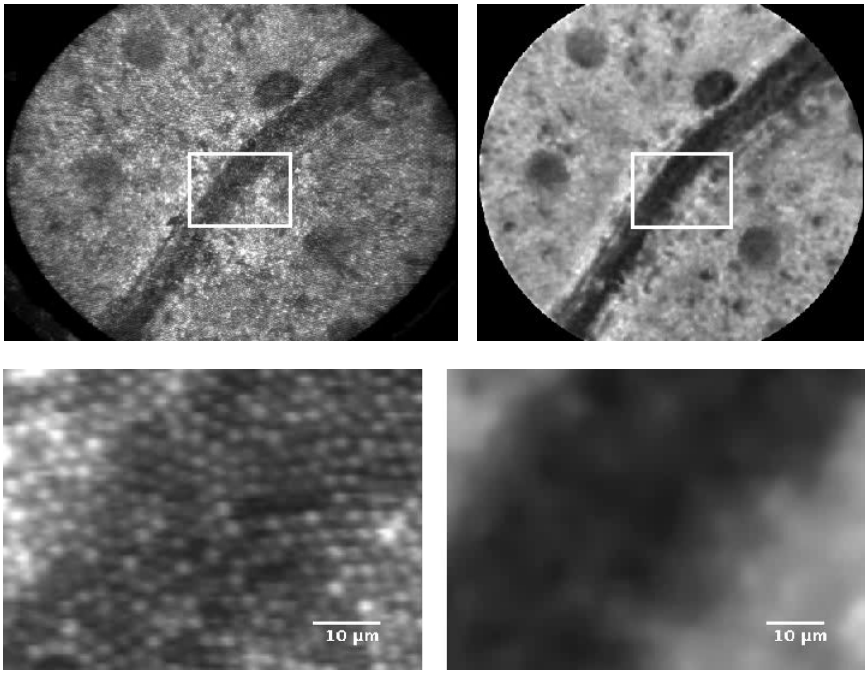


Figure 2.8: Autofluorescence FCM images of a *Ficus Benjamina* leaf. Left: Raw data. Right: Reconstructed images. Top: Complete images. Bottom: Zoom on rectangle. Note that the non-uniform honeycomb modulation and the geometric distortions on the raw data have been corrected in the reconstructed image.

resulting lateral resolution of such a system is then given by $2.d_{ic}/M$. For a user to benefit from the full spatial resolution allowed by the sampling theorem, an optimal fiber pattern removal and geometric distortion compensation scheme is needed.

2.4.2 From Raw Data to Irregularly Sampled Images

The task of the on-the-fly image reconstruction module is to restore, at a rate of 12 frames per second, the true physical signal from the raw data by removing the fiber bundle honeycomb modulation and the scanning distortion.

As mentioned previously, the fiber inter-core distance and the PSF of each fiber have been tailored to basically meet the sampling theorem requirements. We can thus consider that each fiber of the bundle provides one and only one sampling point on the tissue. Associated with these sampling points comes a signal that depends on the imaged tissue and on the single fiber characteristics. The role of the image processing is first to build a mapping between the FCM raw image and the fibers composing the flexible optical microprobe. Once this mapping is obtained, characteristics of each fiber are measured and the effective signal coming back from the tissue is estimated. We then have non-uniformly sampled frames where each sampling point corresponds to a center of a fiber in the flexible optical microprobe. In this section, we will give an overview of the techniques employed to go from the raw data to an irregularly sampled image. We refer the reader to

[Le Goualher 04, Perchant 05] for more detailed information.

Calibration

As a preliminary step, a mapping between the FCM raw image and the fibers composing the flexible optical microprobe is built. This is achieved by using segmentation algorithms that have specifically been designed for our fiber optics bundle. After segmentation, we have access to the position of the fibers within the raw image and to the scanning distortion. Therefore, it becomes possible to create a point set where, for each fiber in the bundle, we have one and only one point whose coordinates are given by the exact, distortion-free, position of the corresponding fiber in the bundle.

This mapping allows us to compute the signal measured by a single fiber. When combined and averaged together, the 15 to 50 pixels corresponding to one single fiber lead to a signal I with a much better SNR.

Since the signal that is measured by a single fiber depends on the imaged biological sample signal α_{tissue} (fluorescence or reflectance) and on the fiber itself, we need to estimate some characteristics of each single fiber such as its gain and background (Raman diffusion, autofluorescence or reflectance). For this purpose, the procedure is to acquire an image of a neutral sample, $\alpha_{tissue} = 0$, such as water, and compute the background signal I_b measured by each fiber. An image of a sample of constant signal, $\alpha_{tissue} = \text{const}$, is also acquired and the signal I_s associated to each fiber is computed.

Imaging Model

In fluorescence mode, the relationship between the actual biological sample fluorescence of interest α_{tissue} and the raw signal I measured by each fiber of the flexible optical microprobe, is given by the model from [Le Goualher 04]:

$$I = I_0.\big(a.\tau_{inj}.\tau_{col}.\alpha_{tissue} + b.\tau_{inj}.\alpha_{back}\big) \tag{2.3}$$

where a and b are constants, τ_{inj} and τ_{col} are the injection rate and collection rate of the fiber, α_{back} takes into account intrinsic characteristics of the fiber such as autofluorescence or Raman diffusion, and I_0 is the intensity of the laser source.

Given the raw signal and the calibration data, the true physical measure α_{fluo} we would like to have cannot be directly estimated. It is however possible to recover it up to a scaling factor with the following equation:

$$I_{restored} = \frac{I - I_b}{I_s - I_b} = K.\alpha_{tissue}, \tag{2.4}$$

where K is a constant independent of the considered fiber. As mentioned earlier, in reflectance mode the model should be more complex. However, (2.3) still provides a first order approximation that enables adequate image reconstruction in practice.

In any case, the calibration model (2.4) cannot account for all the physical phenomenon. Moreover, the processing steps (fiber segmentation, signal estimation)

involved in computing the required quantities (positions of the fibers, background and foreground signal, pre-restoration fiber signal) are rather complex. It is thus necessary to evaluate the accuracy of this distortion-corrected irregularly sampled image extraction scheme. In this case a simple procedure could be used to evaluate the ability of the algorithm to recover correct intensities. A set a fluorescent samples of known concentration were prepared. The irregularly sampled image extraction was performed on all the samples. For each sample, the consistency of the extracted fiber measurements was evaluated. A flat image is indeed required. It was also possible to check that across the different samples, the extracted intensities were proportional to the known fluorophore concentration. To evaluate the correctness of the geometric distortion compensation scheme, the recovered fiber inter-core distance was measured locally and compared to the manufacturer data sheet. It was shown in [Le Goualher 04] that it was indeed possible to validate the complete procedure.

2.4.3 Real-time Image Reconstruction

Motivations

At this step, we have access to the distortion-free position of each fiber composing the image bundle as well as the restored intensity, $I_{restored}$, for each fiber. From an information theoretic point of view, we have thus recovered all the information we could get from the raw data. However, to visualize the data, it is necessary to reconstruct an image on a regular square grid from the irregularly sampled point set defined by the fibers. Such a reconstruction is also very useful if further image processing or image analysis is required. Most of the available image analysis algorithms have indeed been developed for regularly sampled data. And even if a scattered data version exists, it is often much more efficient in terms of computational time to take advantage of the regularity of the sampling grid.

The final process of the real-time image processing scheme used in Cellvizio will thus be the interpolation or approximation, into an image on a square grid, of the point set we get after the calibration and imaging model inversion steps. In the literature, this operation is usually referred to as scattered data interpolation or approximation, function reconstruction, resampling, etc.

For this step also, many factors have to be taken into account if we want to aim for a unerring interpolation or approximation algorithm. First of all, the reconstruction should be pleasing from a user point of view. Second, the reconstruction algorithm should be as fast as possible since we want the complete image processing pipeline to run in real time at 12 frames per second. If possible we would also like to spare as much CPU as possible for other tasks. We should also consider that visualization is only one of the requirements. We need, for example, the reconstruction for storage and further processing. It would indeed be inefficient to store only the point set and recompute the reconstruction each time we need a regular image. Moreover, from a disk usage point of view, we would like to avoid storing

the information twice. It would therefore be interesting to be able to recover, from the reconstructed image, a point set which is as close as possible to the initial one.

Common Reconstruction Methods

The problem of scattered data reconstruction (interpolation or approximation) in two or more independent variables has been addressed many times in various scientific fields. We refer the reader to [Amidror 02, Lodha 99], and references therein, for good reviews of the most classical methods.

Scattered data reconstruction can be seen as the problem of finding a continuous function $I(p)$ such that, given a set of points p_k and corresponding values I_k, we have $I(p_k)$ that is close to, or equal to, I_k. Without any further constraint, it is thus an ill-posed problem that admits many solutions. Let us briefly mention some widely used methods and point out their characteristics.

The simplest interpolation methods use a partitioning of the space into cells on which interpolation becomes easy. They can be related to finite element methods. This approach involves creating some type of optimal neighborhoods, using for example some triangulation or Voronoï diagrams, over which surface patches are defined. Piecewise constant functions can be used with a nearest-neighbor scheme to get the fastest reconstruction by assigning a common value to all neighboring points of a data point. C^0 continuity can be obtained by using piecewise linear reconstruction over a given triangulation on the data points. The Clough-Tocher method is certainly the most common C^1 method. It uses a rather ad-hoc triangle subdivision scheme to go beyond the simple C^0 continuity of linear interpolants. These methods are often rather efficient even if they need a computation of a space partition such as a Delaunay triangulation or Voronoï diagram. They are however often considered to be insufficient because they are sensitive to the distribution of the data points. Long and thin triangles can often not be avoided and produce low quality results. The most interesting variant of these methods is referred to as natural neighbor interpolation [Boissonnat 02]. The interpolated function is defined for each point p by adding p as a site to the Voronoï diagram of the original data points, and averaging the data points' values weighted by the fraction of the cell for p previously covered by each other cell. This procedure can result in a continuous function that is smooth everywhere.

Shepard's method is another classical option that is based on inverse distance weighting of data. The most basic version produces C^0 continuity but it can be modified to get a C^1 continuous interpolation. This technique suffers from several shortcomings such as the production of flatness artifacts at the data points. The most important issue of this method is however that it is a global method. All the points should be used to compute the value of the function at a given point.

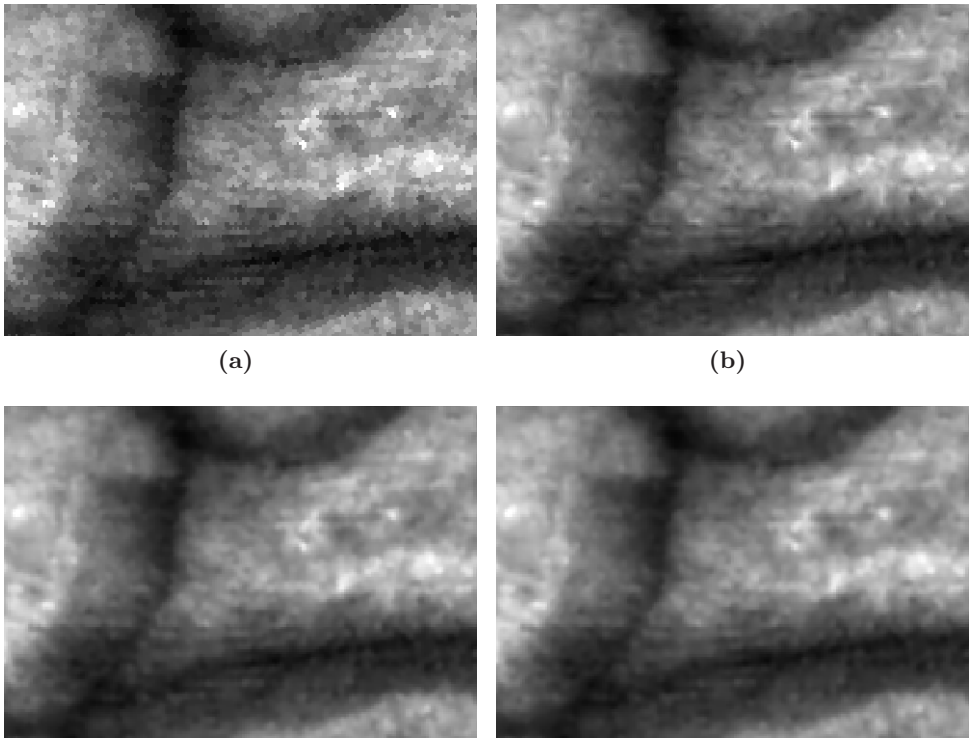
While these reconstruction methods often offer viable algorithmic alternatives to the problem of scattered data interpolation or approximation, from a theoretical point of view, it is not easy to justify them and thus to choose among them. A more theoretically well founded approach relies on the optimization of some smoothness

function of the set of functions that satisfy the interpolation constraints in a given Hilbert space. In [Duchon 77], Duchon showed that this formulation leads to a solution based on radial basis functions (RBF). RBF methods are often thought of to be the preferred interpolation techniques in terms of reconstruction quality. However, they rely on the resolution of a non-sparse linear system of equations with the dimension equal to the number of data points. This makes them intractable for large problems. Fast solutions exist [Cherrie 02] but they are still unfit for real-time applications. These methods can be extended to use compactly supported basis functions or provide approximations instead of interpolation by using moving least-squares (see [Fasshauer 06] and references therein). Nonetheless, in practice the computational requirements remains high. Other mathematically well-founded but computationally expensive methods have attacked the problem from the sampling theorem point of view [Aldroubi 01, Strohmer 97]. Their aim is to find a bandlimited function whose samples are as close as possible to the measured samples.

Note that most of the previously mentioned reconstruction algorithm aim at finding a continuous function. What we actually need in our case is a resampling of the data onto a regular rectangular grid. A very interesting approach used in [Arigovindan 05, Bernard 99, Lee 97, Vázquez 05] to attack this problem directly is to model the space of allowed function as a discrete B-Spline or wavelet basis. The goal is then to find the best coefficients in this basis. While these algorithms provide a very sound approach to our reconstruction problem, they are still not perfectly fit for real-time application.

Our Approach to Reconstruction

Our problem is however somewhat simpler than the generic resampling one since our data points are on a quasi-hexagonal grid. It is also possible to choose an adequate reconstruction grid. The approach we take is thus very simple. On average, the distance between the neighbor fibers is given by d_{ic}. We choose the reconstruction grid such that, if we map any given fiber to its closest reconstruction grid point (or pixel), the distance between the initial data-point and the resulting discretized data-points is less than some given fraction of d_{ic}. We also require that no two fibers are assigned the same pixel. These rather restrictive conditions can always be met in our case because of the quasi-hexagonal pattern and because the diameter of the fibers in the bundle is bounded by design. This results in choosing d_{ic} to span around 3 pixels. One could argue that our approach amounts to modifying the fibers positions. However, the perturbation is very local. Moreover, since the fibers positions were extracted from a segmentation algorithm, there is uncertainty in the fibers positions in the first place. From this discrete mapping, we can come up with very efficient algorithms. The nearest-neighbor reconstruction algorithm can use a fast discrete distance map scheme to compute the neighboring areas. This approach also allows to get interpolations in the discrete setting. Most of the available continuous interpolation methods will indeed not allow us to recover the



(a) **(b)**

(c) **(d)**

Figure 2.9: Visual comparison of the tested reconstruction schemes using a close up on a typical image. (a) Nearest neighbor interpolation. (b) Linear triangular interpolation. (c) Smoothed nearest neighbor approximation. (d) Gaussian Shepard approximation.

initial data point values if the continuous function is then sampled onto a rectangular reconstruction grid. By using this discretization, the continuous interpolation methods will however also provide a discrete interpolation. Note also that since our data point mesh is quasi-hexagonal, the Delaunay triangulation will not produce any problematic long and thin triangles. Moreover, the discretization can be used to provide more efficient schemes. In [Vercauteren 06] and Section 5, we provide an efficient Shepard's like method that relies on a Gaussian weighting influence of the data points instead of the usual inverse-distance. With this scheme, a mesh-free approximation can be obtained by using two Gaussian filtering operations and one image division. This provides a rather low computational complexity.

We first did some preliminary evaluation of the different methods to get a feel of their adequacy and computational requirements. Because of our strict real-time computation requirement, only four methods remained as potential candidates: the nearest neighbor reconstruction, the linear triangular interpolation using Delaunay triangulation, a smoothed version of the nearest neighbor method and our Gaussian Shepard-like algorithm. An equal emphasis was then put on their algorithmic implementation to get a fair comparison and include the right tool within Mauna

Table 2.1: Comparison of the tested reconstruction schemes. The reconstruction time was measured on a laptop Pentium 4. The reconstruction RNMSE is given by $\sqrt{\sum_k (I_k - I_k^{recons})^2 / \sum_k I_k^2}$ and measures the amount of approximation. The reconstruction visual quality was assessed by expert users.

	NN	Smoothed NN	Linear triangular	Gauss. Shepard
Recons. time (ms/s)	1.5	4.8	3.4	7.6
Visual quality	Bad	Good	Sufficient	Good
Recons. RNMSE	0%	3.4%	0%	4.2%

Kea Technologies proprietary softwares. All the implementations heavily rely on precomputations.

Figure 2.9 shows an example image reconstructed with these four algorithms. We evaluated these algorithms on a set of images to get the results shown in Table 2.1. As expected, the nearest neighbor method is the fastest but provides insufficient results in terms of image quality. The visual quality of the nearest neighbor scheme can be enhanced at the price of losing the discrete interpolation property and of adding some computational time by apply a Gaussian smoothing after the NN reconstruction. Our Gaussian Shepard-like method performs very similarly to the smoothed NN algorithm. Finally, the discretized linear triangular interpolation performs the best. Its computational time is the second best. It allows us to retrieve, from the reconstructed image, the initial data points values. It also provides a sufficient visual quality for real-time applications. Therefore, this scheme is now used for every single image seen by a Cellvizio user.

It should be noted that, if the reconstruction requirements changed, this ranking would need to be changed. For example for the image mosaicing problem we address in [Vercauteren 06] and Chapter 5, we only need to reconstruct one image. Furthermore, the data points do not form a quasi-hexagonal mesh anymore. This implies that, in such an application, the precomputation time has to be taken into account. Also, the Delaunay triangulation will produce unwanted long and thin triangles. This is why we developed the Gaussian Shepard-like method.

2.5 The Need for Image Registration

We have seen in this chapter that fibered confocal microscopy devices are becoming a standard tool to perform *in vivo* and *in situ* imaging both in research applications on living animals and in the clinical setting. These new imaging technologies allow for the acquisition and visualization of microscopic images with cellular resolution in any part of the living body, in real time, and without removal of biological tissue.

We have shown that even if the raw data that Cellvizio produces is difficult to interpret, adequate real-time image processing can provide readily interpretable smooth motion movies. The first contribution of this chapter has been to provide a detailed description of fibered confocal microscopy. The main contribution ensued from this insight: Several efficient image reconstruction schemes adapted to the

particular setting of FCM were proposed and fairly compared. This comparison led to a well-motivated choice and to a change of Mauna Kea Technologies proprietary softwares. The resulting video sequences can be used for visualization and quantification.

The classical way of using the data produced by biomedical imaging device is still mainly based on human interpretation. In the next decades this is however expected to change. In many medical centers in the United States, computer-aided diagnosis has for example already become a part of the routine clinical operation for detection of breast cancers. Most biologists and clinical practitioners are starting to face the need for a more quantitative and automated path for image interpretation. Typical biomedical imaging users will greatly benefit from advanced image analysis tools capable of extracting the pertinent information.

We do believe that image registration algorithms are among the most important building blocks that will enable the design of such automated image analysis tools for fibered confocal microscopy. Even if great results have already been achieved on quantification applications on single frames acquired with Cellvizio [Bourgeais 05, Vincent 06], there are many applications where either tissue or imaging device motions are major problems to perform more accurate quantification. The quest for noninvasiveness (organs should not be damaged) can lead to tissue or imaging device motions that result in motion artifacts and possible misquantifications. This noninvasiveness goal is also compounded by the need for a large field-of-view which implies a trade-off between the size of the imaged region and the size of the optical microprobe. Both the miniaturization of the flexible optical microprobe and the access difficulties are responsible for such limitations of the imaging device.

We will show in this thesis how image registration tools can help us move beyond these limitations. By allowing the stabilization of a given region of interest in a sequence acquired on a moving organ, image alignment schemes can support more accurate measurements and can help us track the evolution of a biological process that takes place during the acquisition. To track the evolution of a disease, it is also necessary to perform longitudinal studies and register the images acquired at different time steps. With the advent of biomedical imaging, many users are also using several imaging modalities at the same time to get different type of information simultaneously. To take full advantage of these multiple sources of information, it is often necessary to align and fuse the different images. We will also show in this thesis that, through video mosaicing, image registration can help us reach a wide field of view while keeping the imaging procedure non-invasive.

Chapter 3

Optimization Methods for Linear Image Registration

Table of Contents

3.1 Motivation: Fast and Robust Alignment of Pairs of Images

As we mentioned previously, image registration has become a fundamental tool in many biomedical image analysis problems. The process of image registration, which is also referred to as image fusion, superimposition or matching, aims at finding an optimal spatial transformation that will align the given images. Image registration should map each point in one image onto the corresponding point in the second image. Of course, the registration schemes that can be used depend on the type of images we want to align and on the type of spatial transformations we need. This chapter is focused on the registration of images acquired with the same imaging modality (mono-modal problem) under the assumption of a low-dimensional parametric space of spatial transformations. More specifically, we will mostly consider linear spatial transformations such as translations, rigid-body transformations and affine transformations. We also focus on registration methods that use the image

intensity directly rather than relying on a set of image features. We refer the reader to [Maintz 98] for a survey of image registration methods and a discussion on the different properties of intensity-based methods versus feature-based methods.

The main goal of this chapter is to provide a fast and robust approach to the intensity-based registration problem. If registration of pairs of images is to be a building block of more elaborate Cellvizio video sequences analysis pipelines, the efficiency of the schemes we use is important. We will indeed often need to register many pairs of frames to process a complete sequence. Furthermore, as the integration of information from multiple images finds more and more applications in the fields of biomedical research and clinical applications, the efficiency of the image registration procedures becomes a crucial point for the end-users of many other biomedical imaging devices. Consequently there is a growing interest from the scientific community to better understand and optimize the registration procedures [Farnebäck 06, Modersitzki 04].

In this chapter, we use the work we developed in [Vercauteren 07b] and present an efficient approach to image registration with a focus on mono-modal image registration. In this setting, registration is classically performed by optimizing a similarity criterion such as the mean squared error. Literature on image registration and optimization theory already provides a wealth of algorithms that can be used to solve this problem. However, they do not always use all the specificity of mono-modal image registration. Our main contribution in this chapter is to shed a new light on this problem by showing that the tools that have recently been developed in [Benhimane 04, Benhimane 06, Malis 04] in the field of vision-based robot control can be used for biomedical image registration and that they outperform the well-known optimizers. The efficient second-order minimization (ESM) technique of [Benhimane 04] takes advantage of the specificity of mono-modal image registration to boost its convergence rate. It is not tailored to a particular class of spatial transformations and can thus be used for a broad class of problems.

In Section 3.4 we concentrate on one particular application of this efficient registration scheme. The high resolution images provided by Cellvizio are mostly acquired on living organs, therefore specific image processing tools are required to cope with the natural movements of the tissues being imaged. We address the case where a Cellvizio user wants to focus on a small region of the living tissue that is difficult to stabilize mechanically. For instance *in vivo* and *in situ* acquisition on the liver, the bladder or even the heart can be unstable. Such organs receive a growing interest among biologists to assess pharmacokinetics parameters of molecules, to screen the changing morphology of the anatomy, or to measure bio-distribution parameters. Moreover, if the images of a sequence were stabilized, measurements of various image parameters would become possible or easier and could be carried out for many applications. We propose a region-of-interest tracker that provides one possible solution to the problem of image stabilization. This region-of-interest tracker and its application to blood velocity estimation in capillaries of moving organs form our second main contribution in this chapter. This work was first presented in [Perchant 07].

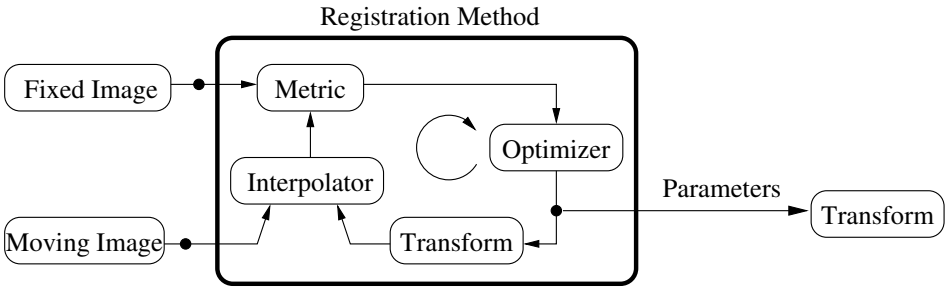


Figure 3.1: Schematic view of intensity-based image registration. The optimizer aims at finding an optimal spatial transformation with respect to the similarity criterion or metric.

3.2 Rigorous Mathematical Framework For Intensity-based Image Registration

3.2.1 Image Registration Model

Given a *fixed image* $F(.)$ and a *moving image* $M(.)$ in a D-dimensional space, intensity-based image registration is treated as an optimization problem that aims at finding the spatial mapping that will align the fixed and moving images. The transformation $s(.)$: $\mathbb{R}^D \to \mathbb{R}^D$, $p \mapsto s(p)$, models the spatial mapping of points from the fixed image space to the moving image space. The similarity criterion $\mathrm{Sim}\,(F, M \circ s)$ measures the quality of a given transformation. In this chapter we will only consider the mean squared error (MSE) similarity measure which forms the basis of the intensity-based image registration algorithms:

$$\mathrm{Sim}\,(F, M \circ s) = \frac{1}{2}\left\|F - M \circ s\right\|^2 = \frac{1}{2\left|\Omega_P\right|}\sum_{p\in\Omega_P}\left|F(p) - M(s(p))\right|^2, \qquad (3.1)$$

where Ω_P is the region of overlap between F and $M \circ s$. Figure 3.1 shows a schematic view of the registration process.

In order to register the fixed and moving images, we need to optimize (3.1) over a given space of spatial transformations. This can often be done by parameterizing the transformations. However, most of the spatial transformations we use do not form vector spaces but only Lie groups (e.g. rigid body, projective, diffeomorphisms...), meaning that we have a smooth manifold where one can invert or compose transformations and obtain a spatial transformation of the same type. We thus need to perform an optimization procedure on a Lie group such as in [Benhimane 04, Lee 05, Mahony 02].

3.2.2 Newton Methods for Lie Groups

Optimization problems on Lie groups can often be related to constrained optimization by embedding the Lie group in an Euclidean space. The classical way of dealing with the geometric structure of the Lie group is to use Lagrange multipliers or when

the constraints are simple to have an *ad hoc* procedure to preserve the constraints (e.g. renormalize a quaternion to have a unit quaternion). In this work we use an alternative strategy known as geometric optimization which uses local canonical coordinates [Mahony 02]. This strategy intrinsically takes care of the geometric structure of the group and allows the use of unconstrained optimization routines.

Let us first recall that a Lie group $\mathcal{G}$ is a smooth manifold together with a smooth composition map usually denoted as multiplication ($x \mapsto s \circ x$ for x and s in $\mathcal{G}$), and a smooth inverse map ($x \mapsto x^{-1}$ for x in $\mathcal{G}$), that satisfy the group axioms: closure, associativity, existence of a neutral element (denoted hereafter as Id) and existence of an inverse. We refer the reader to the standard textbooks for a detailed treatment of Lie groups, see e.g. [Helgason 01]. To any Lie group can be associated a Lie algebra $\mathfrak{g}$, whose underlying vector space is the tangent space of $\mathcal{G}$ at the neutral element Id. This Lie algebra captures the local structure of $\mathcal{G}$. The Lie group and the Lie algebra are related through the group exponential which is a diffeomorphism from a neighborhood of 0 in $\mathfrak{g}$ to a neighborhood of Id in $\mathcal{G}$. Let $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ be a basis of the Id-tangent space $T_{\mathrm{Id}}(\mathcal{G})$ corresponding to a basis of $\mathfrak{g}$. Canonical coordinates provide local coordinate charts so that for any $x \in \mathcal{G}$ in some neighborhood of s, there exists a vector $\boldsymbol{u} = \sum_i u_i \boldsymbol{e}_i \in T_{\mathrm{Id}}(\mathcal{G})$ such that $x = s \circ \exp(\boldsymbol{u}) = s \circ \exp(\sum_i u_i \boldsymbol{e}_i)$. These coordinates can be used to get the Taylor expansion of a smooth function φ on the Lie group $\mathcal{G}$:

$$\varphi\left(s \circ \exp(\boldsymbol{u})\right) = \varphi(s) + J_s^{\varphi}.\boldsymbol{u} + \frac{1}{2}\boldsymbol{u}^T.H_s^{\varphi}.\boldsymbol{u} + O(\|\boldsymbol{u}\|^3), \tag{3.2}$$

where $[J_s^{\varphi}]_i = \frac{\partial}{\partial u_i}\varphi(s \circ \exp(\boldsymbol{u}))\big|_{\boldsymbol{u}=0}$ and $[H_s^{\varphi}]_{ij} = \frac{\partial^2}{\partial u_i \partial u_j}\varphi(s \circ \exp(\boldsymbol{u}))\big|_{\boldsymbol{u}=0}$.

This approximation is used in [Mahony 02] to adapt the classical Newton-Raphson method by using an intrinsic update step:

$$s \leftarrow s \circ \exp(\boldsymbol{u}), \tag{3.3}$$

where $\boldsymbol{u}$ solves $H_s^{\varphi}.\boldsymbol{u} = -{J_s^{\varphi}}^T.\varphi(s)$.

Algorithm 1 (Newton-Raphson Method on a Lie Group)

- Choose a starting point s
- Iterate until convergence:
  - Given s, compute J_s^{φ} and H_s^{φ}
  - Compute the update $\boldsymbol{u}$ by solving the linear system $H_s^{\varphi}.\boldsymbol{u} = -{J_s^{\varphi}}^T.\varphi(s)$ using e.g. a Cholesky factorization of H_s^{φ}
  - Let $s \leftarrow s \circ \exp(\lambda \boldsymbol{u})$, with $\lambda = 1$ for the classical case but a line search can also be used ∎

As in the vector space case, this algorithm has a local quadratic convergence. Another important property of this algorithm is that, similar to the classical Newton-Raphson method, it is independent of the basis of $\mathfrak{g}$ that we use.

In many cases, using the Newton-Raphson method is not advocated or simply not possible. The Hessian matrix is indeed often difficult or impossible to compute, is not numerically well-behaved and convergence problems may arise when it is not positive-definite. To address these problems in the context of non-linear least squares optimization, most of the available efficient methods (e.g. Levenberg-Marquardt) are related to the Gauss-Newton method [Madsen 99].

Let $\phi(.) = \frac{1}{2}\left\|\boldsymbol{\varphi}(.)\right\|^2 = \frac{1}{2}\sum_p \varphi_p(.)^2$ be a sum of squared smooth functions. The Gauss-Newton method is based on a linear approximation of $\boldsymbol{\varphi}$ in a neighborhood of the current estimate. From (3.2), we have $\boldsymbol{\varphi}\left(s \circ \exp(\boldsymbol{u})\right) = \boldsymbol{\varphi}(s) + J_s^{\boldsymbol{\varphi}}.\boldsymbol{u} + O(\|\boldsymbol{u}\|^2)$. By keeping only the linear part, we obtain a quadratic approximation that we use to derive the Gauss-Newton method on a Lie group:

$$\phi(s \circ \exp(\boldsymbol{u})) = \frac{1}{2}\left\|\boldsymbol{\varphi}(s \circ \exp(\boldsymbol{u}))\right\|^2 \approx \frac{1}{2}\left\|\boldsymbol{\varphi}(s) + J_s^{\boldsymbol{\varphi}}.\boldsymbol{u}\right\|^2. \tag{3.4}$$

It is well known that if $J_s^{\boldsymbol{\varphi}}$ has full rank, this equation admits a unique minimizer which is the solution of the following equations known as normal equations:

$$\left(J_s^{\boldsymbol{\varphi}T}.J_s^{\boldsymbol{\varphi}}\right).\boldsymbol{u} = -J_s^{\boldsymbol{\varphi}T}.\boldsymbol{\varphi}(s).$$

By using this solution in the intrinsic update step, $s \leftarrow s \circ \exp(\boldsymbol{u})$, we get the Gauss-Newton method for Lie Groups.

Algorithm 2 (Gauss-Newton Method on a Lie Group)

- Choose a starting point s
- Iterate until convergence:
  - Given s, compute $J_s^{\boldsymbol{\varphi}}$
  - Compute the update $\boldsymbol{u}$ by solving the linear system $\left(J_s^{\boldsymbol{\varphi}T}.J_s^{\boldsymbol{\varphi}}\right).\boldsymbol{u} = -J_s^{\boldsymbol{\varphi}T}.\boldsymbol{\varphi}(s)$ using for example a Cholesky decomposition of $J_s^{\boldsymbol{\varphi}T}.J_s^{\boldsymbol{\varphi}}$ or a QR factorization of $J_s^{\boldsymbol{\varphi}}$
  - Let $s \leftarrow s \circ \exp(\lambda\boldsymbol{u})$, with $\lambda = 1$ for the classical case but a line search can also be used ■

In a vector space, the local convergence of the Gauss-Newton method (and the Levenberg-Marquardt) is in general not quadratic. In the Lie group setting, we also see that (3.4) is only a first-order approximation. We must therefore also expect only local linear convergence.

3.2.3 Gauss-Newton for Image Registration

For the registration problem (3.1), the Gauss-Newton algorithm can be used with the following function involved in the nonlinear least squares problem:

$$\varphi_p(s \circ \exp(\boldsymbol{u})) = F(p) - M \circ s \circ \exp(\boldsymbol{u})(p). \tag{3.5}$$

We now need to compute the Jacobian $J_s^{\varphi_p}$ of this function. In general, we will be able to compute the exponential numerically but we will not have a closed-form formula for it. An attractive way to compute the Jacobian will therefore rely on the chain rule and the fact that the differential map of the exponential at Id is the identity [Helgason 01].

In practice, we need a computational representation of the Lie group and the Lie algebra. By Whitney's theorem, we know that there exists an embedding Θ, $\mathcal{G} \to \mathbb{R}^N$, $s \mapsto \Theta(s)$ of the Lie group in an Euclidean space. This embedding also allows us to represent the Lie algebra. An example is the matrix representation of the common spatial transformations (e.g. rigid body, affine, projective) in homogeneous coordinates. In practice, this Euclidean representation is used to compute the spatial transformation (e.g. using matrix multiplication in homogeneous coordinates). Let us denote $w(\Theta(s), p)$ the representation, in the Euclidean embedding space $\mathbb{R}^N$, of the transformation of a point $p \in \mathbb{R}^D$ through the mapping $s \in \mathcal{G}$. Because of the embedding properties, $w(., p)$ is well defined in a neighborhood of $\mathcal{G}$ in $\mathbb{R}^N$ (e.g. the matrix multiplication in homogeneous coordinates need not be restricted to a specific kind of matrices). Using this representation, the chain rule and the fact that the differential map of the exponential at Id is the identity, the Jacobian of (3.5) can be decomposed as (cf. below):

$$J_s^{\varphi_p} = \frac{\partial}{\partial \boldsymbol{u}^T} \varphi_p(s \circ \exp(\boldsymbol{u}))\Big|_{\boldsymbol{u}=0} = -\nabla_p^T (M \circ s).J^{w_p}.\boldsymbol{e}_\Theta, \tag{3.6}$$

where $\nabla_p(M \circ s) = \frac{\partial M \circ s(q)}{\partial q^T}|_{q=p}$ is the gradient of the warped moving image ($D \times 1$ vector), $J^{w_p} = \frac{\partial w(X,p)}{\partial X^T}\big|_{X=\Theta(\mathrm{Id})}$ is the derivative of the mapping action expressed in the Euclidean embedding space ($D \times N$ matrix) and $\boldsymbol{e}_\Theta = [\Theta(\boldsymbol{e}_1), \ldots, \Theta(\boldsymbol{e}_n)]$ stacks the basis vectors of $\mathfrak{g}$ expressed in the Euclidean embedding space ($N \times n$ matrix). An example of such a decomposition is given in Section 3.3.2 for the rigid body case.

Derivation of (3.6) We apply the chain rule to $J_s^{\varphi_p} = -\frac{\partial M \circ s \circ e^{\boldsymbol{u}}(p)}{\partial \boldsymbol{u}^T}|_{\boldsymbol{u}=0}$, the Jacobian of (3.5), by using the following decomposition:

$$M \circ s \circ e^{\boldsymbol{u}}(p) = M \circ s\Big(w\big(\Theta(e^{\boldsymbol{u}}), p\big)\Big)$$

where $w(\Theta(e^{\boldsymbol{u}}), p) = e^{\boldsymbol{u}}(p)$ is the expression, in the Euclidean embedding space, of the transformation of the point p through the mapping $e^{\boldsymbol{u}}$ and where $\Theta(e^{\boldsymbol{u}})$ is the representation of $e^{\boldsymbol{u}}$ in the embedding space. We then get

$$\begin{aligned}[J_s^{\varphi_p}]_i &= -\frac{\partial M \circ s(q)}{\partial q^T}\Big|_{q=e^0(p)=p}.\frac{\partial w(X,p)}{\partial X^T}\Big|_{X=\Theta(e^0)=\Theta(\mathrm{Id})}.\frac{\partial \Theta(\exp(u_i \boldsymbol{e}_i))}{\partial u_i}\Big|_{u_i=0} \\ &= -\nabla_p^T (M \circ s).J^{w_p}.\Theta(\boldsymbol{e}_i),\end{aligned}$$

where we used the fact that the differential map of the exponential at Id is the identity and where by definition $J^{w_p} \triangleq \frac{\partial w(X,p)}{\partial X^T}\Big|_{x=\Theta(\mathrm{Id})}$.

3.3 Efficient Second-Order Minimization (ESM)

Image registration (especially mono-modal) is not just a generic optimization problem. Algorithms can take advantage of the specificity of the problem to develop more efficient schemes. In this section, we show that the efficient second-order minimization (ESM) procedure proposed in [Benhimane 04] to solve tracking problems in the field of vision-based robot control can be used to outperform classical image registration algorithms in the field of biomedical imaging.

The ESM scheme uses the fact that when the images are aligned with the optimal spatial transformation s^{opt}, the fixed image and the warped image as well as their gradient should be very close to each other: $\nabla_p(M \circ s^{\text{opt}}) \approx \nabla_p F$ where the equality holds up to a noise term. The main idea is that we can use this information to improve the search direction of the Newton methods.

The optimizers we presented above all work by building a polynomial approximation of the cost function. The Newton-Raphson uses the value of φ_p, its first and second derivatives around 0 to build a second-order polynomial approximation of φ_p. The Gauss-Newton and Levenberg-Marquardt methods discard the second derivative of φ_p and can thus only build a first-order polynomial approximation of it. Note that since the cost function uses only the squared norm of φ_p, Gauss-Newton and Levenberg-Marquardt still have some second-order information in their approximation of the cost function. These methods are however in general only first-order minimization routines. In constrast, the ESM uses the value of φ_p, its first derivative around 0 as well as its first derivative around s^{opt} to build a second-order polynomial without the need of second derivative information.

To get a feel of it, let $\varphi(x)$ be a simple real-valued second-order polynomial. We know that in this polynomial case, $\varphi(x) = \varphi(0) + \varphi'(0)x + \varphi''(0)x^2$ is a true equality and not only an approximation. It is also trivial to see that the following equality holds:

$$\varphi(x) = \varphi(0) + \frac{1}{2}\big(\varphi'(x) + \varphi'(0)\big)x.$$

In [Hummel 49], such generalized Taylor expansions are provided in the general vector space case. For any smooth function $\varphi(x)$ (and not only polynomials), there exists a point $y \in [0; x]$ such that the following equality holds:

$$\varphi(x) = \varphi(0) + \frac{1}{2}\big(\varphi'(x) + \varphi'(0)\big)x + \frac{1}{12}\varphi'''(y)x^3.$$

By using such a generalized Taylor expansion on Lie groups, the ESM provides a second-order minimization method that does not need the computation of the Hessian matrix.

3.3.1 A Second-Order Linearization

With the ESM, the information about the Hessian that is discarded with the Gauss-Newton iteration is recovered with a Taylor expansion of a Jacobian calculated at the optimal transformation. In a generic optimization problem, such

an information is not available. However, for image registration, we should have $\nabla_p(M \circ s^{\text{opt}}) \approx \nabla_p F$ up to a noise term. To use this very special property, let us define a generalization of the Jacobian used in Section 3.2.2:

$$J_s^{\boldsymbol{\varphi}}(\boldsymbol{u}) = \frac{\partial}{\partial \boldsymbol{v}^T} \boldsymbol{\varphi}(s \circ \exp(\boldsymbol{v}))\Big|_{\boldsymbol{v}=\boldsymbol{u}}. \tag{3.7}$$

Note that at $\boldsymbol{u} = 0$ we get the initial Jacobian: $J_s^{\boldsymbol{\varphi}}(0) = J_s^{\boldsymbol{\varphi}}$. By using a first-order expansion of (3.7) around 0 we have:

$$J_s^{\boldsymbol{\varphi}}(\boldsymbol{u}) = J_s^{\boldsymbol{\varphi}}(0) + \boldsymbol{u}^T.H_s^{\boldsymbol{\varphi}} + O(\|\boldsymbol{u}\|^2),$$

that can be rewritten as $\boldsymbol{u}^T.H_s^{\boldsymbol{\varphi}} = J_s^{\boldsymbol{\varphi}}(\boldsymbol{u}) - J_s^{\boldsymbol{\varphi}} + O(\|\boldsymbol{u}\|^2)$. By incorporating this expression into (3.2), this provides us with a true second-order Hessian-free approximation:

$$\begin{aligned} \boldsymbol{\varphi}(s \circ \exp(\boldsymbol{u})) &= \boldsymbol{\varphi}(s) + J_s^{\boldsymbol{\varphi}}.\boldsymbol{u} + \frac{1}{2}\left(J_s^{\boldsymbol{\varphi}}(\boldsymbol{u}) - J_s^{\boldsymbol{\varphi}}\right).\boldsymbol{u} + O(\|\boldsymbol{u}\|^3) \\ &= \boldsymbol{\varphi}(s) + \frac{1}{2}\left(J_s^{\boldsymbol{\varphi}}(\boldsymbol{u}) + J_s^{\boldsymbol{\varphi}}\right).\boldsymbol{u} + O(\|\boldsymbol{u}\|^3) \end{aligned} \tag{3.8}$$

The non-linear least squares problem of Section 3.2.2 can thus be revisited to get a second-order approximation of (3.4):

$$\phi(s \circ \exp(\boldsymbol{u})) = \frac{1}{2}\|\boldsymbol{\varphi}(s) + \frac{1}{2}\left(J_s^{\boldsymbol{\varphi}}(\boldsymbol{u}) + J_s^{\boldsymbol{\varphi}}\right).\boldsymbol{u}\|^2 + O(\|\boldsymbol{u}\|^3) \tag{3.9}$$

The computation of $J_s^{\boldsymbol{\varphi}}(\boldsymbol{u})$ is a difficult problem in the general setting. Even if we get a closed-form expression of it, a minimization problem that involves this term might not be easy to solve in practice. To be able to use (3.9), we need to use the special properties of our problem.

From the current transformation s, the optimal step $\boldsymbol{u}_s^{\text{opt}}$ that an optimizer can make would bring us to the optimum: $s^{\text{opt}} = s \circ \exp(\boldsymbol{u}_s^{\text{opt}})$. From a computational point of view, the main property behind the ESM procedure is that, for this optimal step, the product $J_s^{\boldsymbol{\varphi}}(\boldsymbol{u}_s^{\text{opt}}).\boldsymbol{u}_s^{\text{opt}}$ is linear in $\boldsymbol{u}_s^{\text{opt}}$ [Benhimane 06]. This property leads to a simple minimization of (3.9).

In order to compute the product $J_s^{\boldsymbol{\varphi}}(\boldsymbol{u}_s^{\text{opt}}).\boldsymbol{u}_s^{\text{opt}}$, we need to realize that it is related to the moving image warped with the optimal transformation. The idea is thus to replace the gradient of the optimally warped image $M \circ s^{\text{opt}} = M \circ s \circ \exp(\boldsymbol{u}_s^{\text{opt}})$ by its *equivalent*, the gradient of the fixed image F. We then get a simple linear approximation: $J_s^{\boldsymbol{\varphi}}(\boldsymbol{u}_s^{\text{opt}}).\boldsymbol{u}_s^{\text{opt}} \approx \nabla_p^T F.J^{w_p}.\boldsymbol{e}_\Theta.\boldsymbol{u}_s^{\text{opt}}$ as shown in the appendix page 48. This approximation can be used with (3.8) to get:

$$\begin{aligned} \boldsymbol{\varphi}\left(s \circ \exp(\boldsymbol{u}_s^{\text{opt}})\right) &= \boldsymbol{\varphi}(s) + J_s^{\text{ESM}}.\boldsymbol{u}_s^{\text{opt}} + O(\|\boldsymbol{u}_s^{\text{opt}}\|^3) \\ J_s^{\text{ESM}_p} &\triangleq -\frac{1}{2}\left(\nabla_p^T F + \nabla_p^T(M \circ s)\right).J^{w_p}.\boldsymbol{e}_\Theta \end{aligned} \tag{3.10}$$

where we omit the image noise and where J^{w_p} and $\boldsymbol{e}_\Theta$ are defined in Section 3.2.3.

An efficient image registration algorithm is thus obtained by choosing an initial spatial transformation and then performing the following iteration steps until convergence.

Algorithm 3 (ESM and Gauss-Newton for Registration)

- Given the current transformation s, compute the Jacobians to be used in the normal equations. Let $J^p = -\frac{1}{2}\big(\nabla_p^T F + \nabla_p^T (M \circ s)\big).J^{w_p}.\boldsymbol{e}_\Theta$ for ESM or $J^p = -\nabla_p^T (M \circ s).J^{w_p}.\boldsymbol{e}_\Theta$ for Gauss-Newton

- Compute the update $\boldsymbol{u}$ by solving the linear system $(J^T.J).\boldsymbol{u} = -J^T.\boldsymbol{\varphi}(s)$ using e.g. a Cholesky decomposition of $J^T.J$ or a QR factorization of J

- Let $s \leftarrow s \circ \exp(\lambda \boldsymbol{u})$, with $\lambda = 1$ or found by a line search procedure ∎

Note that both the Gauss-Newton and the ESM have the same computational complexity since $\nabla_p F$ needs only be computed once during initialization.

3.3.2 Example: 2D Rigid Body Transformations

Let us now focus on the optimization of (3.1) for the Lie group $SE(2)$ of 2D rigid body transformations. To use the optimization method presented in Algorithm 3, we need to know what the corresponding Lie algebra $\mathfrak{se}(2)$ is. We also need to be able to compute the exponential map and the necessary Jacobian.

Most of the Lie groups we consider can be represented as matrix groups, i.e., subgroups of invertible matrices. The composition and inversion operations are simply matrix multiplication and matrix inversion. In this setting, the exponential map is given explicitly by the standard matrix power series: $\exp(A) = \sum_{k=0}^{\infty} \frac{A^k}{k!}$. In general, the matrix exponential does not have a closed-form formula but efficient methods exists to compute it [Higham 05].

A 2D rigid body transformation r is composed of a rotation of angle α followed by a translation $\tau = (\tau_x, \tau_y)$. This Lie group $SE(2)$ can be represented using homogeneous coordinates by a 3×3 matrix group of the form

$$\Theta(r) = \left[\begin{smallmatrix} R_\alpha & \tau \\ 0 & 1 \end{smallmatrix}\right],$$

where $R_\alpha = \left[\begin{smallmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{smallmatrix}\right]$ is a rotation matrix. Thanks to this matrix representation (which is the Euclidean embedding space used in Section 3.2.3), we see that the Lie Algebra can be represented by the following vector space of matrices (see appendix page 49):

$$\mathfrak{se}(2) = \left\{ \left[\begin{smallmatrix} dR_\alpha & d\tau \\ 0 & 0 \end{smallmatrix}\right] \;\middle|\; dR_\alpha \text{ skew-symmetric} \right\} \tag{3.11}$$

where $dR_\alpha = \left[\begin{smallmatrix} 0 & -\alpha \\ \alpha & 0 \end{smallmatrix}\right]$ is any skew-symmetric matrix (α need not be restricted to $[0, 2\pi]$) and $d\tau$ is any vector. From this expression, we see that a convenient basis of the Lie algebra $\mathfrak{se}(2)$ is given (in matrix form) by

$$\Theta(\boldsymbol{e}_1) = \left[\begin{smallmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{smallmatrix}\right], \quad \Theta(\boldsymbol{e}_2) = \left[\begin{smallmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{smallmatrix}\right], \quad \Theta(\boldsymbol{e}_3) = \left[\begin{smallmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{smallmatrix}\right]. \tag{3.12}$$

We work with a Lie algebra that is a vector space of matrices. It is however sometimes more comfortable to work with columns vectors. In this case we have:

$$\begin{aligned}\Theta(\boldsymbol{e}_1) &= [0,1,0,-1,0,0,0,0,0]^T \\ \Theta(\boldsymbol{e}_2) &= [0,0,0,0,0,0,1,0,0]^T \\ \Theta(\boldsymbol{e}_3) &= [0,0,0,0,0,0,0,1,0]^T.\end{aligned} \tag{3.13}$$

To use Algorithm 3, we still need to compute is J^{w_p}. The spatial transformation $r(p)$ of a point p through a 2D rigid body transformation r is a simple matrix multiplication where, contrary to projective transformations, no rescaling is necessary. This leads to

$$J^{w_p} = \begin{bmatrix} p_x & 0 & 0 & p_y & 0 & 0 & 1 & 0 & 0 \\ 0 & p_x & 0 & 0 & p_y & 0 & 0 & 1 & 0 \end{bmatrix}.$$

After some basic simplifications shown in the appendix page 49, we obtain the expression of interest:

$$J^{w_p}.\boldsymbol{e}_\Theta = \begin{bmatrix} -p_y & 1 & 0 \\ p_x & 0 & 1 \end{bmatrix}. \tag{3.14}$$

Finally, in order to use the intrinsic update step, $s \leftarrow s \circ \exp(\lambda \boldsymbol{u})$, we need a computationnal scheme to get the exponential. This can be done using a generic matrix exponential tool such as the one available in MATLAB®, `expm`. However, in the special case of rigid body transformations, we have a closed-form expression of the matrix exponential [Arsigny 06b]:

$$\exp\left(\begin{bmatrix} dR_\alpha & d\tau \\ 0 & 0 \end{bmatrix}\right) = \begin{bmatrix} R_\alpha & A_\alpha.d\tau \\ 0 & 1 \end{bmatrix} \quad \text{where} \quad A_\alpha = \alpha^{-1} \begin{bmatrix} \sin(\alpha) & \cos(\alpha)-1 \\ 1-\cos(\alpha) & \sin(\alpha) \end{bmatrix}.$$

Registration results: In the context of tracking for vision-based robot control, a detailed comparison of the optimization schemes showed that, for the space of homographies, the ESM outperformed classical solutions [Benhimane 07]. In this section, we compare the performance of the ESM optimizer with respect to the Gauss-Newton optimizer on a real-life biomedical image registration problem. A $2D + t$

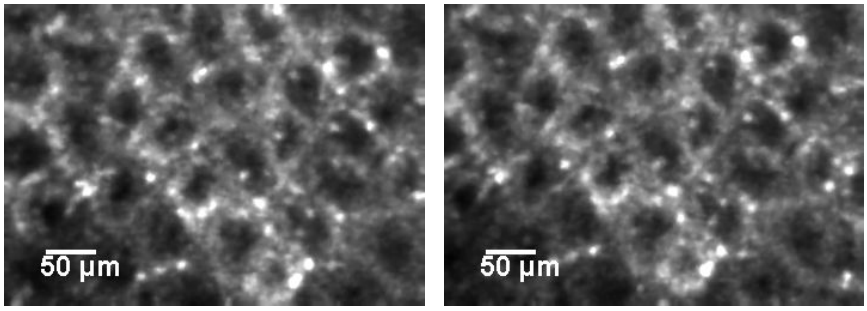
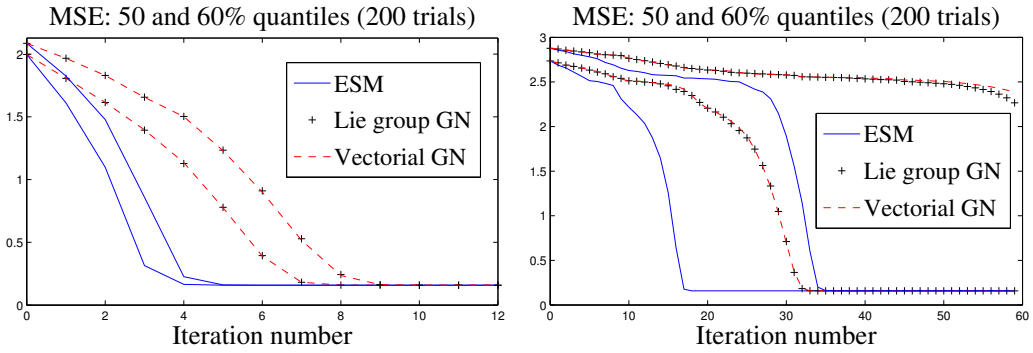


Figure 3.2: Two consecutive images of a dynamic sequence of live mouse colon acquired with a fibered confocal microscope. Images are courtesy of D. Vignjevic, S. Robine, D. Louvard, Institut Curie, Paris, France.



(a) $\sigma_\alpha = 0.15$ rad and $\sigma_\tau = 5$ pixels **(b)** $\sigma_\alpha = 0.3$ rad and $\sigma_\tau = 10$ pixels

Figure 3.3: Simple convergence experiments using the images of Fig. 3.2. We randomly generate initial rigid body transformations and compare the different optimizers. The random generator is Gaussian centered around the optimal transformation (validated by an expert), uses σ_α for the rotation part and σ_τ for the translation parts. Note that in the far initialization case on the right, the ESM is faster to converge with 50% of the trials converging in less than 19 iterations vs. 34 for the Gauss-Newton. It is also more robust as 60% converge in less than 36 iterations with ESM but we never reach 60% of convergence with the Gauss-Newton.

dynamic sequence is acquired with a fibered confocal microscope (FCM). We performed a rigid body registration between the consecutive frames [Vercauteren 06]. To get a statistically meaningful example, we chose two representative frames and compared the optimizers with random starting points. Since the emphasis is on the comparison of the various schemes and not on the final performance, no multi-resolution scheme was used.

Our results in Fig. 3.3 show that the analysis of [Benhimane 07] can be extended to the problem of biomedical image registration. We indeed see that for rigid body registration the ESM has a faster convergence rate and is more robust than the Gauss-Newton optimizer. In this special case, we can also see that the main advantage we get from the ESM comes from the Hessian-free second-order approximation rather than the intrinsic update step. We indeed see that the Lie group and vectorial Gauss-Newton schemes perform similarly. The main reason is that the Lie group of 2D rigid body transformations is very close to a Euclidean space. We work with a null curvature. Furthermore, with a simple parameterization using an angle and a translation vector, it is impossible to be mapped out from the Lie group. This 2D rigid body transformations setting needs thus be seen as an illustrative example of the power of the ESM. In more complex spaces, such as those required for projective geometry and those we use in the next chapter, improvement comes both from the intrinsic update step and from the Hessian-free second-order approximation.

3.4 Region Tracking Algorithms Applied to Cell Traffic Analysis

In vivo and *in situ* confocal images are often distorted by motion artifacts and soft tissue deformations. To measure small amplitude phenomena on this type of images, we have to compensate for such artifacts. One way of doing it is by using image registration schemes such as the one we showed previously in this chapter.

We present in this section a region-of-interest (ROI) tracking algorithm, that we developed in [Perchant 07], which is specialized for confocal imaging using a scanning device. One typical application of this tool is provided: the blood velocity estimation inside a capillary on a moving organ.

Cell trafficking in micro vessels is an important research field in cellular biology and molecular imaging. Numerous algorithms have been developed to assess cell motion, or blood flow, see e.g. [Sato 97, Savoire 04]. To the best of our knowledge, there had been no previous attempt to perform blood flow velocity measurement on a sequence acquired *in vivo* with global tissue motion. This application illustrates the usefulness of the region-of-interest (ROI) tracking we propose. First results show that the method permits accurate estimation of blood cell velocity even in presence of motion artifacts.

3.4.1 Region-of-Interest Tracker

Using Cellvizio, the hand-held flexible optical microprobe can freely glide along the soft tissue while keeping contact with it. The spatial transformation between two consecutive frames will thus be composed of a translation, a possible rotation, a possible little scaling if the tissues are compressed and even some residual non-rigid tissue deformation. Furthermore, an interesting point of scanning imaging devices is that the output image is not a representation of a given instant, but a juxtaposition of points acquired at different times. Instead of motion blur, we get geometric distortions, e.g. a circle is distorted into an ellipse. If, within the field of view and during the acquisition of an image, an object moves with a constant translational velocity, the imaged object gets distorted through a skew transformation [Savoire 04]. This feature has been successfully used for red blood cell velocimetry on single images [Savoire 04] and mosaicing of *in vivo* video confocal images [Vercauteren 06].

As a first approximation, we can ignore the non-rigid tissue deformation and represent the spatial transformation between two consecutive frames of a laser scanning confocal sequence by an affine transformation that takes into account the motion of the optical microprobe, the motion artifacts and the possible small scaling of the tissue. We know however that the most important component of the inter-frame spatial transformation we seek is given by the translation component.

The region-of-interest tracking scheme we designed aims at recovering the position of a template image in each frame of the acquired Cellvizio video sequence. The template is initialized by using a user-defined region of interest. When the

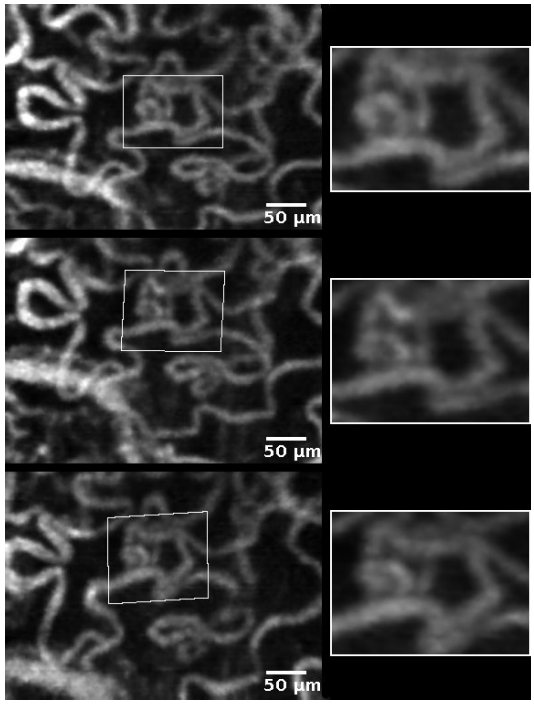


Figure 3.4: Tracking a selected ROI on tumor vasculature acquired *in vivo*. The upper frame is the reference frame, the other rows represent the frames 11 and 15 at from a 200 frame sequence. The tracked ROIs are shown in the left column with the corresponding warped region in the right column.

tissue does not change too fast, the same template can be kept over the complete video sequence. However, if the molecular dynamics evolve to quickly, it is necessary to update the image template. Unfortunately, this procedure can introduce a drift during tracking.

To ensure faster convergence of the registration scheme, and avoid as much as possible local optima, our algorithm uses a hierarchical framework that recovers the translation component first and then feeds an iterative multi-scale affine registration scheme with it. A globally optimal translation guess is computed by using a fast normalized correlation [Lewis 95]. Similarly to the rigid-body case, we have adapted the ESM registration scheme to work with the Lie group of affine transformation. A multi-scale pyramidal scheme avoids some local minima and provides some robustness. The computational requirements of the affine registration have also been lowered by using the fibered nature of our input images. We know from Chapter 2 that the only informative pixels in the reconstructed images are given by the ones that correspond to a fiber center. It is thus not necessary to use the other pixels for the computation of the similarity metric.

The complete steps of our ROI tracker can be summarized into the following algorithm:

Algorithm 4 (ROI Tracking Scheme)

- Start with an image sequence and a user-defined ROI in the first frame
- Let the user-defined ROI be the image template
- For each frame perform the following step
  - Predict the current template transformation from the previous one using for example a naive prediction or an autoregressive model
  - Resample the current image with respect to the predicted spatial transformation
  - Find a globally optimal translation using the fast normalized cross-correlation
  - Find an optimal affine transformation by registering the current image with the template with an iterative registration scheme initialized with the composition of the predicted transformation and the optimal translation
  - If necessary, update the template using the current registered ROI ■

3.4.2 Application to Cell Trafficking

The analysis of the behavior of blood cells and vessels is a very important topic of physiology research. As such, *in vivo* measurements made for example by intravital fluorescence microscopy have proved since the early seventies to be crucial to the understanding of the physiology and pathophysiology of micro-circulation. More recently, fibered confocal microscopy has been shown to allow for the observation and measurement of several characteristics of microcirculation with the clear benefit of reducing the invasiveness to its bare minimum [Laemmel 04].

However, as for any quantitative measurement from images, the automated analysis of the data poses a number of new challenges that need to be addressed with specific solutions. In this section, we show how the ROI tracking scheme we presented above was used in [Perchant 07] to assess cell trafficking in a capillary using Cellvizio.

The aim is to measure a slow blood velocity in a capillary. Classical methods for velocity measurements of blood cells in micro-vessels are often based on the processing of 2D temporal image sequences obtained in the field of intravital microscopy. Line shift diagram, spatio-temporal analysis or blood cell tracking are used in such setting to process the sequences generated by CCD-based video microscopes [Sato 97]. For this range of blood velocity and this size of capillary these conventional methods could not directly be used for Cellvizio because of the unstable nature of many imaged tissues such as tumoral grafts. This is why we chose to first track a given region of interest and then use a cross-correlation scheme on the stabilized sequence.

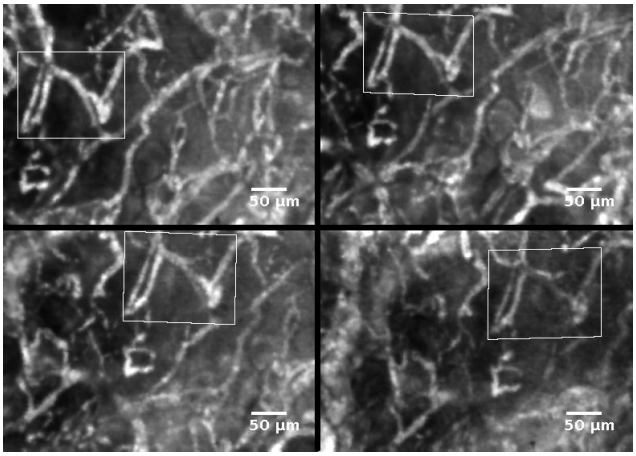


Figure 3.5: ROI tracking using affine transformations: 4 frames (index 1, 51, 101, 151) from the same sequence are displayed with the registered ROI. The complete sequence includes 237 frames.

The user can select manually a rectangular ROI on the image. Figure 3.5 shows the tracking of this region on a sequence acquired with a hand-held probe on a tumoral skin xenograft. Vessels were stained using dextran fluorescein from Invitrogen. Using the tracking results, we resampled each frame to compose a stabilized sequence. Three frames of this sequence of 5 seconds (acquired at 12 Hz) appear in Fig. 3.6.

On the temporal mean frame of the stabilized sequence, we have segmented the vessels using a 2D adaptation of the multi-scale tubular vessel detection algorithm of [Krissian 00]. We used this adaptation on the same type of images in [Lin 06] to perform morphometric analysis of the vascular network. The upper left frame of Fig. 3.6 shows the result of the detection: the medial axis and the vessel borders. The mean vessel diameter is 10.7 µm, which is roughly the size of a red blood cell.

The medial axis of the vessel in the ROI displayed in Fig. 3.6 was used to extract the vessel intensity in the center line. The normalized cross correlation of these two lines allows for the estimation of the velocity of the blood in the capillary. Two temporally contiguous lines are displayed in Fig. 3.7. Figure 3.8 shows the estimation of the velocity of the blood in the capillary.

The range of velocities that we can address depends on the scanning period, and the amplitude. Typical values on the device we used, Cellvizio, are 12 Hz frame rate, with a field-of-view of 0.5×0.6 mm^2, and a 3.5 µm spatial sampling distance, i.e., lateral resolution. The velocity precision is given by the minimum translation observable between two frames: $\delta v = 0.012$ mm/s. The velocity interval computed using a maximum detectable translation of half the horizontal field-of-view is $[0, 3.6]$ mm/s. The same device can perform fast scanning at a frequency up to 200 Hz with the same resolution by reducing the vertical field-of-view. The velocity interval in this case is maximal when the vessel is horizontal: $[0, 60]$ mm/s with a precision of 0.2 mm/s.

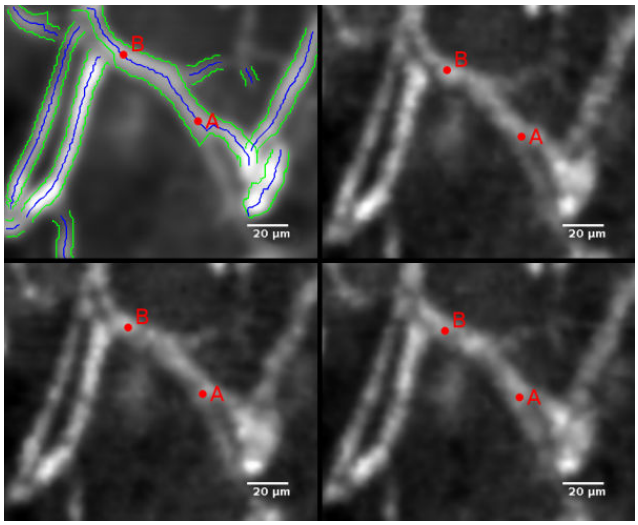


Figure 3.6: Upper-left: vessel detection on the temporal mean frame after stabilization. Other images: three contiguous frames of the stabilized sequence (12 Hz). Blood velocity was acquired on the medial axis segment [AB].

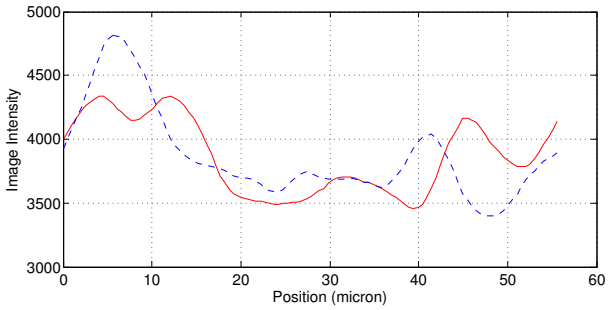


Figure 3.7: The medial axis intensity of the same vessel on two contiguous frame. The correlation of the two signals is visible.

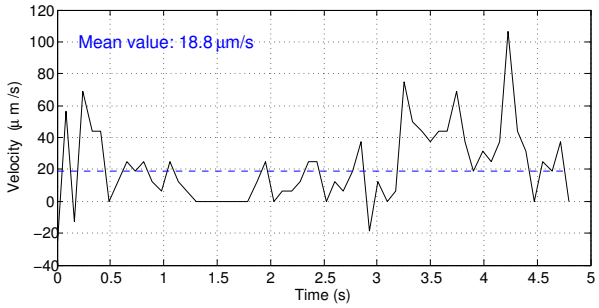


Figure 3.8: Velocities computed using the correlation method on the registered ROI. The dashed line indicates the mean value of the velocities.

An additional interesting feature of this tracker is that it also enables the reconstruction of images on the region of interest with an enhanced resolution. This is made possible, when the kinetics of the signal is slow enough, thanks to the noise reduction provided by the processing of several registered noisy images of the same region and to a small remaining aliasing of the input images.

3.5 Conclusions

We showed in this chapter that some tools that have recently been developed for tracking problems in the field of vision-based robot control can outperform classical biomedical image registration algorithms by exploiting the special nature of the image registration problem. We have focused on mono-modal registration but the ESM scheme can also be extended to address more complex intensity relationships. Robust estimation techniques can be used to account for outliers in the cost function and we plan to investigate on iterative intensity matching for the optimization of other simple similarity metrics such as the correlation coefficient and the correlation ratio.

A biologically relevant application of this efficient registration scheme has also been proposed. We presented a framework to accurately measure blood velocity inside a small capillary on a moving region of interest in a field-of-view. The region of interest is stabilized using a specialized tracking algorithm that use an ESM registration scheme. On the stabilized region, the capillaries were detected using tubular model-based segmentation. The medial axis signal level was extracted from the image. The spatio-temporal correlation was finally used to estimate the blood velocity in the capillary. Now that the feasibility of the framework has been proved, the next step would be the validation on both numerical and real data which is left for future work.

Appendix

Derivation of (3.10)

We detail here how $J_s^{\varphi_p}(\boldsymbol{u}_s^{\text{opt}}) = -\frac{\partial M \circ s \circ e^{\boldsymbol{v}}(p)}{\partial \boldsymbol{v}^T}\big|_{\boldsymbol{u}=\boldsymbol{u}_s^{\text{opt}}}$ can be decomposed into a product of three terms. We start by incorporating the optimal update step $\boldsymbol{u}_s^{\text{opt}}$ into $M \circ s \circ e^{\boldsymbol{v}}(p)$:

$$\begin{aligned} M \circ s \circ e^{\boldsymbol{v}}(p) &= M \circ s \circ e^{\boldsymbol{u}_s^{\text{opt}}} \circ e^{-\boldsymbol{u}_s^{\text{opt}}} \circ e^{\boldsymbol{v}}(p) \\ &= M \circ s^{\text{opt}}\Big(w\big(\Theta(e^{-\boldsymbol{u}_s^{\text{opt}}} \circ e^{\boldsymbol{v}}), p\big)\Big), \end{aligned}$$

where we used the fact that, by definition of $\boldsymbol{u}_s^{\text{opt}}$, we have $s \circ e^{\boldsymbol{u}_s^{\text{opt}}} = s^{\text{opt}}$; where as previously $w(\Theta(e^{-\boldsymbol{u}_s^{\text{opt}}} \circ e^{\boldsymbol{v}}), p) = e^{-\boldsymbol{u}_s^{\text{opt}}} \circ e^{\boldsymbol{v}}(p)$ is the expression, in the Euclidean embedding space, of the transformation of the point p through the mapping $e^{-\boldsymbol{u}_s^{\text{opt}}} \circ e^{\boldsymbol{v}}$; and where $\Theta(e^{-\boldsymbol{u}_s^{\text{opt}}} \circ e^{\boldsymbol{v}})$ is the representation of $e^{-\boldsymbol{u}_s^{\text{opt}}} \circ e^{\boldsymbol{v}}$ in the embedding space. By using the chain rule we get:

$$\begin{aligned} J_s^{\varphi_p}(\boldsymbol{u}_s^{\text{opt}}) = \frac{\partial M \circ s^{\text{opt}}(q)}{\partial q^T}\Big|_{q=e^{-\boldsymbol{u}_s^{\text{opt}}} \circ e^{\boldsymbol{u}_s^{\text{opt}}}(p)} \\ \cdot \frac{\partial w(X,p)}{\partial X^T}\Big|_{X=\Theta(e^{-\boldsymbol{u}_s^{\text{opt}}} \circ e^{\boldsymbol{u}_s^{\text{opt}}})} \\ \cdot \frac{\partial \Theta(e^{-\boldsymbol{u}_s^{\text{opt}}} \circ e^{\boldsymbol{u}})}{\partial \boldsymbol{u}^T}\Big|_{\boldsymbol{u}=\boldsymbol{u}_s^{\text{opt}}} \end{aligned}$$

The first term is simply given by the gradient of the optimally warped moving image which is approximately the gradient of the fixed image:

$$\frac{\partial M \circ s^{\text{opt}}(q)}{\partial q^T}\Big|_{q=e^{-\boldsymbol{u}_s^{\text{opt}}} \circ e^{\boldsymbol{u}_s^{\text{opt}}}(p)} = \nabla_p^T (M \circ s^{\text{opt}}) = \nabla_p^T F + \varepsilon,$$

where ε is a noise term. The second term is the same as the one appearing in (3.6):

$$\frac{\partial w(X,p)}{\partial X^T}\Big|_{X=\Theta(e^{-\boldsymbol{u}_s^{\text{opt}}} \circ e^{\boldsymbol{u}_s^{\text{opt}}})} = \frac{\partial w(X,p)}{\partial X^T}\Big|_{X=\Theta(\text{Id})} \triangleq J^{w_p}.$$

The last term is in general very difficult to compute. However, we only need to compute its product with $\boldsymbol{u}_s^{\text{opt}}$ which appears to be a directional derivative. We can thus also write it as a rate of change and see it as another directional derivative:

$$\begin{aligned} \frac{\partial \Theta(e^{-\boldsymbol{u}_s^{\text{opt}}} \circ e^{\boldsymbol{u}})}{\partial \boldsymbol{u}^T}\Big|_{\boldsymbol{u}=\boldsymbol{u}_s^{\text{opt}}} . \boldsymbol{u}_s^{\text{opt}} &= \frac{\partial \Theta(e^{-\boldsymbol{u}_s^{\text{opt}}} \circ e^{\boldsymbol{u}_s^{\text{opt}} + t\boldsymbol{u}_s^{\text{opt}}})}{\partial t}\Big|_{t=0} \\ &= \frac{\partial \Theta(e^{-\boldsymbol{u}_s^{\text{opt}} + \boldsymbol{u}_s^{\text{opt}} + t\boldsymbol{u}_s^{\text{opt}}})}{\partial t}\Big|_{t=0} = \frac{\partial \Theta(e^{t\boldsymbol{u}_s^{\text{opt}}})}{\partial t}\Big|_{t=0} \\ &= \frac{\partial \Theta(e^{\boldsymbol{u}})}{\partial \boldsymbol{u}^T}\Big|_{\boldsymbol{u}=0} . \boldsymbol{u}_s^{\text{opt}} = \boldsymbol{e}_\Theta . \boldsymbol{u}_s^{\text{opt}}. \end{aligned}$$

In this equation, we used the fact that the exponential defines one-parameter subgroups. This implies that for any scalars α and β we have $e^{\alpha \boldsymbol{u}} \circ e^{\beta \boldsymbol{u}} = e^{(\alpha+\beta)\boldsymbol{u}}$. As previously, $\boldsymbol{e}_\Theta$ stacks the basis vectors of the Lie algebra $\mathfrak{g}$ expressed in the Euclidean embedding space.

Lie Algebra for $SE(N)$ (3.11)

This appendix is designed to provide an intuition of how one can easily find the Lie algebra of a given Lie group. It is definitely not meant to be a formal proof.

If $\mathcal{G}$ is a closed subgroup of $GLn(\mathbb{R})$ then the Lie algebra of $\mathcal{G}$ can be thought of informally as the set of matrices dX of $Mn(\mathbb{R})$ such that $\mathrm{Id} + \varepsilon dX$ is in $\mathcal{G}$, with ε being an infinitesimal small positive number.

Since $SE(N) = \left\{ \left[\begin{smallmatrix} R & \tau \\ 0 & 1 \end{smallmatrix}\right] \mid R \text{ rotation matrix} \right\}$ is a closed subgroup of $GLn(\mathbb{R})$, we can find the form of it Lie algebra by looking at $\mathrm{Id} + \varepsilon dX$. In order for $\mathrm{Id} + \varepsilon dX$ to be in $SE(N)$ we obviously need dX to be of the form $\left[\begin{smallmatrix} dR & d\tau \\ 0 & 0 \end{smallmatrix}\right]$. The only remaining constrain is to get a rotation matrix in the upper left block. It is thus required that $\mathrm{Id} + \varepsilon dR$ be in the rotation matrix group. We thus only need that $(\mathrm{Id} + \varepsilon dR).(\mathrm{Id} + \varepsilon dR)^T = \mathrm{Id}$ in the first order sense. This is achieved if and only if $dR + dR^T = 0$, i.e., dR is a skew-symmetric matrix.

Derivation of (3.14)

Let r be a 2D rigid body transformation represented in the Euclidean embedding space of 3×3 matrices by $\Theta(r) = \left[\begin{smallmatrix} R_\alpha & \tau \\ 0 & 1 \end{smallmatrix}\right]$. The mapping of a point $p = [p_x; p_y]$ through the transformation r can be expressed using the embedding space of 3×3 matrices as

$$w(\Theta(r), p) = \begin{bmatrix} \cos(\alpha)p_x - \sin(\alpha)p_y + \tau_x \\ \sin(\alpha)p_x + \cos(\alpha)p_y + \tau_y \end{bmatrix}$$

Let $X = \left[\begin{smallmatrix} a_1 & a_4 & a_7 \\ a_2 & a_5 & a_8 \\ a_3 & a_6 & a_9 \end{smallmatrix}\right]$ be an element of the embedding space. We have

$$w(X, p) = \begin{bmatrix} a_1 p_x + a_4 p_y + a_7 \\ a_2 p_x + a_5 p_y + a_8 \end{bmatrix}$$

and thus

$$J^{w_p} = \frac{\partial w(X,p)}{\partial [a_1, \ldots, a_9]} = \left[\begin{smallmatrix} p_x & 0 & 0 & p_y & 0 & 0 & 1 & 0 & 0 \\ 0 & p_x & 0 & 0 & p_y & 0 & 0 & 1 & 0 \end{smallmatrix}\right].$$

By stacking the basis vectors of the Lie Algebra given in (3.13) we get:

$$\boldsymbol{e}_\Theta = \left[\begin{smallmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{smallmatrix}\right].$$

Using a simple matrix multiplication we find:

$$J^{w_p}.\boldsymbol{e}_\Theta = \left[\begin{smallmatrix} -p_y & 1 & 0 \\ p_x & 0 & 1 \end{smallmatrix}\right]. \tag{3.15}$$

Chapter 4

Efficient Diffeomorphic Image Registration

Table of Contents

4.1 Motivation: Fast Compensation of Tissue Deformation

In the previous chapter, we have seen that linear transformation spaces are sufficient for many applications of biomedical image registration and we presented an efficient framework for solving this problem. There are however a number of cases where linear transformations are simply inadequate. There are three main causes [Goshtasby 03, Hill 01] that can require the use of non-rigid registration[1]. The first is linked to inherent changes of the imaged object that might be due to biological

[1]Throughout the manuscript we use the widely accepted term of *non-rigid* image registration to refer to the class of methods where the images to be registered have geometric differences that cannot be accounted for by simple transformations such as global translation, affine and projective transformations.

processes such as microcirculation, absorption of the fluorescent dye or that might be related to morphological changes such as the one appearing because of tissue motion and compression, surgery, aging or disease evolution. Secondly, non-rigid registration can help us measure structural variability among different subjects such as the difference in shapes of subcortical structures. The third type of non-rigid deformations that might appear within biomedical images is caused by the imaging devices. We have previously mentioned that Cellvizio is a laser scanning device and that, as such, the displacement of an imaged object with respect to the optical microprobe results in geometric distortions instead of a simple blur. Severe geometric deformations can also appear in macroscopic images such as with echo planar image (EPI) data obtained in functional imaging or PET [Qiao 07]. In this thesis which is focused on the registration of fibered confocal microscopy images, we have been confronted to both the first and third causes of geometric deformations because we have mainly been working with single video sequences. In the future however, it might well be possible to face the second type of deformations, for example if we were to look at computer aided diagnosis tools for Cellvizio.

Although non-rigid image registration has been a very active area of research for some time, it is still widely accepted that more work is needed [Goshtasby 03, Hill 01, Pennec 06b]. In this chapter, we build on the efficient tools we presented in Chapter 3 and show that our framework can be extended to provide a very efficient non-parametric diffeomorphic image registration algorithm. This work was first proposed in [Vercauteren 07b, Vercauteren 07d]. The goal of this chapter is mainly to provide a powerful non-rigid registration scheme that can be used to compensate for tissue deformation in Cellvizio images and can be integrated within more complex processing pipelines such as the mosaicing algorithm we present in Chapter 5. However, since the non-rigid registration scheme we designed has proved to be very effective for other biomedical imaging problems, this chapter is written with a larger perspective in mind. We focus on both Cellvizio images and macroscopic 3D clinical data. Furthermore, to make this work available to a wider audience, we proposed in [Vercauteren 07c] an open-source implementation based on the Insight Toolkit (ITK) [Ibáñez 05]. It can be downloaded at `http://hdl.handle.net/1926/510`.

We have seen that the ESM provides a very efficient alternative for the problem of linear image registration. Looking at non-rigid image registration, one of the most efficient methods is the demons algorithm proposed by Thirion [Thirion 98]. Several variants of the algorithm have been proposed depending on how the *forces* are computed. In [Wang 05, Rogelj 06] an *ad hoc* symmetrization of the demons forces similar to the one proposed by Thirion was shown to improve the results of the original demons algorithm. In [Pennec 99] the authors showed that the demons algorithm had connection with gradient descent schemes. However, to the best of our knowledge, the different variants of the demons have not been given a strong unified theoretical justification.

Our first contribution in this chapter, Section 4.2, is to show that the image registration framework presented in Chapter 3 provides strong theoretical justification for the demons algorithm and that the different variants are related to the use

of different optimizers. One of the main results of this theoretical analysis we presented in [Vercauteren 07b] is to show that the symmetric forces variant is related to the ESM scheme of [Benhimane 04, Malis 04]. This study thus explains why, from a theoretical point of view, the symmetric forces demons algorithm seems to be more efficient in practice. *Another contribution* of this chapter in Section 4.3 is to provide evidence that, in practice, using symmetric forces indeed leads to a higher convergence rate.

One of the main limitations of the demons algorithm is that it does not provide diffeomorphic transformations contrarily to the algorithms developed in [Beg 05, Marsland 04]. Diffeomorphic transformations are a requirement in the growing field of computational anatomy. They are powerful for other problems as well, as they preserve the topology of the objects and prevent from introducing folding which is often physically impossible. Finally, diffeomorphisms are considered to be a good working framework when no additional information about the spatial transformation is available. *The main contribution* of this chapter in Section 4.4 is to show that the alternate optimization scheme of the demons algorithm can be used in combination with the Lie group structure on diffeomorphic transformations of [Arsigny 06a]. In contrast to many diffeomorphic registration schemes, our diffeomorphic demons algorithm, which we first proposed in [Vercauteren 07d], is computationally efficient since in practice it only replaces an addition of displacement fields by a few compositions of non-parametric transformations. Our approach is evaluated in Section 4.5 in both a simulated and a realistic registration setup. We show that in addition to being diffeomorphic, our algorithm provides results that are similar to the ones from the demons but with transformations that are much smoother and closer to the true ones in terms of Jacobians.

4.2 An Insight into the Demons Algorithm

In [Thirion 98], the author proposed to consider non-parametric non-rigid registration as a diffusion process. He introduced *demons* that push according to local characteristics of the images in a similar way Maxwell did for solving the Gibbs paradox. The forces are inspired from the optical flow equations [Barron 94] and the method alternates between computation of the forces and regularization by a simple Gaussian smoothing. This results in a computationally efficient algorithm compared to other non-rigid registration procedures such as those based on linear elasticity [Christensen 97]. Several teams [Bro-Nielsen 96, Cachier 03, Modersitzki 04, Pennec 99] have worked towards providing a theoretical framework for the demons in order to understand and potentially modify the underlying assumptions.

The goal of this section is not to propose a novel non-rigid image registration algorithm but rather to build on the previous work and improve the understanding of the demons algorithm. We first expand the alternate optimization framework of [Pennec 99] and show that the different variants of the algorithm can all be cast into the image registration framework of Chapter 3. The symmetric forces variant

was first proposed (but not analyzed) by Thirion as one possible expression of the demons forces. In [Rogelj 06, Wang 05] a similar *ad hoc* symmetrization of the forces proved to boost the results of the demons algorithm. One of the main results of our theoretical analysis is to show that the symmetric forces demons can be cast to the ESM optimization method of [Benhimane 04, Malis 04]. We therefore have both empirical and theoretical evidence that this variant should be the most efficient one. Our second goal is to verify this evidence in a practical case study.

4.2.1 A Deeper Understanding of the Alternate Optimization of the Demons

Given a *fixed image* $F(.)$ and a *moving image* $M(.)$, non-parametric image registration is treated as an optimization problem that aims at finding the displacement of each pixel so as to get a reasonable alignment of the images. Similarly to Chapter 3, the transformation $s(.)$, $p \mapsto s(p)$, models the spatial mapping of points from the fixed image space to the moving image space. In many cases, non-parametric spatial transformations will be described by a displacement field $\boldsymbol{s}$ which is simply added to an identity transformation to get the non-parametric transformation s:

$$s : p \mapsto p + \boldsymbol{s}(p)$$

The similarity criterion $\mathrm{Sim}\,(.,.)$ measures the resemblance of two images. In this chapter we will again only consider the mean squared error which forms the basis of intensity-based registration:

$$\mathrm{Sim}\,(F, M \circ s) = \frac{1}{2}\left\|F - M \circ s\right\|^2 = \frac{1}{2\left|\Omega_P\right|}\sum_{p\in\Omega_P}\left|F(p) - M(s(p))\right|^2, \qquad (4.1)$$

where Ω_P is the region of overlap between F and $M \circ s$.

A simple optimization of (4.1) over the space of non-parametric transformations leads to an ill-posed problem with unstable and non-smooth solutions. To avoid this and possibly add some *a priori* knowledge, a regularization term $\mathrm{Reg}\,(s)$ is often introduced to get the global energy

$$E(s) = \frac{1}{\sigma_i^2}\mathrm{Sim}\,(F, M \circ s) + \frac{1}{\sigma_T^2}\mathrm{Reg}\,(s),$$

where σ_i accounts for the noise on the image intensity, and σ_T controls the amount of regularization we need.

This energy indeed provides a well-posed framework but the mixing of the similarity and the regularization terms leads in general to computationally intensive optimization steps. On the other hand the demons algorithm of [Thirion 98] provides a very efficient registration scheme but has often been considered as somewhat *ad hoc*.

In order to cast the demons algorithm into a minimization of a well posed criterion, it was proposed in [Cachier 03] to introduce a hidden variable in the registration process: correspondences. The idea is to consider the regularization criterion

as a prior on the smoothness of the transformation s. Instead of requiring that point correspondences between image pixels, a non-parametric spatial transformation c, be exact realizations of the spatial transformation s, one allows some error at each image point. Considering a Gaussian noise on displacements, we end up with the global energy:

$$E(c,s) = \|\frac{1}{\sigma_i}(F - M \circ c)\|^2 + \frac{1}{\sigma_x^2}\,\mathrm{dist}\,(s,c)^2 + \frac{1}{\sigma_T^2}\,\mathrm{Reg}\,(s) \tag{4.2}$$

where σ_x accounts for a spatial uncertainty on the correspondences. We classically have $\mathrm{dist}\,(s,c) = \|c - s\|$ and $\mathrm{Reg}\,(s) = \|\nabla s\|^2$ but the regularization can, for example, also be modified to address fluid-like constraints [Cachier 03].

The interest of this auxiliary variable is that an alternate optimization over c and s decouples the complex minimization into simple and very efficient steps. The first step solves for the correspondences by optimizing $\|\frac{1}{\sigma_i}(F - M \circ c)\|^2 + \frac{1}{\sigma_x^2}\,\mathrm{dist}\,(s,c)^2$, with respect to c and with s being given, by making a step from $c = s$. The second step solves for the regularization by optimizing $\frac{1}{\sigma_x^2}\,\mathrm{dist}\,(s,c)^2 + \frac{1}{\sigma_T^2}\,\mathrm{Reg}\,(s)$, with respect to s and with c being given.

This minimization has a closed-form solution using a single convolution when the regularization is quadratic and uniform. Given the harmonic regularization criteria $\|\nabla s\|^2$, for example, it can be shown that the optimal regularized deformation field is the convolution of the correspondence field by a Gaussian kernel. More elaborate regularization terms can lead to advanced vectorial filters [Cachier 04]. In this work, we focus on the first step of this alternate minimization and refer the reader to [Bro-Nielsen 96, Cachier 03, Modersitzki 04] for a detailed coverage of the regularization questions.

4.2.2 Compositive and Additive Demons

Let us consider the complete space of non-parametric spatial transformations. As with any given spatial transformation, the most natural operation we can endow this space with, is given by the composition of spatial transformations. To optimize for the correspondences, it is thus relevant to look for a small deformation that will be composed with the current estimate. Let this small deformation be described by a dense displacement field $\boldsymbol{u}$. In its seminal paper [Thirion 98], the author proposed to make, at each iteration what we call a compositive adjustment: $s \circ (\mathrm{Id} + \boldsymbol{u})$. The compositive demons algorithm can then be described by the following iterations:

Algorithm 5 (Compositive Demons Algorithm)

- Given the current transformation s, compute a correspondence update field $\boldsymbol{u}$ by minimizing $E_s^{\mathrm{corr}}(\boldsymbol{u}) = \|F - M \circ s \circ (\mathrm{Id} + \boldsymbol{u})\|^2 + \frac{\sigma_i^2}{\sigma_x^2}\|\boldsymbol{u}\|^2$ with respect to $\boldsymbol{u}$
- For a fluid-like regularization let $\boldsymbol{u} \leftarrow K_{\mathrm{fluid}} \star \boldsymbol{u}$. The convolution kernel will typically be a Gaussian kernel.

- Let $c \leftarrow s \circ (\mathrm{Id} + \boldsymbol{u})$
- For a diffusion-like regularization let $\boldsymbol{s} \leftarrow K_{\text{diff}} \star \boldsymbol{c}$ (else let $\boldsymbol{s} \leftarrow \boldsymbol{c}$). The convolution kernel will also typically be a Gaussian kernel. ∎

A different approach relies on the fact that the space of dense displacement fields forms a simple vector space with respect to the addition. Several teams have thus been using an additive adjustment rule $\boldsymbol{s} + \boldsymbol{u}$ instead of the compositive update rule [Cachier 03, Ibáñez 05, Modersitzki 04, Pennec 99]. This type of additive update is closer to the classical update rules used in Newton methods on vector spaces. It should however be noted that it disregards the fact that we work on spatial transformations. While it is natural to compose spatial transformations, their addition has no geometric meaning. From the practical point of view, it can be argued that the composition $s \circ (\mathrm{Id} + \boldsymbol{u})$ requires to warp the dense displacement field $\boldsymbol{s}$ with $\boldsymbol{u}$ and to add the result with $\boldsymbol{u}$. The addition rule is therefore less computationally expensive. We show however in our experiments that using this additive structure, which is not consistent with our transformation space, leads to slower convergence and less accurate solutions. Let us however give an overview of how these additive adjustments have been used within the demons algorithm. From an initial non-parametric transformation s, the following iterations are performed until convergence:

Algorithm 6 (Additive Demons Algorithm)

- Given the current transformation s, compute a correspondence update field $\boldsymbol{u}$ by minimizing $E_s^{\text{corr}}(\boldsymbol{u}) = \|F - M \circ (s + \boldsymbol{u})\|^2 + \frac{\sigma_i^2}{\sigma_x^2} \|\boldsymbol{u}\|^2$ with respect to $\boldsymbol{u}$
- For a fluid-like regularization let $\boldsymbol{u} \leftarrow K_{\text{fluid}} \star \boldsymbol{u}$. The convolution kernel will typically be a Gaussian kernel.
- Let $\boldsymbol{c} \leftarrow \boldsymbol{s} + \boldsymbol{u}$
- For a diffusion-like regularization let $\boldsymbol{s} \leftarrow K_{\text{diff}} \star \boldsymbol{c}$ (else let $\boldsymbol{s} \leftarrow \boldsymbol{c}$). The convolution kernel will also typically be a Gaussian kernel. ∎

4.2.3 Demons Forces

We see that, in Algorithm 5 and Algorithm 6, the minimization of $E_s^{\text{corr}}(\boldsymbol{u})$ is very close to the mean squared error image registration problem of (4.1). The goal is to find an optimal update field $\boldsymbol{u}$. Since we deal with a least-square problem, all the methods we present in this chapter rely on a linearization of the first inner term in $E_s^{\text{corr}}(\boldsymbol{u})$ and use a Gauss-Newton like approach.

Let us consider the intensity difference at a given point, $\varphi_p(s) = F(p) - M \circ s(p)$. Let $\varphi_p^s(\boldsymbol{u}) = F(p) - M \circ s \circ (\mathrm{Id} + \boldsymbol{u})(p)$ in the compositive case and $\varphi_p^s(\boldsymbol{u}) = F(p) - M \circ (s + \boldsymbol{u})(p)$ in the additive case. Let us assume that the following linearization

is available:

$$\varphi_p^s(\boldsymbol{u}) \approx \varphi_p^s(0) + J^p.\boldsymbol{u}(p) = F(p) - M \circ s(p) + J^p.\boldsymbol{u}(p).$$

We will see in Section 4.2.4 that such a linearization can for example be given by a Taylor expansion or an ESM scheme. The approximation order will then depend on this choice of J^p. Such a linearization can be used to rewrite the correspondence energy used in the demons algorithm:

$$E_s^{\text{corr}}(\boldsymbol{u}) \approx \frac{1}{2\,|\Omega_P|} \sum_{p \in \Omega_P} \left\| \begin{bmatrix} F(p) - M \circ s(p) \\ 0 \end{bmatrix} + \begin{bmatrix} J^p \\ \frac{\sigma_i(p)}{\sigma_x} I \end{bmatrix} .\boldsymbol{u}(p) \right\|^2,$$

where we recall that Ω_P is the overlap between F and $M \circ s$.

As opposed to the global transformation case (e.g. 2D rigid body transformations) we see that here, the approximations given for each pixel are independent from each other. This greatly simplifies the minimization of E_s^{corr} by splitting it into very simple systems for each pixel. We indeed only need to solve, at each pixel p, the following normal equations:

$$\begin{bmatrix} J^{pT} & \frac{\sigma_i(p)}{\sigma_x} I \end{bmatrix} . \begin{bmatrix} J^p \\ \frac{\sigma_i(p)}{\sigma_x} I \end{bmatrix} .\boldsymbol{u}(p) = - \begin{bmatrix} J^{pT} & \frac{\sigma_i(p)}{\sigma_x} I \end{bmatrix} . \begin{bmatrix} F(p) - M \circ s(p) \\ 0 \end{bmatrix}$$

which simplifies into

$$\left(J^{pT}.J^p + \frac{\sigma_i^2(p)}{\sigma_x^2} I \right) .\boldsymbol{u}(p) = -(F(p) - M \circ s(p)).J^{pT}.$$

From the Sherman-Morrison formula, a.k.a. matrix inversion lemma, shown in the appendix page 76, we finally have:

$$\boldsymbol{u}(p) = -\frac{F(p) - M \circ s(p)}{\|J^p\|^2 + \frac{\sigma_i^2(p)}{\sigma_x^2}} J^{pT} \tag{4.3}$$

If we use the local estimation of the image noise $\sigma_i(p) = |F(p) - M \circ c(p)|$, and the ESM approximation $J^p = \left(\nabla_p^T F + \nabla_p^T (M \circ s)\right)/2$ shown in Section 4.2.4, we end up with the exact expression of the symmetric forces demons algorithm. Note that as shown in the appendix page 76, with this choice of image noise, σ_x controls the maximum step length

$$\|\boldsymbol{u}(p)\| \leq \frac{\sigma_x}{2}. \tag{4.4}$$

4.2.4 Linearization of the Intensity Difference

We have shown that if a linearization of the intensity difference is available, we can explain the forces used in the demons algorithm. In this section, we will show how the different forces can be connected to different types of linearization. Obviously the linearization depends on the type of update rule, compositive or additive, that we use. The reader not interested in the technical details will find in Table 4.1 a summary of the conclusions of the derivations we show here and may jump directly to Section 4.3.

Table 4.1: Demons variants according to the different adjustment rules and the demons forces. In this table, *Thirion* refers to the variants proposed in [Thirion 98], *ITK* refers to the implementation of the Insight Toolkit [Ibáñez 05], and *Pennec 99* refers to the variant proposed in [Pennec 99]. The schemes that can readily be explained by a Gauss-Newton scheme are appended with *(GN)* while those that almost fit in the ESM framework are denoted by *(ESM-like)*. The variants that are implemented in ITK can only be seen as approximate gradient schemes as they use an additive adjustment rule with forces derived for the compositive demons.

Used J^p \ Adjustment	Compositive $c \leftarrow s \circ (\mathrm{Id} + \boldsymbol{u})$	Additive $\boldsymbol{c} \leftarrow \boldsymbol{s} + \boldsymbol{u}$
$\nabla_p^T (M \circ s)$	Thirion Mov. (GN)	
$\nabla_p^T F$	Thirion Fix.	ITK Fix.
$\big(\nabla_p^T F + \nabla_p^T (M \circ s)\big)/2$	Thirion Sym. (ESM-like)	ITK Sym.
$\nabla_{s(p)}^T M$		Pennec 99 (GN)

Compositive Adjustments

In Chapter 3 and [Vercauteren 07b], we have shown that for any Lie group of spatial transformations endowed with the composition operation, a Hessian-free second order approximation of the mean squared error similarity criterion could be obtained using the Lie Algebra and some nice properties of the image registration problem.

A very similar analysis can be used here. It should however be noted that since we deal with the complete space of non-parametric spatial transformations, not all transformations are invertible. Thus, we do not have a Lie group but only a monoid. It is not too problematic in this case since we can avoid an actual use of the invertibility property. One of the key technical points in the ESM framework relied on the embedding of the Lie group within an Euclidean space. In the current setting, such an embedding is obvious as we can simply use the complete space of dense displacement fields. The derivations can thus be simplified. This will also help us understand a little more the ESM framework.

As previously, let $\varphi_p^s(\boldsymbol{u}) = F(p) - M \circ s \circ (\mathrm{Id} + \boldsymbol{u})(p)$, we can use a Taylor expansion of it to get the following linearization:

$$\varphi_p^s(\boldsymbol{u}) = \varphi_p^s(0) + J_s^{\varphi_p}.\boldsymbol{u} + \frac{1}{2}\boldsymbol{u}^T.H_s^{\varphi_p}.\boldsymbol{u} + O(\|\boldsymbol{u}\|^3),$$

where $[J_s^{\varphi_p}]_i = \frac{\partial}{\partial u_i}\varphi_p^s(\boldsymbol{u})\big|_{\boldsymbol{u}=0}$ and $[H_s^{\varphi_p}]_{ij} = \frac{\partial^2}{\partial u_i \partial u_j}\varphi_p^s(\boldsymbol{u})\big|_{\boldsymbol{u}=0}$. Let us first look at the first-order terms. We see that:

$$\begin{aligned}
\frac{\partial \varphi_p^s(\boldsymbol{u})}{\partial \boldsymbol{u}(q)^T}\Big|_{\boldsymbol{u}=0} &= -\frac{\partial M \circ s\big((\mathrm{Id}+\boldsymbol{u})(p)\big)}{\partial \boldsymbol{u}(q)^T}\Big|_{\boldsymbol{u}=0} \\
&= -\frac{\partial M \circ s(p+\boldsymbol{u}(p))}{\partial \boldsymbol{u}(q)^T}\Big|_{\boldsymbol{u}=0} = -\delta_{p,q}\frac{\partial M \circ s(\rho)}{\partial \rho^T}\Big|_{\rho=p} \\
&= -\delta_{p,q}\nabla_p^T (M \circ s),
\end{aligned}$$

where $\delta_{p,q}$ is the Kronecker delta. This gives us a first order expansion of φ_p:

$$\varphi_p^s(\boldsymbol{u}) = \varphi_p^s(0) - \nabla_p^T(M \circ s).\boldsymbol{u}(p) + O(\|\boldsymbol{u}\|^2). \tag{4.5}$$

We see in this case that, by plugging $J^p = -\nabla_p(M \circ s)$ into (4.3), we get a Gauss-Newton step for the compositive update rule.

We also see that since the Jacobian of φ_p is zero for indexes that does not correspond to the pixel p, it will also be the case for the Hessian matrix[2]. We denote by H^p the $D \times D$ sub-matrix containing the non-zero elements.

Following the ESM framework, let us express the Hessian matrix through a Jacobian expressed at the optimal update step $\boldsymbol{u}_s^{\text{opt}}$. This step is defined through the optimal spatial transformation: $s^{\text{opt}} = s \circ (\text{Id} + \boldsymbol{u}_s^{\text{opt}})$. Note that this definition may not be well-defined since we are not on a Lie group anymore. Nevertheless, we assume that we do have such an optimal update step. This assumption proves to be very efficient in practice. Let us use a Taylor expansion of an extension of the Jacobian:

$$\begin{aligned}\frac{\partial \varphi_p^s(\boldsymbol{v})}{\partial \boldsymbol{v}(p)^T}\Big|_{\boldsymbol{v}=\boldsymbol{u}} &= \frac{\partial \varphi_p^s(\boldsymbol{v})}{\partial \boldsymbol{v}(q)^T}\Big|_{\boldsymbol{v}=0} + \boldsymbol{u}(p)^T.H^p + O(\|\boldsymbol{u}\|^2) \\ &= -\nabla_p^T(M \circ s) + \boldsymbol{u}(p)^T.H^p + O(\|\boldsymbol{u}\|^2).\end{aligned}$$

By using the fact that $\boldsymbol{u}(p)^T.H^p$ can now be expressed by the difference of two Jacobians, we get:

$$\varphi_p^s(\boldsymbol{u}) = \varphi_p^s(0) + \frac{1}{2}\left(\frac{\partial \varphi_p^s(\boldsymbol{v})}{\partial \boldsymbol{v}(p)^T}\Big|_{\boldsymbol{v}=\boldsymbol{u}} - \nabla_p^T(M \circ s)\right)\boldsymbol{u}(p) + O(\|\boldsymbol{u}\|^3).$$

By the simple change of variable $\boldsymbol{w} = \boldsymbol{v} - \boldsymbol{u}$, we also have:

$$\begin{aligned}\frac{\partial \varphi_p^s(\boldsymbol{v})}{\partial \boldsymbol{v}(p)^T}\Big|_{\boldsymbol{v}=\boldsymbol{u}} &= -\frac{\partial M \circ s(p + \boldsymbol{v}(p))}{\partial \boldsymbol{v}(p)^T}\Big|_{\boldsymbol{u}=\boldsymbol{v}} \\ &= -\frac{\partial M \circ s(p + \boldsymbol{u}(p) + \boldsymbol{w}(p))}{\partial \boldsymbol{w}(p)^T}\Big|_{\boldsymbol{w}=0} \\ &= -\nabla_{p+\boldsymbol{u}(p)}^T(M \circ s) \\ &= -\nabla_p^T(M \circ s \circ (\text{Id} + \boldsymbol{u})) \;.\; \text{Jac}(\text{Id} + \boldsymbol{u})(p)^{-1},\end{aligned}$$

where $\text{Jac}(\text{Id} + \boldsymbol{u})(p)$ is the Jacobian of the adjustment $\text{Id} + \boldsymbol{u}$ evaluated at p.

With the optimal step $\boldsymbol{u}_s^{\text{opt}}$, the fixed and moving image should be equivalent. We thus have, up to a noise term,

$$\nabla_p^T(M \circ s \circ (\text{Id} + \boldsymbol{u}_s^{\text{opt}})) = \nabla_p^T(M \circ s^{\text{opt}}) \approx \nabla_p^T F.$$

[2]In theory, on a smooth manifold, the Hessian matrix should account for the metric. Cross-terms might thus appear if we do not use L^2 norm. These technical details are however out of the scope of this thesis.

Moreover, since we only look at small deformations, we can make the hypothesis that the Jacobian of the adjustment $\mathrm{Jac}(\mathrm{Id}+\boldsymbol{u}_s^{\mathrm{opt}})$ is not to far from the identity. We then find that

$$\frac{\partial \varphi_p^s(\boldsymbol{v})}{\partial \boldsymbol{v}(p)^T}\Big|_{\boldsymbol{v}=\boldsymbol{u}_s^{\mathrm{opt}}} \approx -\nabla_p^T F.$$

Which leads to the following Hessian-free linearization:

$$\varphi_p^s(\boldsymbol{u}) \approx \varphi_p^s(0) - \frac{1}{2}(\nabla_p^T(M\circ s) + \nabla_p^T F).\boldsymbol{u}(p). \tag{4.6}$$

By plugging $J^p = -\frac{1}{2}(\nabla_p^T(M\circ s) + \nabla_p^T F)$ into (4.3), we thus get an ESM-like step for the compositive update rule.

In [Thirion 98], the author actually suggested the use of $J^p = -\nabla_p^T F$. This can be justified by a combination of (4.5) and (4.6):

$$\varphi_p^s(\boldsymbol{u}) \approx \varphi_p^s(0) - \nabla_p^T F.\boldsymbol{u}(p). \tag{4.7}$$

A different justification of a similar force can also be found in [Pennec 99].

Additive Adjustments

In the setting of additive adjustments, we also need a linearization of $\varphi_p^s(\boldsymbol{u}) = F(p) - M\circ(s+\boldsymbol{u})(p)$. We see that since $(s+\boldsymbol{u})(p) = p+\boldsymbol{s}(p)+\boldsymbol{u}(p) = s(p)+\boldsymbol{u}(p)$, we have:

$$\frac{\partial M\circ(s+\boldsymbol{u})(p)}{\partial \boldsymbol{u}(p)^T}\Big|_{\boldsymbol{u}(p)=0} = \frac{\partial M(s(p)+\boldsymbol{u}(p))}{\partial \boldsymbol{u}(p)^T}\Big|_{\boldsymbol{u}(p)=0} = \frac{\partial M(q)}{\partial q^T}\Big|_{q=s(p)} = \nabla_{s(p)}^T M$$

We then get the following first-order expansion:

$$\varphi_p^s(\boldsymbol{u}) = \varphi_p^s(0) - \nabla_{s(p)}^T M.\boldsymbol{u}(p) + O(\|\boldsymbol{u}\|^2). \tag{4.8}$$

By plugging $J^p = -\nabla_{s(p)}^T M$ into (4.3), we thus get a Gauss-Newton step for the additive update rule. A similar derivation was proposed in [Pennec 99] but as they did not see that the matrix inversion lemma could be used, they proposed an unnecessary approximation of the Hessian matrix with a scalar matrix.

We should also mention that it is very common to find implementations of the demons algorithm that use an additive adjustment rule with the forces that we justified in the compositive adjustment case. This is for example the case in the Insight Toolkit (ITK) [Ibáñez 05] which is often considered as a reference implementation in the biomedical imaging community. These variants can thus only be seen as approximate gradient schemes. One potential way of justifying them is to look at a first order approximation of the compositive adjustment rule:

$$s\circ(\mathrm{Id}+\boldsymbol{u})(p) \approx s(p) + \mathrm{Jac}(s)(p).\boldsymbol{u}(p)$$

We can see that using the compositive forces with the additive adjustment rule amounts to discarding the influence of the Jacobian $\mathrm{Jac}(s)(p)$ of the transformation

s at point p. If this seems reasonable for small deformations, it is however a very gross approximation for most of the problems we are interested in. This approach might however be reasonnable if low computationnal requirement is the priority. Additive adjustments coupled with fixed image gradient forces indeed provide the lowest complexity per update step.

4.3 Experiments: Practical Advantage of the Symmetric Forces

We have just demonstrated the theoretical advantage of the symmetric forces variant of the demons algorithm in the compositive adjustment case. On the other hand, previous studies that have used (but not theoretically justified) a very similar symmetrization of the demons forces have reported the practical advantage of this variant [Rogelj 06, Wang 05] even with the additive adjustment rule. In this section, we provide some more evidence by comparing the different variants of the demons algorithm on both a synthetic and a realistic case study.

We used the same set of parameters for all the experiments: a maximum step length of 2 pixels, a Gaussian *fluid-like* regularization with $\sigma_{\text{fluid}} = 1$ and a Gaussian *diffusion-like* regularization with $\sigma_{\text{diff}} = 1$. As previously, the emphasis is on the comparison of the various schemes and not on the final performance. Therefore, no multi-resolution scheme such as in [Hellier 01] was used.

The first experiments provide a completely controlled setup. We use a fibered confocal microscopy image as our original image. For each random experiment, we generate a smooth displacement field with a Markov random filed (MRF) sampler and warp the original image. We add some random noise both to the original and the warped image. We then run the different demons algorithms starting with an identity spatial transformation. Two conclusions can be drawn from Fig. 4.2. First, we see that, for a given type of demons forces, the compositive and additive demons converge almost at the same pace in terms of MSE. However, if we look at the harmonic energy and at the distance to the actual field, we see that, for all demons forces, the compositive demons converges faster than its additive counterpart. The second interesting fact we see here, for both for the compositive and the additive demons, is the fact that the symmetric forces variant converges faster in terms of MSE. It also behaves well in terms of smoothness of the displacement field. Finally, it is the fastest to converge in terms of distance to the actual field.

Our second setup is a more realistic case study were a gold standard is still available. We use synthetic T1 MR images from two different anatomies available from BrainWeb [Aubert-Broche 06]. These datasets are distributed along with a segmentation of eleven different tissue classes. The goal of this experiment is to show that on this more realistic case we can also draw the conclusion that symmetric forces are an advantageous option in practice. To ease the reading of the results, we simply compare the most widely used version of the demons, additive update rule and fixed image force, with is symmetric forces counterpart. We see on Fig. 4.4,

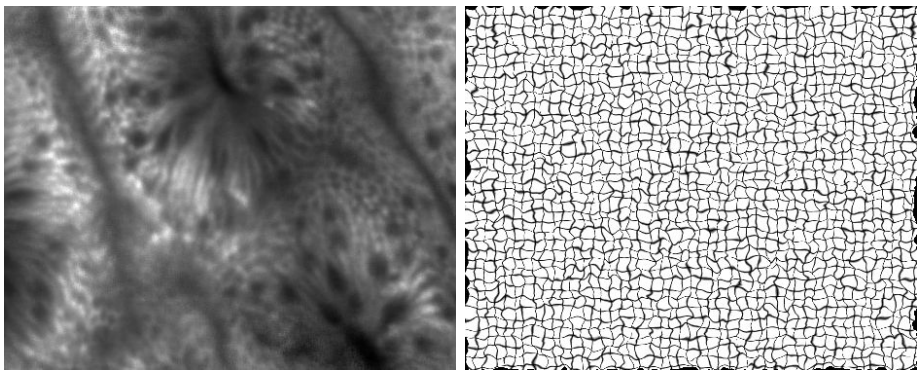


Figure 4.1: Original image (FCM) of a normal human colonic mucosa image (image courtesy of PD. Dr. A. Meining, Klinikum rechts der Isar, Munich) and one example random warp used in our controlled experimental setup.

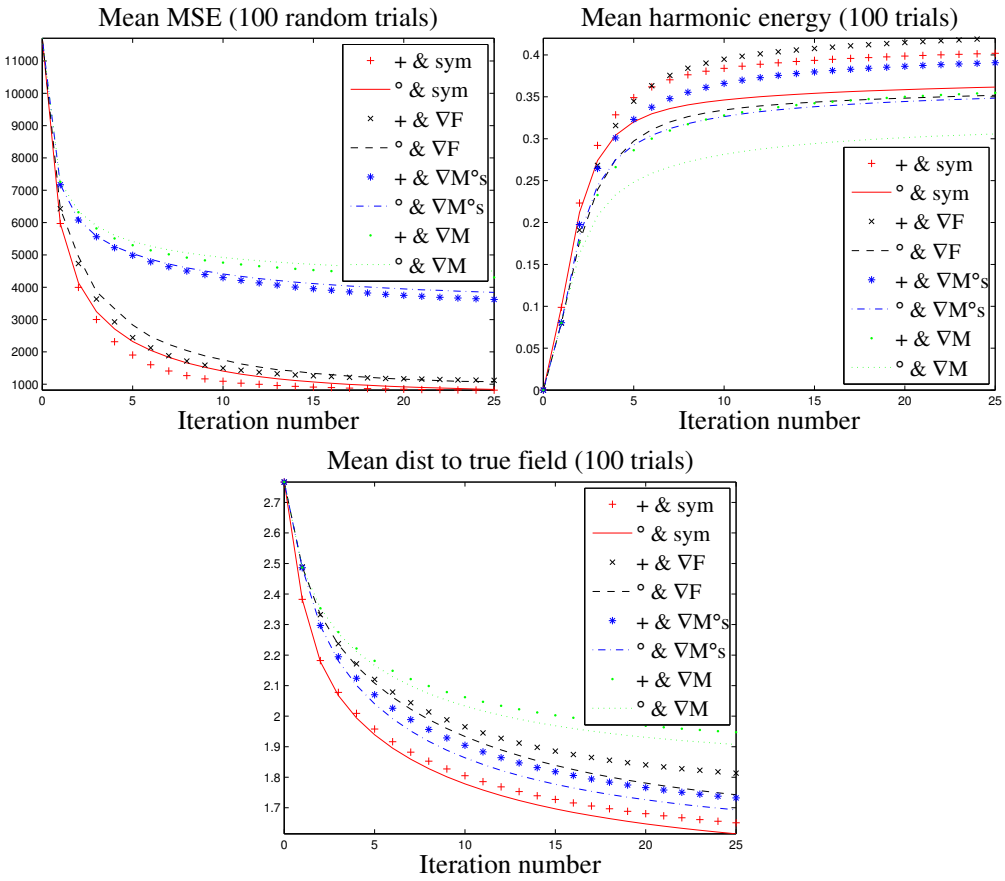


Figure 4.2: Registration on 100 random experiments such as the one in Fig. 4.1. Note the faster convergence of the symmetric forces demons in terms of images intensities agreement (MSE), smoothness of the non-rigid spatial transformation (harmonic energy) and more importantly in terms of distance to the actual spatial transformation.

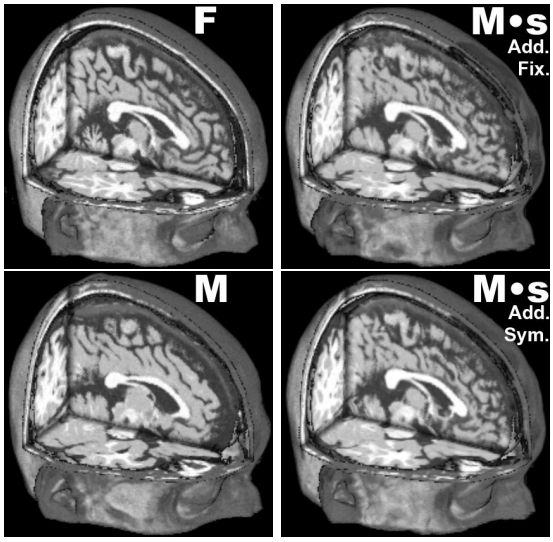


Figure 4.3: Registration of two synthetic T1-weighted MR images of distinct anatomies from BrainWeb [Aubert-Broche 06].

Table 4.2: Comparison (Dice similarity coefficient * 100) of the discrete segmentations obtained from the registration with the additive demons of the synthetic T1-weighted MR images shown in Fig. 4.3. For each tissue class, the best segmentation is obtained with the symmetric forces variant.

	CSF	GM	WM	Fat	Muscle	Skin	Skull	Vessels	Fat2	Dura	Marrow
No reg.	41.73	63.06	61.51	19.30	20.14	66.65	42.75	14.26	6.06	14.74	28.19
Add Fix.	63.41	78.99	79.23	47.74	36.40	78.57	64.91	27.21	14.75	23.13	45.05
Add Sym.	69.75	83.78	84.58	52.81	41.41	82.94	71.28	35.21	17.25	29.79	51.85

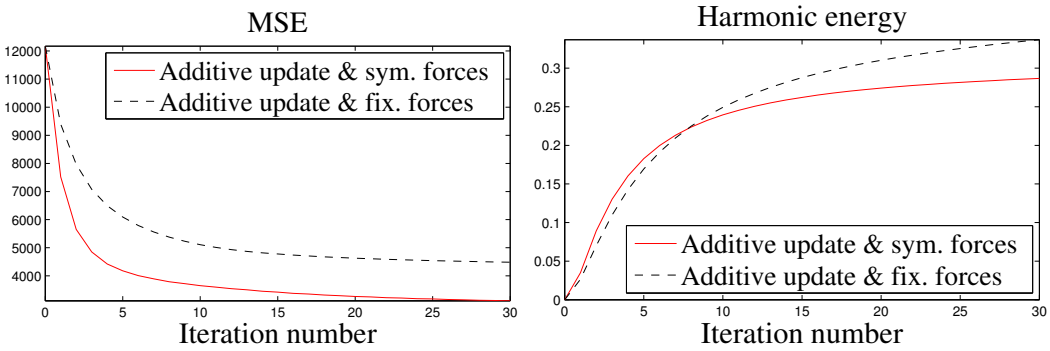


Figure 4.4: Comparison of Thirion's demons algorithm with the symmetric forces demons algorithm on the BrainWeb images shown in Fig. 4.3. Note the faster convergence of the symmetric forces demons in terms of images intensities agreement (MSE) and smoothness of the non-rigid transformation (harmonic energy).

that, on this dataset also, the symmetric forces variants converges faster in terms of MSE on the images intensities and smoothness of the displacement field. In Table 4.2, we compare the agreement between the segmentation of the fixed image and the segmentation of the moving image warped by the non-parametric spatial transformation found by the registration of the T1-weighted MR images. We show the Dice similarity index of the eleven tissue classes before any registration and after registration with the two variants of the demons we consider. We see that, for each tissue class, the best segmentation is obtained with the symmetric forces variant.

4.4 Introducing Diffeomorphisms into the Demons

One of the main limitations of both the additive and compositive demons algorithm is that it does not ensure the invertibility of the output transformations contrarily to diffeomorphic image registration algorithms. It may not be a requirement to get such diffeomorphic transformations. However, this framework may be relevant and powerful for many image registration problems. It indeed preserves the topology of the objects in the image and prevents from introducing folding which is often physically impossible. Diffeomorphisms are also considered to be a good working framework when no additional information about the spatial transformation is available. With the development of computational anatomy and in the absence of a justified physical model of inter-subject variability, statistics on diffeomorphisms also become an important topic [Arsigny 06a, Vaillant 04, Xue 06]. Diffeomorphic registration algorithms are at the core of this research field since they often provide the *input data*.

Diffeomorphic image registration usually relies on the computationally heavy solution of some partial differential equations [Beg 05, Christensen 96, Joshi 00, Marsland 04, Miller 98, Trouvé 98, Vaillant 04] or uses very small optimization steps such as in the approach of [Chefd'hotel 02]. In [Rueckert 06], the authors proposed a parametric approach by composing a set of constrained B-spline transformations. Since the composition and inversion of B-spline transformations cannot be expressed on a B-spline basis, the advantage of using a parametric approach is not clear in this case. A complex optimization scheme on constrained B-splines has also been proposed in [Noblet 05] for this problem.

In this section, we propose a non-parametric diffeomorphic image registration algorithm based on the demons algorithm. We have just shown that the original demons algorithm could be seen as an optimization procedure on the entire space of displacement fields. The main idea of our algorithm is to adapt this optimization procedure to a space of diffeomorphic transformations. We show that a Lie group structure on diffeomorphic transformations that has recently been proposed in [Arsigny 06a] can be used in combination with optimization tools on Lie groups to derive our diffeomorphic image registration algorithm.

4.4.1 A Lie Group Structure on Diffeomorphisms

Like most spatial transformation spaces, diffeomorphisms do not form a vector space. However, they can be smoothly composed and inverted. It is therefore possible to provide a Lie group structure on the space of diffeomorphisms[3]. The most straightforward way to adapt the demons algorithm to make it diffeomorphic, is to optimize the global energy (4.2) over a space of diffeomorphisms instead of the complete space of non-parametric spatial transformations. We thus need to perform an optimization procedure on a Lie group such as in [Benhimane 04, Mahony 02].

In Chapter 3, we have seen a number of Newton methods for optimization problems on Lie groups. The main idea of these methods is to find, from the current transformation s, an update step $\boldsymbol{u}$ on the Lie algebra and to use an intrinsic update rule on the Lie group through the exponential map:

$$s \leftarrow s \circ \exp(\boldsymbol{u}).$$

Such Newton methods for Lie groups are in theory well fit for diffeomorphic image registration. In practice however, it can only be used if a fast and tractable numerical scheme for the computation of the exponential is available. We would indeed have to use it at each iteration. Such an efficient scheme clearly relies on a good parameterization of the Lie group and the Lie algebra.

In the context of image registration, it has been proposed, in [Miller 98] for landmarks and [Beg 05] for images, to parameterize the space of diffeomorphic transformations using *time-varying* speed vector fields. This has the advantage of fully using the group structure. However, the computation of a displacement field requires the numerical integration of a time-varying ODE [Trouvé 98]. In [Arsigny 06a] the authors proposed a practical approximation of such a Lie group structure on diffeomorphisms by using *stationary* speed vector fields only. This has the significant advantage of yielding very fast computations of exponentials. It becomes indeed possible to use the scaling and squaring method and compute the exponential with just a few compositions of spatial transformations.

By generalizing to vector fields the equivalence that exists in the finite dimensional case between one-parameter subgroups and the exponential map, the exponential $\exp(\boldsymbol{u})$ of a smooth vector field $\boldsymbol{u}$ is defined in [Arsigny 06a] as the flow at time one of the stationary ODE,

$$\frac{\partial p(t)}{\partial t} = \boldsymbol{u}(p(t)).$$

From the properties of one-parameters subgroups ($t \mapsto \exp(t\boldsymbol{u})$), we see that, for any integer K, we have

$$\exp(\boldsymbol{u}) = \exp(K^{-1}\boldsymbol{u})^K,$$

where the power operation relates to the composition of spatial transformations. This yields the following efficient algorithm for the computation of vector fields exponentials:

[3] In theory, we should clearly state that we do not have an actual Lie group but only a pseudo-group. These technical details should be addressed by future work.

Algorithm 7 (Fast Computation of Vector Field Exponentials)

- Choose N such that $2^{-N}\boldsymbol{u}$ is close enough to 0, e.g. $\max_p \left\|2^{-N}\boldsymbol{u}(p)\right\| \leq 0.5$
- Perform an explicit first order integration: $\boldsymbol{v}(p) \leftarrow 2^{-N}\boldsymbol{u}(p)$ for all pixels. As previously we use $v = \mathrm{Id} + \boldsymbol{v}$
- Do N (not 2^N!) recursive squarings of v: $v \leftarrow v \circ v$ ∎

It should be noted that since we deal with an infinite dimensional group, this framework poses some theoretical problems that are yet to be completely solved. For example, we do not strictly deal with a Lie group but only with a pseudo-group. We do however have sufficient evidence of its well-foundedness and its effectiveness to consider this framework as one of the most efficient ways of dealing with diffeomorphisms.

4.4.2 Diffeomorphic Demons Algorithm

Let us now derive our non-parametric diffeomorphic image registration algorithm. As in Section 4.2.1, we focus on the first step of the minimization rather than on the regularization step. Let us first assume that we have a first order expansion of the intensity difference of the form:

$$F(p) - M \circ s \circ \exp(\boldsymbol{u})(p) \approx F(p) - M \circ s(p) + J^p.\boldsymbol{u}(p).$$

To get a computationally tractable expression of the correspondence energy and optimize (4.2), we choose the following distance between two diffeomorphisms: $\mathrm{dist}\,(s,c) = \left\|\mathrm{Id} - s^{-1} \circ c\right\|$. We then get

$$\mathrm{dist}\,(s, s \circ \exp(\boldsymbol{u})) = \|\mathrm{Id} - \exp(\boldsymbol{u})\| \approx \|\boldsymbol{u}\|\,.$$

By using these two expansions, we see that we get the same expression for the approximation of the correspondence energy

$$\begin{aligned} E_s^{\mathrm{corr}}(\boldsymbol{u}) &= \|F - M \circ s \circ \exp(\boldsymbol{u})\|^2 + \frac{\sigma_i^2}{\sigma_x^2}\,\mathrm{dist}\,(s, s \circ \exp(\boldsymbol{u}))^2 \\ &\approx \frac{1}{2\,|\Omega_P|} \sum_{p \in \Omega_P} \left\| \begin{bmatrix} F(p) - M \circ s(p) \\ 0 \end{bmatrix} + \begin{bmatrix} J^p \\ \frac{\sigma_i(p)}{\sigma_x} I \end{bmatrix} .\boldsymbol{u}(p) \right\|^2. \end{aligned}$$

We thus have the same expression (4.3) as in the classical demons for the demons forces. The difference is in how we compute the Jacobian J^p and how we consider $\boldsymbol{u}$. In the classical demons $\boldsymbol{u}$ is a dense displacement field whereas in the diffeomorphic demons, $\boldsymbol{u}$ is considered as a speed vector field.

From these derivations, we thus obtain our non-parametric diffeomorphic image registration algorithm:

Algorithm 8 (Diffeomorphic Demons Algorithm)

- Given the current transformation s, compute a correspondence update field $\boldsymbol{u}$ by minimizing $E_s^{\mathrm{corr}}(\boldsymbol{u})$ with respect to $\boldsymbol{u}$

- For a fluid-like regularization let $\boldsymbol{u} \leftarrow K_{\text{fluid}} \star \boldsymbol{u}$. The convolution kernel will typically be a Gaussian kernel.
- Let $c \leftarrow s \circ \exp(\boldsymbol{u})$
- For a diffusion-like regularization let $\boldsymbol{s} \leftarrow K_{\text{diff}} \star \boldsymbol{c}$ (else let $\boldsymbol{s} \leftarrow \boldsymbol{c}$). The convolution kernel will also typically be a Gaussian kernel. ∎

4.4.3 Linearization of the Intensity Difference

For the diffeomorphic demons, we also need a linearization of the intensity difference to put in the expression of the demons forces (4.3).

In Chapter 3, we have provided a rigorous derivation of this linearization for any Lie group. Contrarily to the additive and compositive demons, we now have a Lie group structure on the space of diffeomorphisms. Adapting the ESM framework is thus just a matter of instantiating the expression we showed previously.

One of the key technical points in the ESM framework relies on the embedding Θ of the Lie group within an Euclidean space. In the current diffeomorphic setting, such an embedding is quite simple. We only need to work with the complete space of dense displacement fields which obviously allows us to represent any diffeomorphic spatial transformation. With the previous notation, we simply have $\Theta(s) = \boldsymbol{s}$. Let us denote $w(\Theta(s), p)$ the expression, in the Euclidean embedding space of dense displacement fields, of the transformation of a point p through the mapping s:

$$w(\Theta(s), p) = w(\boldsymbol{s}, p) = s(p) = p + \boldsymbol{s}(p)$$

By using a Taylor expansion, we have seen in Chapter 3 that $\varphi_p^s(\boldsymbol{u}) = F(p) - M \circ s \circ \exp(\boldsymbol{u})(p)$ could be approximated as follows:

$$\varphi_p^s(\boldsymbol{u}) = \varphi_p^s(0) - \nabla_p^T(M \circ s).J^{w_p}.\boldsymbol{e}_\Theta + O(\|\boldsymbol{u}\|^2),$$

where $\nabla_p(M \circ s)$ is the gradient of the warped moving image, $J^{w_p} = \frac{\partial w(\boldsymbol{s},p)}{\partial \boldsymbol{s}^T}\big|_{\boldsymbol{s}=\Theta(\text{Id})=0}$ is the derivative of the mapping action expressed the Euclidean embedding space and $\boldsymbol{e}_\Theta$ stacks the basis vectors of the Lie algebra expressed in the Euclidean embedding space. It is easy to see that, in the current case, we have

$$\frac{\partial w(\boldsymbol{s}, p)}{\partial \boldsymbol{s}(q)^T}\Big|_{\boldsymbol{s}=0} = \frac{\partial p + \boldsymbol{s}(p)}{\partial \boldsymbol{s}(q)^T}\Big|_{\boldsymbol{s}=0} = \delta_{p,q}.\,\text{Id},$$

where $\delta_{p,q}$ is the Kronecker delta. We thus have in this special case $J^{w_p} = \text{Id}$. Furthermore, since the Lie algebra spans all the vector space of dense displacement fields, we simply use the same basis vectors for the Lie algebra and the full embedding space, meaning that in this special case, $\boldsymbol{e}_\Theta = \text{Id}$.

From this derivation, we see that a Gauss-Newton approach in the diffeomorphic demons simply provides us a demons force based on the gradient of the warped moving image:

$$J^p = -\nabla_p^T(M \circ s). \tag{4.9}$$

The most important conclusion of the ESM framework we presented in Chapter 3 was that a Hessian-free second-order linearization was possible. The expression of this approximation is given by:

$$\varphi_p^s(\boldsymbol{u}) = \varphi_p^s(0) - \frac{1}{2}\big(\nabla_p^T F + \nabla_p^T (M \circ s)\big).J^{w_p}.\boldsymbol{e}_\Theta + O(\|\boldsymbol{u}\|^3),$$

In our special case, we again find a very simple expression that can be used for a second-order approximation in the demons force:

$$J^p = -\frac{1}{2}\big(\nabla_p^T F + \nabla_p^T (M \circ s)\big). \tag{4.10}$$

By combining (4.9) and (4.10), we also see that we get a justification of the forces proposed by Thirion that are based on the gradient of the fixed image only. The complete set of variants of the demons is summarized in Table 4.3.

Table 4.3: Demons variants according to the different adjustment rules and the demons forces. In this table, *Th.* refers to the variants proposed in [Thirion 98], *ITK* refers to the implementation of the Insight Toolkit [Ibáñez 05], *Diffeo. Dem.* refers to our proposed diffeomorphic demons [Vercauteren 07d] and *Pennec 99* refers to the variant proposed in [Pennec 99]. The schemes that can readily be explained by a Gauss-Newton scheme are appended with *(GN)* while those that fit in the ESM framework are denoted by *(ESM)*. The variants that are implemented in ITK can only be seen as approximate gradient schemes as they use an additive adjustment rule with forces derived for the compositive demons.

Used J^p \ Adjustment	Compositive $c \leftarrow s \circ (\mathrm{Id} + \boldsymbol{u})$	Additive $\boldsymbol{c} \leftarrow \boldsymbol{s} + \boldsymbol{u}$	Diffeomorphic $c \leftarrow s \circ \exp(\boldsymbol{u})$
$\nabla_p^T (M \circ s)$	Th. Mov. (GN)		Diffeo. Mov. (GN)
$\nabla_p^T F$	Th. Fix.	ITK Fix.	Diffeo. Fix.
$\big(\nabla_p^T F + \nabla_p^T (M \circ s)\big)/2$	Th. Sym. (ESM-like)	ITK Sym.	Diffeo. Sym. (ESM)
$\nabla_{s(p)}^T M$		Pennec 99 (GN)	

4.5 Experiments: Diffeomorphic Registration Can Be Fast

To evaluate the performance of the diffeomorphic demons algorithm with respect to the additive and compositive demons algorithm, two sets of results are presented. We used the same set of parameters for all the experiments: Thirion's rule ($J^p = \nabla_p^T F$) with a maximum step length of 2 pixels was used in the demons force (4.3), a Gaussian *fluid-like* regularization with $\sigma_{\text{fluid}} = 1$ and a Gaussian *diffusion-like* regularization with $\sigma_{\text{diff}} = 1$ were used. Since the emphasis is on the comparison of the various schemes and not on the final performance, no multi-resolution scheme was used.

The first experiments are very similar to the ones presented in Section 4.3 and Fig. 4.1. In this section, we use this setup to compare the different variants of the

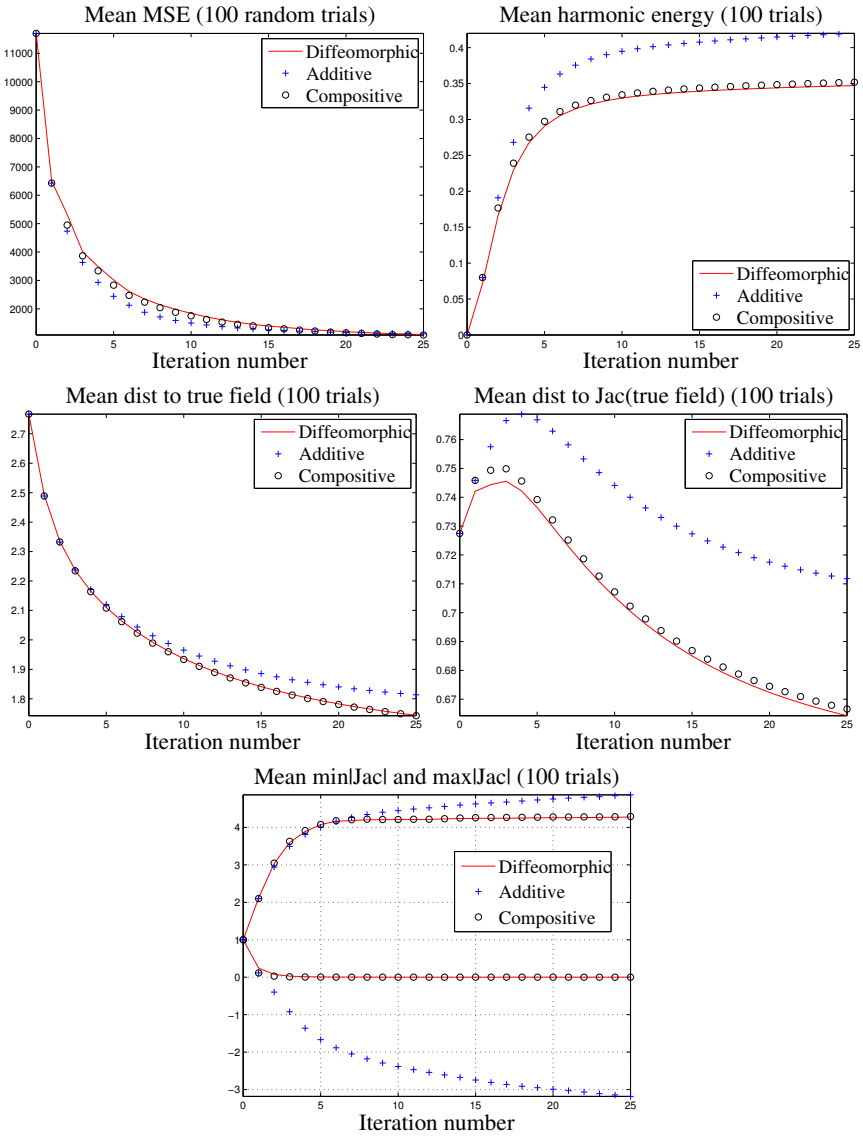


Figure 4.5: Registration on 100 random experiments such as the one presented in Fig. 4.1. We compare the three types of update rule with Thirion's demons forces based on the gradient of the fixed image. Note that for similar performance in terms of MSE and distance to the true field, the compositive and diffeomorphic demons algorithm provides much smoother results than the additive demons algorithm. We also see that we provide diffeomorphic transformations whereas $\min(|\mathrm{Jac}(s)|)$ goes way below zero with the additive demons algorithm. Furthermore, in terms of distance to the true Jacobian of the transformation, the diffeomorphic demons provide a large gain with respect to the additive demons and a small gain with respect to the compositive demons.

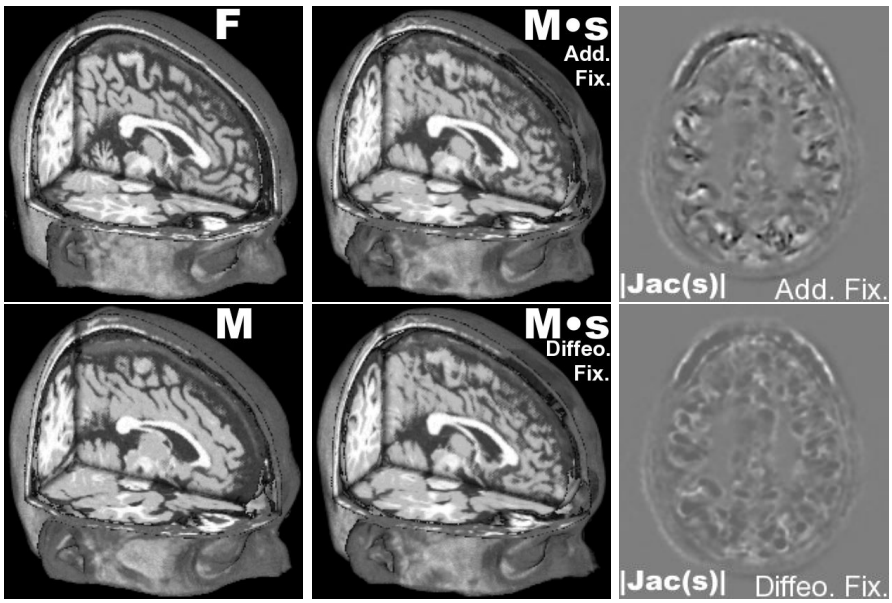


Figure 4.6: Registration of two synthetic T1 MR images of distinct anatomies. For visually similar results, our algorithm provides smoother diffeomorphic transformations.

Table 4.4: Comparison (Dice similarity coefficient * 100) of the discrete segmentations obtained from the registration of the synthetic T1-weighted MR images in Fig. 4.6.

	CSF	GM	WM	Fat	Muscle	Skin	Skull	Vessels	Fat2	Dura	Marrow
No reg.	41.73	63.06	61.51	19.30	20.14	66.65	42.75	14.26	6.06	14.74	28.19
Add Fix	63.41	78.99	79.23	47.74	36.40	78.57	64.91	27.21	14.75	23.13	45.05
Diffeo Fix	64.37	78.94	78.43	47.22	36.11	79.39	65.02	27.25	14.70	24.56	43.92

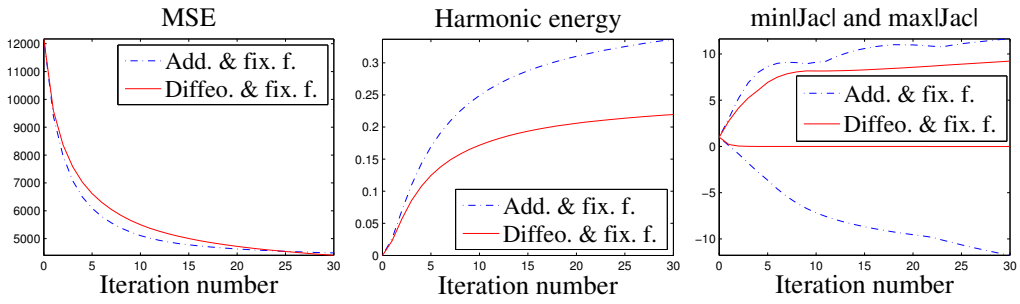


Figure 4.7: Comparison of the additive demons algorithm with the diffeomorphic demons algorithm on the BrainWeb images shown in Fig. 4.6. For similar performance in terms of MSE, our algorithm provides much smoother transformations than the additive demons algorithm. Most importantly we see that we provide diffeomorphic transformations whereas $\min(|\mathrm{Jac}(s)|)$ goes way below zero with the additive demons.

adjustment rule. They provide a completely controlled setup. We use a fibered confocal microscopy image as our original image. For each experiment, we generate a random diffeomorphic spatial transformation (by passing a Markov random field through the exponential) and warp the original image. We add some noise both to the original and the warped image. We then run the registration algorithms starting with an identity spatial transformation. We see on Fig. 4.5 that, in terms of MSE, the performance of the additive demons, the compositive demons and the diffeomorphic demons algorithm are similar. However, the distance to the true field, the harmonic energy and the minimum and maximum values of the determinant of the Jacobian of the transformation show that both the compositive demons and the diffeomorphic demons clearly outperform the additive scheme by providing much smoother spatial transformations. We also see that our diffeomorphic algorithm provides better results in terms of distance to the true Jacobian of the transformation. Note that this is accomplished with a reasonable 50% increase of computation time per iteration with respect to the computationally efficient additive demons algorithm. We could also have presented for all the adjustment rules, a comparison of the different demons forces. We chose however not to show these results as the conclusions are very similar to the ones found in Section 4.3. The symmetric forces variant outperforms the other demons forces.

Our second setup is the same as the second setup of Section 4.3. It is a more realistic case study were a gold standard is still available. We use synthetic T1 MR images from two different anatomies available from BrainWeb [Aubert-Broche 06]. These datasets are distributed along with a segmentation of eleven different tissue classes. The goal of this experiment is to show that, on this more realistic case, we can also draw the conclusion that diffeomorphic demons are an advantageous option in practice. To ease the reading of the results, we simply compare the most widely used version of the demons, additive update rule and fixed image force with its diffeomorphic counterpart. We see on Fig. 4.7 and Table 4.4 that, on this dataset also, the additive demons algorithm and our algorithm provide very similar results in terms of visual appearance, MSE and segmentation accuracy. However, we see that our algorithm does it with much better spatial transformations. We indeed get smoother deformations that are diffeomorphic.

Thanks to our open-source implementation of the diffeomorphic demons algorithm [Vercauteren 07c], our registration scheme has also been tested by an independent group, cf. [Urschler 07]. They proposed a non-rigid registration evaluation framework and benchmarked six different open-source non-rigid registration algorithms. According to their findings, “the overall best results on the evaluation experiments are given by the diffeomorphic demons algorithm.” Given the small number of tested algorithms and the current evaluation framework used in [Urschler 07], these results do not, by any means, provide a definite ranking. They can however be considered as a further evidence to support the value of our approach. A more advanced tool for benchmarking non-rigid registration algorithms might also be available in the near future [Christensen 06].

4.6 Discussion

In this section, we discuss some of the theoretical issues of our diffeomorphic demons. A recent work proposed a diffeomorphic image registration algorithm that is close to our proposition in the sense that it uses the exponential of a stationary vector field to get a diffeomorphic spatial transformation [Ashburner 07]. To provide a more objective view, the present discussion is organized around how the work of [Ashburner 07] and our diffeomorphic demons address some of the theoretical issues differently.

Discretization Questions

We have provided a principled approach to the problem of non-parametric diffeomorphic registration. However our discussion is based on spatial transformations and vector fields defined on a continuous domain. In practice of course, we need to discretize the problem. In this work, this is simply done by using the sampling grid of the images to represent the spatial transformations and the vector fields. Trilinear interpolation is then commonly used to interpolate between the sampling point. It should however be noted that this discretization is, in theory, not necessarily consistent with the continuous diffeomorphic framework. It might even be argued that diffeomorphisms are not well-posed in the context of discrete grid [Bazin 07]. A somewhat more rigorous approach might be to integrate the interpolation, within the modeling of the problem such as in [Ashburner 07] and some finite element based methods. However, since our approach is based on the composition of transformations and that the composition of finite element based transformations cannot in general be represented by a similar finite element based transformations it is not clear what the advantage would be. Furthermore, this would induce a larger computational cost.

In our results, we have shown that, in general, the diffeomorphic demons provided spatial transformations whose Jacobians remain positive. It might however not be always the case. The way we compute the Jacobian is by finite difference on the sampling grid. This computational framework is however, in theory, not necessarily consistent with the continuous diffeomorphic setup. It might indeed well be the case that a true continuous diffeomorphism has negative Jacobians when evaluated by finite difference on a sampling grid. This would simply mean that, in some sense, our sampling grid does not meet the Nyquist criterion for the spatial transformation we consider. In [Ashburner 07], the author proposed to use a computation of the Jacobian of the transformation that is more consistent with the composition of spatial transformations. If we know the Jacobian of two transformations, the analytical Jacobian of the composition at a given point can be computed through matrix multiplication. This has the advantage of better estimating the true Jacobian. However, since we still represent the composed transformation on the sampling grid, it can in this case be argued that the Jacobian is not consistent with the representation of the transformation. This would also result in a

more computationally intensive scheme as we would need to compute and store the Jacobian at each iteration. In our setting, we also use a smoothing of the transformation. Because of this smoothing, it is also difficult to get an analytical form of the Jacobian of the spatial transformation.

Compositive Demons as an Approximate Diffeomorphic Demons

In the results we showed in Section 4.5, we see that the compositive demons algorithm provides results that are very close to the results of the diffeomorphic demons, with a slight advantage to the diffeomorphic demons.

This conclusion should in fact not be so surprising. In Section 4.2, we provided one possible theoretical derivation of the compositive demons. Our explanation was based on an ESM view of the registration problem. It was however not completely possible to cast the compositive registration scheme into the ESM framework as we do not have a Lie group structure on the complete space of non-parametric spatial transformations. This is why we had to go through the derivation of the compositive demons forces rather than instantiating the ESM framework to a given transformation space. Moreover, some further approximations were necessary. In contrast, in the case of the diffeomorphic demons, the Lie group structure on diffeomorphisms, allows us to derive a very principled approach to non-rigid registration that nicely fits into the ESM framework. A different way of looking at the compositive demons is to realize that the compositive and diffeomorphic adjustment rules are equivalent up to the first-order:

$$s \circ \exp(\boldsymbol{u}) = s \circ (\mathrm{Id} + \boldsymbol{u}) + O(\|\boldsymbol{u}\|^2).$$

And since we only use small adjustments, the first-order approximation can provide good results. On a similar point, it is shown in [Arsigny 06a, Ashburner 07] that, for a given speed vector field, there is an optimal number of steps to compute the exponential. If too many steps are used, we introduce too much numerical errors. Because, in our case, we only use small speed vector fields, it might be better to use very small number of scaling and squaring steps. As $\mathrm{Id} + \boldsymbol{u}$ can also be seen as an exponential computed with zero scaling and squaring steps, it might explain that the compositive demons performs well.

In contrast, if we try to get such a retrospective view of the additive demons, we see that a crude approximation is necessary. We indeed have at the first order,

$$s \circ \exp(\boldsymbol{u})(p) \approx s(p) + \mathrm{Jac}(s)(p).\boldsymbol{u}(p),$$

which basically implies that the additive demons can be seen as an approximation of the diffeomorphic demons if we disregard the influence of the Jacobian of the current transformation. This Jacobian can actually be quite far from the identity. This can explain the divergence of the Jacobian when using additive demons.

All in all, we do believe that even if the results of the diffeomorphic demons are only slightly better when compared to the compositive demons, the algorithm

we developed provides a well-posed framework for dealing with diffeomorphisms. Future work should aim at finding how to automatically compute the number of steps in the exponential. It might well be the case that, in our case, most of the time the optimum number is zero.

Different Ways of Using the Exponential

We have seen that the parameterization of diffeomorphic transformations through a stationary speed vector field presented in [Arsigny 06a], provides a very efficient framework for dealing with diffeomorphisms. It is thus not surprising to see that several groups have started using it. In the diffeomorphic demons, we chose to use this parameterization to encode an adjustment that needs to be made to the current transformation. In [Ashburner 07], the author proposed to use a complete transformation parameterized by a speed vector field.

The approach of [Ashburner 07] has several advantages as well as several drawbacks. First of all, it is argued in [Arsigny 06a] that the exponential is a smooth one-to-one mapping between an open neighborhood of the zero speed vector field and an open neighborhood of the identity transformation. It is still not clear which elements of the group of diffeomorphic transformations can be reached with the exponential, i.e., what is the size of the target open neighborhood. In our approach, as we recursively compose the current transformation with a small exponential update, this may not be problematic. In the finite-dimensional case, for example, even if all the elements of the Lie group cannot be directly reached through the exponential, they can still all be reached through a composition of two exponentials [Wüstner 03]. When using a single parameterization in the Lie algebra, it might therefore be useful to investigate the image of the exponential map.

In the diffeomorphic demons, we chose to use a regularization through a Gaussian smoothing on the spatial transformation. This may, in theory, not always be consistent with the diffeomorphic framework. It does however provide a very efficient regularization in terms of computation time. It is also familiar to many people in the field of biomedical image analysis. The most important benefit we get from it is, however, that it allows us to decouple the minimization problem into an alternate scheme composed of very easy steps. In [Ashburner 07], the author chose to introduce a regularization term on the speed vector field that is used to parameterize the complete diffeomorphisms. While this is consistent with its use of the exponential and allows him to propose an inverse consistent scheme, it is somewhat difficult to interpret the meaning of this regularization and, most importantly, it introduces coupling in the minimization problem. Similarly, since the exponential is not used around zero, we need to compute the derivative of the exponential far from the identity. This introduces additional tight coupling. It is therefore necessary to solve a very large system of equations with a non-trivial sparsity pattern. In our case, one optimization step requires a time of the order of a second, whereas in [Ashburner 07] the author reports a computation time of around a minute.

4.7 Conclusions

The adequacy of the ESM for linear image registration led us to revisit non-rigid registration and especially Thirion's demons algorithm. We showed that the demons algorithm could be seen as an optimization procedure on the entire space of displacement fields. By using the ESM, the matrix inversion lemma and a local estimation of the image noise, we improved our understanding of the demons algorithm. A pertinent comparison between the different variants of the demons was provided. Our analysis predicted a theoretical advantage to the symmetric forces variant of the demons algorithm which we confirmed on the practical side.

The final goal of understanding an algorithm is to improve it. One of the main limitations of the demons algorithm is that it does not provide diffeomorphic transformations contrarily to the algorithms developed in [Beg 05, Marsland 04]. By combining a recently developed Lie group framework on diffeomorphisms and the optimization procedure for Lie groups we presented in Chapter 3, we showed that the framework in which we cast the demons algorithm could be adapted to provide non-parametric diffeomorphic transformations. Our experiments have shown that our algorithm provides, with respect to the additive demons algorithm, very similar results in terms of MSE. This is however achieved with diffeomorphic transformations that are smoother and closer to the true transformations in terms of Jacobians.

Thanks to the open-source implementation of our diffeomorphic demons we proposed in [Vercauteren 07c], our algorithm has been successfully tested by several independent groups. In [Urschler 07], the authors reported that our algorithm outperformed several other non-rigid registration schemes. Our algorithm has also been integrated into MedINRIA, the free medical image navigation and research tool of Asclepios research group, INRIA Sophia Antipolis [Toussaint 07]. Thanks to this, it is already used within the European project IST-2004-027749 Health-e-Child.

Appendix

Derivation of (4.3)

The Sherman-Morrison formula, a.k.a. matrix inversion lemma, provides an explicit formula for the inversion of a sum of an invertible matrix A and a dyadic product of a column vector u and a row vector v:

$$(A + u.v)^{-1} = A^{-1} - \frac{A^{-1}.u.v.A^{-1}}{1 + v.A^{-1}.u}.$$

Let J be a row vector. By applying the Sherman-Morrison formula we see that:

$$(\lambda I + J^T.J)^{-1} = \lambda^{-1} I - \lambda^{-2} \frac{1}{1 + \lambda^{-1} J.J^T} J^T.J = \lambda^{-1} \left(I - \frac{1}{\lambda + \|J\|^2} J^T.J \right)$$

and

$$\left(\lambda I + J^T.J\right)^{-1} J^T = \lambda^{-1} \left(1 - \frac{\|J\|^2}{\lambda + \|J\|^2} \right) J^T = \frac{1}{\|J\|^2 + \lambda} J^T.$$

Derivation of (4.4)

Let us look at the update displacement of a single pixel. To simplify the notations, we consider an update vector $\boldsymbol{u} = \boldsymbol{u}(p)$, an intensity difference $\delta = F(p) - M \circ c(p)$ and a Jacobian $J = J^p$. With the local estimation of the image noise $\sigma_i = |\delta|$, we have:

$$\boldsymbol{u} = \frac{\delta}{\|J\|^2 + \sigma_x^{-2}.\sigma_i^2} J^T, \qquad \|\boldsymbol{u}\| = \frac{\sigma_i.\|J\|}{\|J\|^2 + \sigma_x^{-2}.\sigma_i^2},$$

and since we have

$$\begin{aligned} &\left(\|J\| - \sigma_x^{-1}.\sigma_i \right)^2 \geq 0 \\ &\|J\|^2 + \sigma_x^{-2}.\sigma_i^2 - 2.\sigma_x^{-1}.\sigma_i.\|J\| \geq 0 \\ &2.\sigma_x^{-1}.\sigma_i.\|J\| \leq \|J\|^2 + \sigma_x^{-2}.\sigma_i^2 \end{aligned}$$

we find that

$$\|\boldsymbol{u}\| \leq \frac{\sigma_x}{2}$$

Chapter 5

Robust Mosaicing for Fibered Confocal Microscopy

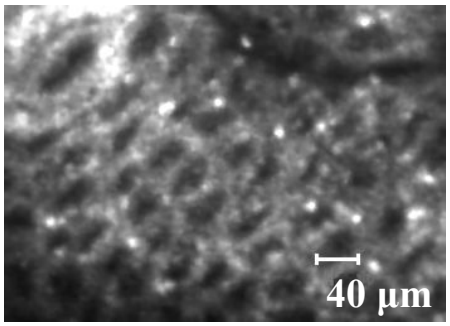
Table of Contents

5.1 Motivation: Improving FOV and Resolution with Image Registration

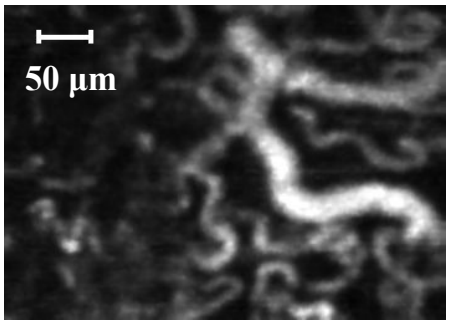
We have previously seen that fibered confocal microscopy (FCM), and especially Cellvizio, is a promising tool for *in vivo* histopathologic examination during an endoscopy or a biological experiment. Since FCM is a contact imaging modality, there is also an inevitable trade-off between resolution, field-of-view and invasiveness. In some cases, the small field-of-view limits the overall imaging possibilities and orientation *in vivo*. Therefore, even if, as illustrated in Fig. 5.1, high-resolution dynamic video sequences already provide interesting and important information, physicians and biologists also call for a complete and accurate representation of the entire region that has been imaged during a procedure.

To image and explore a region of interest, the confocal microprobe is often glided along the soft tissue. In this chapter, we show that video mosaicing techniques can help us move beyond the tradeoffs inherent to the design of the hardware by fusing the information contained in the video sequence into one wide field-of-view image. Several possible applications are targeted. First of all, the rendering of wide-field micro-architectural information on a single image will help experts to interpret the acquired data. This representation will also make quantitative and statistical analysis more accurate and robust. Moreover, mosaicing for microscopic images is a mean of filling the gap between microscopic and macroscopic scales. It allows multi-modality and multi-scale information fusion for the positioning of the optical microprobe.

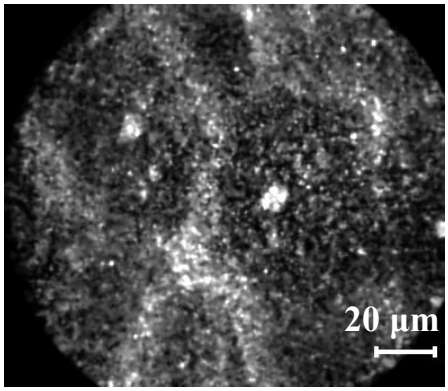
The displacement of the optical microprobe across the tissue can be described by a rigid motion. However, since FCM is a laser scanning device, an input frame does not represent a single point in time. Each sampling point corresponds to a different time instant. This induces motion artifacts when the optical microprobe moves with respect to the imaged tissue. Furthermore, the interaction of the contact optical microprobe with the soft tissue creates additional small non-rigid deformations. Due to these non-linear deformations, motion artifacts and irregular sampling of the input frames, classical video mosaicing techniques need to be adapted. Our approach, which was first presented in [Vercauteren 05, Vercauteren 06], uses the knowledge we have about FCM and the image reconstruction and registration building blocks of the previous chapters to create a robust mosaicing pipeline specific to laser-scanning devices. Our algorithm uses a hierarchical framework that is able to recover a globally consistent alignment of the input frames, to compensate for the motion-induced distortion of the input frames (simply called *motion distortion* hereafter) and to capture the non-rigid deformations. The global positioning is presented as an estimation problem on a Lie group [Vercauteren 05]. An efficient optimization scheme is proposed to solve this estimation problem. Because the motion distortions are induced by the motion of the optical microprobe, we model and use this relationship to recover the motion distortions. An efficient scattered data fitting method is also proposed to reconstruct on a regular grid the irregu-



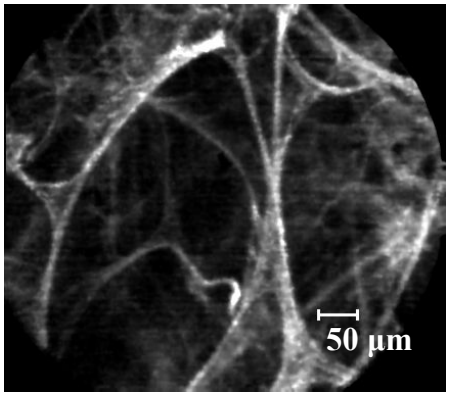
(a) *In vivo* mouse colon after instillation of acriflavine (Courtesy of D. Vignjevic, S. Robine, D. Louvard, Institut Curie, Paris, France).



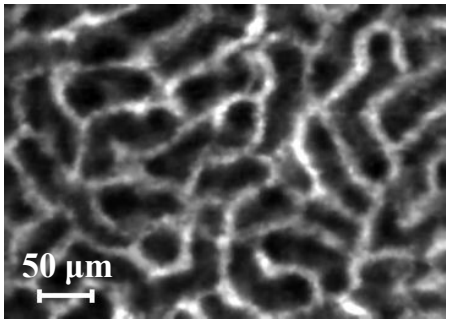
(b) *In vivo* tumoral angiogenesis in mouse with FITC-Dextran high MW (Courtesy of A. Duconseille and O. Clément, Descartes Image, Université Paris V, Paris, France).



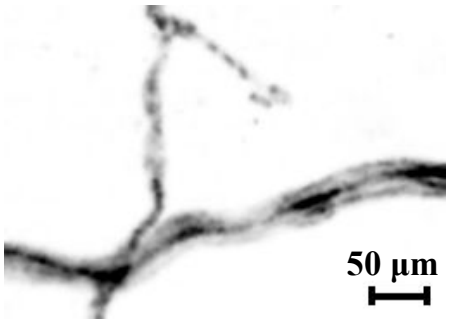
(c) *In vivo* reflectance imaging of human mouth mucosa.



(d) *Ex vivo* Autofluorescence imaging in human lung (Courtesy of Dr. P. Validire, Institut Mutualiste Monsouris, Paris, France).



(e) Microcirculation of the peritubular capillaries of a live mouse kidney with FITC-Dextran high MW.



(f) Dendritic receptors in a live Thy1-YFP mouse (Courtesy of I. Charvet, P. Meda, CMU, Geneva, Switzerland and L. Stoppini, Biocell Interface, Geneva, Switzerland).

Figure 5.1: Different types of images acquired with the Cellvizio.

larly sampled images that arise from the inputs and from the mosaic construction process. This reconstruction method is also used when we recover the non-rigid deformations with an adapted *demons* algorithm. The mosaicing algorithm we propose can generate still images that are comparable to standard histopathologic studies [Becker 07, Meining 07b]. This work also features an important quantitative validation of our mosaicing framework by using rigorous controlled experiments.

The remainder of the chapter is organized as follows. The main steps of our algorithm are described in Section 5.2. Section 5.3 provides a set of basic tools for Lie groups that will be used in Section 5.4 to get a set of globally consistent transformations from pairwise registration results. The motion distortions and non-rigid deformations compensation algorithms are presented in Section 5.5. An efficient scattered data fitting method is proposed in Section 5.6 to reconstruct the irregularly sampled images that arise from FCM and from the mosaic construction process. A controlled evaluation of our method and results on sequences acquired *in vivo* on both human and mouse tissue are presented in Section 5.7. Finally, Section 5.8 concludes the chapter.

Note that this chapter is, almost *in extenso*, the work we previously published in [Vercauteren 06]. There might therefore be some repeated information with respect to the previous chapters. But as redundancy offers opportunities to discern patterns in a variety of ways, this might help better understand the main ideas of this thesis.

5.2 Problem Statement and Overview of the Algorithm

The goal of many existing mosaicing algorithms is to estimate the reference-to-frame mappings and use these estimates to construct the mosaic [Irani 95]. Small residual misregistrations are then of little importance because the mosaic is reconstructed by segmenting the field into disjoint regions that use a single source image for the reconstruction [Davis 98, Levin 04, Peleg 00]. Even if these reconstruction techniques can ignore small local registration errors, a globally consistent alignment framework is needed to avoid large accumulated registration errors as shown in Fig. 5.2.

Since our input frames are rather noisy, we would like to use all the avail-

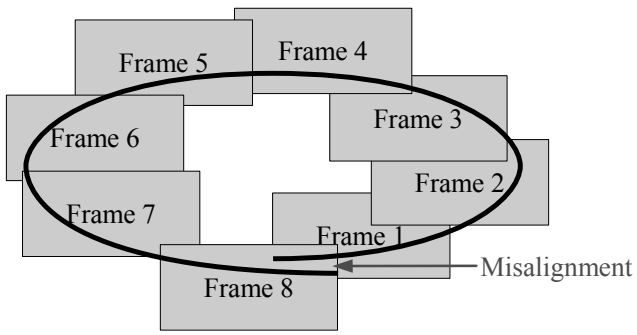


Figure 5.2: Accumulated registration errors result in global misalignment. What should be an ellipse, appears like an open curve due to these errors.

able information to recover an approximation of the true underlying scene. We will therefore estimate the frame-to-reference transformations (instead of the usual reference-to-frame) and consider all the input sampling points as shifted sampling points of the mosaic. This has several advantages for our problem. First of all, this is adapted to irregularly sampled input frames because we will always use the original sampling points and never interpolate the input data. This approach is also more consistent with a model of noise appearing on the observed frames rather than on the underlying truth. Finally, in this framework, it will be possible to get a mosaic at a higher resolution than the input frames. The challenge is that we need an accurate estimate of the unknown transformations.

5.2.1 Observation Model

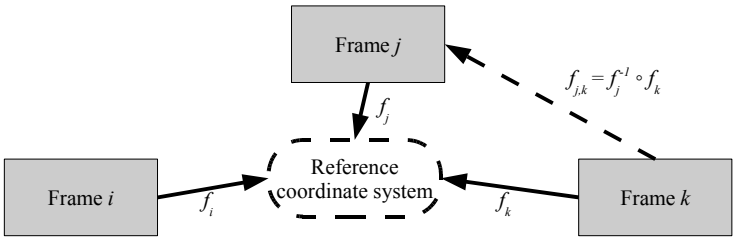


Figure 5.3: Overview of the global reference-to-frame transformations and the local frame-to-frame transformations.

Each frame of the input sequence is modeled as a noisy and deformed partial view of a ground truth 2D scene (the mosaic we want to recover). Let I be the unknown underlying truth and I_n be the observed frames. Each observed sampling point p in the coordinate system Ω_n associated with the n^{th} input frame can be mapped to a point in the reference coordinate system Ω by the n^{th} frame-to-reference mapping f_n. Each observed sampled value $I_n(p)$ can then be seen as a noisy observation of the ground truth signal $I(f_n(p))$:

$$I_n(p) = I(f_n(p)) + \varepsilon_n(p), \quad \forall p \in \Omega_n, \tag{5.1}$$

where $\varepsilon_n(p)$ is a noise term. Note that according to this observation model, we need to recover the frame-to-reference mappings as opposed to many existing mosaicing algorithms that estimate the reference-to-frame mappings.

We have specifically designed our transformation model for fibered confocal microscopy. The displacement of the optical microprobe across the tissue can be described by a rigid shift denoted by r_n. Since FCM is a laser scanning device, this motion of the optical microprobe with respect to the imaged tissue induces some motion artifacts. These distortions can be modeled by a linear transformation v_n that will be described in more details in Section 5.5. Finally, due to the interaction of the contact optical microprobe with the soft tissue, a small non-rigid deformation b_n appears. The frame-to-reference mapping are thus modeled by:

$$f_n(p) = b_n \circ r_n \circ v_n(p). \tag{5.2}$$

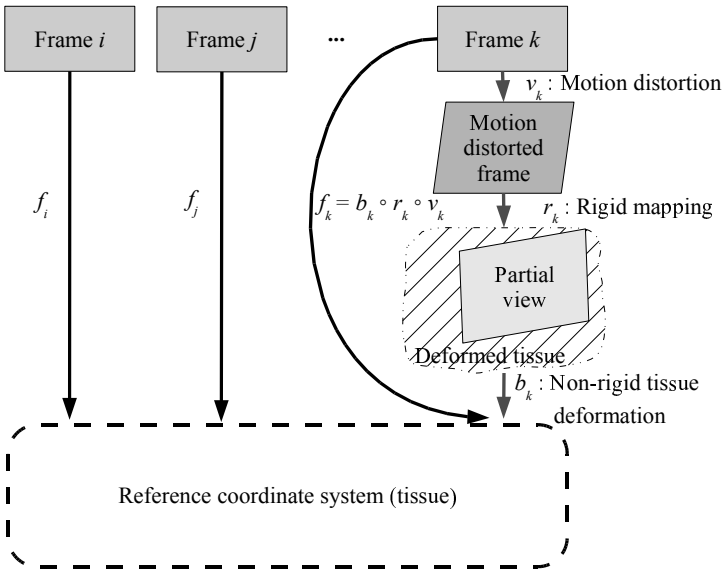


Figure 5.4: A schematic representation of the three-steps composing the global reference-to-frame transformations.

5.2.2 Overview of the Algorithm

As depicted in Fig. 5.4, our complete model of the frame-to-reference transformations is quite complex. A typical approach for dealing with the estimation of such complex models is to have a hierarchical, coarse-to-fine, approach. We will therefore focus on developing a method that iteratively refines the model while always keeping the global consistency of the estimated frame-to-reference transformations. The frame-to-reference mappings are composed of a motion-related distortion, a large rigid mapping and a small non-rigid deformation due to the soft tissue deformation.

We start by assuming that the motion distortions as well as the non-rigid tissue deformations can be ignored. By making the reasonable assumption that consecutive frames are overlapping, an initial estimate of the global rigid mappings can be obtained by using a rigid registration technique to estimate the motion between

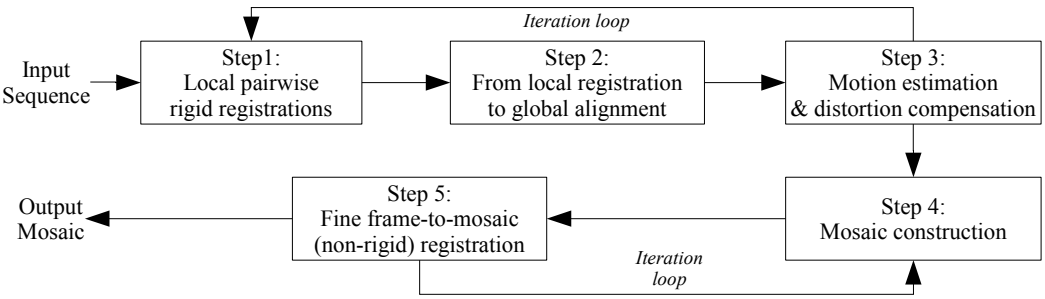


Figure 5.5: Block diagram of the mosaicing algorithm.

the consecutive frames. Global alignment is then obtained by composing the local motions. This initial estimate suffers from the well-known accumulation of error problem illustrated in Fig. 5.2.

The first loop of our algorithm (steps 1,2 and 3 in Fig. 5.5) alternates between three steps. The first step assumes that the motion distortions have been correctly estimated and registers pairs of distortion compensated frames under a rigid body transformation assumption. The second step uses these local pairwise registration results to make a globally consistent estimation of the rigid mappings r_n. The third step uses the relationship between the motion and the motion distortions to provide an updated and consistent set of rigid mappings and motion compensations.

Once a globally consistent set of transformations is found, the algorithm constructs a point cloud by mapping all observed sampling points onto a common reference coordinate system. An efficient scattered data fitting technique is then used on this point cloud to construct an initial mosaic. The residual non-rigid deformations are finally taken into account by iteratively registering an input frame to the mosaic and updating the mosaic based on the new estimate of the frame-to-mosaic mapping.

In step 2 of our algorithm, we use all available pairwise rigid registration results to estimate a set of globally consistent transformations. A sound choice is to consider a least-square approach. However, the space of rigid body transformations is not a vector space but rather a Lie group that can be considered as a Riemannian manifold. Classical notions using distances are therefore not trivial to generalize. In what follows, we provide a basic toolbox for estimation problems on Lie groups.

5.3 Basic Tools for Estimation Problems on Lie Groups

Many sets of primitives used in image processing and computer vision can be considered as real Lie Groups or as quotients of real Lie groups (e.g. 2D rigid body transformations, tensors [Fletcher 04a, Pennec 06c], quaternions, upper triangular matrices, M-reps [Fletcher 04b], vector spaces etc.). Most of them are not vector spaces and paradoxes such as Bertrand's paradox [Papoulis 02] appear when one considers a Lie Group as a vector space within an estimation problem.

The goal of this section is to provide a basic toolbox for estimation problems on real Lie groups. This synthesis uses classical tools from differential geometry and focuses on Lie groups to simplify the results and notations. We refer the reader to the standard textbooks for a detailed treatment of differential geometry (see e.g. [do Carmo 92]) and Lie groups (see e.g. [Helgason 01]. By using differential geometry, we will be able to generalize many algorithms designed for the usual vector space case.

A Lie group $\mathcal{G}$ is a smooth manifold together with a smooth composition map (usually denoted as multiplication) and a smooth inverse map, that satisfy the group axioms: closure, associativity, existence of a neutral element (denoted hereafter as Id) and existence of an inverse. Two important mappings for us are the left-

compositions and right-compositions by an element m:

$$\begin{aligned} L_m : &\qquad x \in \mathcal{G} \mapsto L_m(x) = m \circ x \in \mathcal{G} \\ R_m : &\qquad x \in \mathcal{G} \mapsto R_m(x) = x \circ m \in \mathcal{G}. \end{aligned}$$

They are diffeomorphisms by definition. Thus, they naturally induce the following differential maps (in a particular *local coordinate system* or *chart*) :

$$\begin{aligned} DL_m(x) : &\qquad \boldsymbol{u} \in T_x\mathcal{G} \mapsto DL_m(x) \cdot \boldsymbol{u} = \left.\frac{\partial m \circ y}{\partial y}\right|_{y=x} \cdot \boldsymbol{u} \quad \in T_{m\circ x}\mathcal{G} \\ DR_m(x) : &\qquad \boldsymbol{u} \in T_x\mathcal{G} \mapsto DR_m(x) \cdot \boldsymbol{u} = \left.\frac{\partial y \circ m}{\partial y}\right|_{y=x} \cdot \boldsymbol{u} \quad \in T_{x\circ m}\mathcal{G}, \end{aligned}$$

which allow us to map the space $T_x\mathcal{G}$ of tangent vectors to $\mathcal{G}$ at x to its counterpart $T_{m\circ x}\mathcal{G}$ or $T_{x\circ m}\mathcal{G}$. Similarly, we get for the inversion:

$$\begin{aligned} Inv : &\qquad x \in \mathcal{G} \mapsto Inv(x) = x^{-1} \in \mathcal{G} \\ DInv(x) : &\qquad \boldsymbol{u} \in T_x\mathcal{G} \mapsto DInv(x) \cdot \boldsymbol{u} = \left.\frac{\partial y^{-1}}{\partial y}\right|_{y=x} \cdot \boldsymbol{u} \quad \in T_{x^{-1}}\mathcal{G}. \end{aligned}$$

5.3.1 Left Invariant Metric and Distance

Because many estimation problem involve a measure of discrepancy between two elements, it is natural to look for a definition of a distance between two elements of a Lie group. This can be done by providing the Lie group with a Riemannian metric which is a continuous collection of dot products on the tangent space $T_x(\mathcal{G})$ at x: $\langle \boldsymbol{v} \,|\, \boldsymbol{w} \rangle_x = \boldsymbol{v}^T.G(x).\boldsymbol{w}$. Because of the smoothness of the Lie Group, we can smoothly translate a dot product at the Id-tangent space to any other tangent space by left or right composition thanks to the differential maps above. In the remainder of this chapter, we focus on left invariant metrics. Thanks to the left-composition differential map DL_x, they can be represented by the matrix field,

$$G(x) = DL_x(\mathrm{Id})^{-T}.G(\mathrm{Id}).DL_x(\mathrm{Id})^{-1}. \tag{5.3}$$

The Riemannian metric provides the intrinsic way of measuring the length of any smooth curve on the Lie group. The distance between two points of a Lie Group is then the minimum length among the curves joining these points. The curves realizing this minimum for any two points of the manifold are called geodesics. The calculus of variation shows that the geodesics follow a second order differential system depending on the Riemannian metric. Therefore, we know that there exists one and only one geodesic $\gamma_{(x,\boldsymbol{u})}(\cdot)$ defined for all times (thanks to the left-invariance, see e.g. [do Carmo 92]), going through x at time $t = 0$ and having $\boldsymbol{u}$ as a tangent vector.

5.3.2 Riemannian Exponential and Logarithm Maps

A key notion of differential geometry is the Riemannian exponential map. Let us consider a given point x of the Lie group, a vector $\boldsymbol{u}$ in the corresponding tangent space and the uniquely associated geodesic $\gamma_{(x,\boldsymbol{u})}$. The exponential map is the function that maps $\boldsymbol{u}$ to the point $\gamma_{(x,\boldsymbol{u})}(1)$ reached after a unit time by the geodesic:

$$\exp_x : \left\{ \begin{array}{rcl} T_x(\mathcal{G}) & \rightarrow & \mathcal{G} \\ \boldsymbol{u} & \mapsto & \exp_x(\boldsymbol{u}) = \gamma_{(x,\boldsymbol{u})}(1). \end{array} \right.$$

This function realizes a local diffeomorphism from a neighborhood of 0 to a neighborhood of x by developing the tangent space along the geodesics. Within this neighborhood, the inverse of the exponential map exists and is called the log map $\log_x(\cdot)$:

$$\exp_x \boldsymbol{u} = y \quad \Leftrightarrow \quad \boldsymbol{u} = \log_x y.$$

In the context of Lie groups, the exponential map notion is somewhat ambiguous as one can use the Lie group exponential (as we did in the previous chapters) or the Riemannian manifold one. Both definitions agree if and only if the Lie group admits a left-and-right-invariant Riemannian metric (such as for compact Lie groups) used to define the geodesics. Unfortunately, for the group of rigid body transformations, it can be shown that no bi-invariant metric exists [Arsigny 06b]. In the context of estimation problems, we are mainly interested in distance measurements and we will therefore stick to the Riemannian exponential in this chapter. *Note that the notion of exponential we use in this chapter is thus different from the one used in the previous ones.*

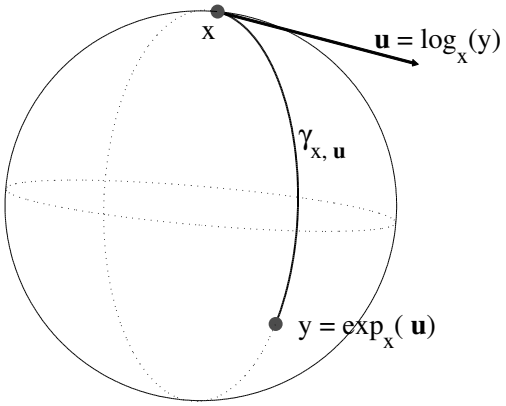


Figure 5.6: Riemannian exponential and log maps on a unit sphere.

When a left-invariant metric is used, the exponential and log maps at any point of a Lie group can be related through left composition to their counterpart at the

identity with the following equations:

$$\log_x(y) = DL_x(\mathrm{Id}) \log_{\mathrm{Id}}(x^{-1} \circ y) \tag{5.4}$$
$$\exp_x(\boldsymbol{u}) = x \circ \exp_{\mathrm{Id}}(DL_x(\mathrm{Id})^{-1}\boldsymbol{u}). \tag{5.5}$$

With the log map, the (geodesic) distance between two points is given by:

$$\begin{aligned} \mathrm{dist}\,(x,y) &= \mathrm{dist}\left(\mathrm{Id}, x^{-1} \circ y\right) = \left\|\log_{\mathrm{Id}}(x^{-1} \circ y)\right\| \\ &= \langle \log_{\mathrm{Id}}(x^{-1} \circ y) \mid \log_{\mathrm{Id}}(x^{-1} \circ y)\rangle^{1/2}. \end{aligned} \tag{5.6}$$

As shown above, we see that even if we mainly use the Lie group as a Riemannian manifold, its additional structure is of great practical interest because it allows us to map every tangent space to the one at the identity. We therefore only need to have computational tools for one tangent space, the tangent space at the identity.

5.3.3 Mean and Covariance Matrix

Lie groups are usually not vector spaces, and the notion of expectation cannot readily be extended to it. However, for any metric space, probabilistic spaces, random elements and probability density functions can be defined [Papoulis 02]. Expectations and other usual tools are then defined for random variables, which are real-valued functions of the probabilistic events, but not directly for random elements of the group [Pennec 98]. However, by changing the definition of the expectation and using the Riemannian geometry tools presented in Section 5.3.2, it turns out that a consistent statistical framework (including the mean, the covariance and the Mahalanobis distance) can be defined [Pennec 06a].

In a vector space, the mean can be seen as the element that minimizes the expected distance to a random vector. This point of view allows us to generalize the mean for Lie groups. Let x be a random element and let

$$\sigma_x^2(y) = \mathbb{E}\left[\mathrm{dist}\,(y,x)^2\right]$$

be its variance at the (fixed) element y. Note that this is well defined because $\mathrm{dist}\,(y,\cdot)$ is a real-valued function.

Let x be a random element of a Lie group $\mathcal{G}$. If the variance $\sigma_x^2(y)$ is finite for one element $y \in \mathcal{G}$, thanks to the triangular inequality, it is finite for all elements $y \in \mathcal{G}$ and every element minimizing the variance is called a Fréchet mean element. The set of Fréchet mean elements is thus given by

$$\mathcal{F}_x = \arg\min_{y \in \mathcal{G}} \left(\mathbb{E}\left[\mathrm{dist}\,(y,x)^2\right]\right). \tag{5.7}$$

It can be shown that under suitable conditions (that are assumed to be fulfilled here), there exists one and only one Fréchet mean which we denote as $\mathbb{E}[x]$.

In [Pennec 06a], an algorithm is provided to compute the (empirical) Fréchet mean of a weighted set of samples or the Fréchet mean of a known distribution (the expectations hereunder are used both for empirical and true distributions).

Algorithm 9 (Fréchet Mean Computation)

- Let $\hat{x}$ be an initial estimate of $\bar{x}$, e.g. any value of the set of samples in the discrete case
- Iterate until some convergence criterion is met, e.g. $\mathbb{E}\left[\mathrm{dist}\left(\hat{x}_{\mathrm{new}}, \hat{x}_{\mathrm{old}}\right)^2\right] < \epsilon$:
  - $\hat{x} \leftarrow \exp_{\hat{x}}\left(\mathbb{E}\left[\log_{\hat{x}}(x)\right]\right) = \hat{x} \circ \exp_{\mathrm{Id}}(\mathbb{E}[\log_{\mathrm{Id}}(\hat{x}^{-1} \circ x)])$ ∎

This algorithm can either be seen as an approximate Newton's method with unitary step length that is used to minimize the variance or as a fixed-point iteration. Fréchet means are indeed fixed-points of $m \mapsto \exp_m\left(\mathbb{E}\left[\log_m(x)\right]\right)$.

To define higher moments of a distribution, on a Lie group, the exponential map at the mean point is used. The random feature is thus represented as a random vector with zero mean in a star-shaped domain. With this representation, the covariance matrix can be defined by:

$$\begin{aligned}\Sigma_{xx} &= \mathbb{E}\left[\log_{\mathbb{E}[x]}(x).\log_{\mathbb{E}[x]}(x)^T\right] \\ &= DL_m(\mathbb{E}[x]).\,\mathbb{E}\left[\log_{\mathrm{Id}}\left(\mathbb{E}[x]^{-1} \circ x\right).\log_{\mathrm{Id}}\left(\mathbb{E}[x]^{-1} \circ x\right)^T\right].DL_m(\mathbb{E}[x])^T.\end{aligned} \tag{5.8}$$

Last but not least, the Mahalanobis distance, which plays a key role to get a robust estimation of the global positioning, can be defined. The Mahalanobis distance is a classical tool to define a statistical distance in a sample space with known covariance matrix. Its definition can easily be extended to Lie groups by using the previous definition of the covariance matrix. The Mahalanobis distance of a point y to a random feature with Fréchet mean $\mathbb{E}[x]$ and covariance matrix Σ_{xx} is given by

$$\mu_x^2(y) = \log_{\mathbb{E}[x]}(y)^T \Sigma_{xx}^{-1} \log_{\mathbb{E}[x]}(y). \tag{5.9}$$

5.4 From Local to Global Alignment

Now that all the necessary tools have been presented in Section 5.3, we will show how the problem of global positioning can be cast to an estimation problems on a Lie group. The first step of our algorithm is to find a globally consistent set of transformations to map the input frames to a common coordinate system. When the input frames arise from a single gliding of the flexible microprobe along a straight line, it may be possible to generate decent alignments by computing only pairwise registrations between the consecutive frames. However, in the general case, all spatial neighbors frames might not be temporal neighbors, and accumulated errors can lead to a poor global alignment as shown in Fig. 5.2. Methods for producing globally consistent alignments are therefore needed. The automatic multi-image alignment algorithms developed so far broadly falls into two categories: feature based [Brown 03, Can 04] and local to global methods [Davis 98, Sawhney 98]. Feature-based methods obviously rely on features extraction from the input frames. To get meaningful extracted information, the chosen features are often tailored to the

types of images we use, see e.g. [Chui 03], and have been developed for classical uniformly sampled images. We want an algorithm that is able to cope with non-uniformly sampled inputs (possibly very noisy) arising from many different types of tissue, as shown in Fig. 5.1. We therefore chose to develop a method using local pairwise alignments to generate consistent global mappings.

5.4.1 Framework for Global Positioning

Given two input frames, we need to estimate the (pairwise) frame-to-frame transformation. At this step of the algorithm, we assume that the non-rigid tissue deformations can be ignored and that the motion distortions have been correctly estimated. With these assumptions, we need to perform rigid registrations of distortion corrected input frames. For that purpose, we use a classical registration framework based on a similarity criterion optimization but any other technique (e.g. block matching framework [Ourselin 00], Mellin transform [Davis 98] etc.) can be used. Let $r_{j,i}^{(\text{obs})}$ be the pairwise rigid registration result between the distortion corrected input frames i and j. This result is considered as a noisy observation of $r_j^{-1} \circ r_i$. Based on the set of all available observations, our algorithm looks for a globally consistent estimate of the global parameters $[r_1, \ldots, r_N]$. This problem is addressed in [Davis 98] where a least squares solution on spatial transformation matrices is given when linear transformations are considered. This technique cannot readily be adapted to rigid body transformations. In [Sawhney 98], the authors propose a more general approach. Some chosen corner points are transformed through r_i and $r_j \circ r_{j,i}^{(\text{obs})}$. The squared distance between the transformed points added to a regularization term is then minimized. These techniques are sensitive to outliers, and are either tailored to a specific type of transformation or need a somewhat *ad hoc* choice of points. In this work, we chose to rely on the tools presented in Section 5.3 to provide consistent and robust estimates of the global rigid body transformations. Practical instantiations of the generic tools presented in Section 5.3 are given, for 2D rigid body transformations, in the appendix page 110.

The computational cost of registering all input frames pairs is prohibitive and not all pairs of input frames are overlapping. It is therefore necessary to choose which pairs could provide informative registration results. For that purpose, we chose the topology refinement approach proposed in [Sawhney 98]. An initial guess of the global parameters $[r_1, \ldots, r_N]$ is obtained by registering the consecutive frames, the algorithm then iteratively chooses a next pair of input frames to register (thus providing a new observation $r_{j,i}^{(\text{obs})}$) and updates the global parameters estimation. As we only consider the pairwise registration results as noisy observations, we need many of them. To minimize the computational cost of those numerous registrations, we use a multi-resolution registration technique using a Gaussian image pyramid that stops at a coarse level of the optimization.

5.4.2 A Lie Group Approach for Global Positioning

Let e be a random error whose Fréchet mean is the identity and whose covariance matrix is Σ_{ee}. The observation model is given by

$$r_{j,i}^{(\text{obs})} = r_j^{-1} \circ r_i \circ e_{j,i}^{(\text{obs})}, \tag{5.10}$$

where $e_{j,i}^{(\text{obs})}$ is a realization of the random error e.

Given the set of all available pairwise rigid registration results Θ, we need to estimate the true transformations. A natural choice is to minimize the statistical distance, i.e., Mahalanobis distance, between the observations $r_{j,i}^{(\text{obs})}$ and the transformations $r_j^{-1} \circ r_i$ predicted by our model. Our global criterion is thus given by:

$$[r_1^*, \ldots, r_N^*] = \arg \min_{[r_1,\ldots,r_N]} \frac{1}{2} \sum_{(i,j)\in\Theta} \mu_e^2(e_{j,i}^{(\text{obs})}). \tag{5.11}$$

It can be seen that this criterion is insensitive to a composition of all transformations with a fixed arbitrary transformation. We can therefore choose any transformation in $[r_1, \ldots, r_N]$ to be the identity transformation Id. Without loss of generality we can look for any minimizer in (5.11) and then compose all estimates with a common rigid body transformation so that we get for example $r_{\lceil \frac{N}{2} \rceil} = \text{Id}$. The covariance matrix used for the Mahalanobis metric depends on the application. We typically start with a diagonal matrix that weights the angles with respect to the translation. This matrix can further be estimated from the data as explained in Section 5.4.3.

According to (5.9), each term of the sum in (5.11) is given by

$$\begin{aligned}\mu_e^2(e_{j,i}^{(\text{obs})}) &= \log_{\text{Id}}(e_{j,i}^{(\text{obs})})^T \Sigma_{ee}^{-1} \log_{\text{Id}}(e_{j,i}^{(\text{obs})}) \\ &= \left(S_{ee}.\log_{\text{Id}}(r_i^{-1} \circ r_j \circ r_{j,i}^{(\text{obs})})\right)^T . \left(S_{ee}.\log_{\text{Id}}(r_i^{-1} \circ r_j \circ r_{j,i}^{(\text{obs})})\right),\end{aligned}$$

where S_{ee} is a matrix square root (e.g. Cholesky factorization) of Σ_{ee}^{-1}. We see that we need to solve the non-linear least squares problem:

$$[r_1^*, \ldots, r_N^*] = \arg \min_{[r_1,\ldots,r_N]} \frac{1}{2} \left\|\phi(r_1, \ldots, r_N)\right\|^2,$$

where $\phi(r_1, \ldots, r_N) = \text{Vect}\left(\left\{S_{ee}.\log_{\text{Id}}(r_i^{-1} \circ r_j \circ r_{j,i}^{(\text{obs})})\right\}_{(i,j)\in\Theta}\right)$. Note that this criterion is indeed non-linear because of the composition involved in the computation of $e_{j,i}^{(\text{obs})}$.

5.4.3 Riemannian Method for Non-linear Least Squares

We have seen in Chapter 3, some well-posed methods for non-linear least-squares problems on Lie groups. The main idea in this setting was to use the Lie group exponential to make a series of intrinsic update steps towards the optimum. Obviously these techniques can readily be applied for the optimization of 5.11. It

might however be somewhat puzzling to the reader to see that our criterion uses the Riemannian exponential whereas our optimization routine uses the Lie group exponential (we recall that these two notion are different in the context of 2D rigid-body transformations). The goal of this section is thus to rephrase the optimization routine from a Riemannian point of view, i.e. using the Riemannian exponential.

In the Riemannian framework, the straightest paths are given by the geodesics. The idea is thus to walk towards an optimum by a series of steps taken along a geodesic of the manifold rather than walking in the tangent vector space. Thanks to the Riemannian exponential map, we have a canonical way to follow a geodesic starting from a given point and having a given initial tangent vector. It is thus possible to combine the power of intrinsic geodesic walking and the ease of use of classical optimization routines in a natural way.

The Gauss-Newton method forms the basis for the efficient methods that have been developed for non-linear least squares optimization. We now show how it can be used in the Riemannian setting. Let x be an element of a Lie group $\mathcal{G}$ and let $\frac{1}{2}\|\phi(x)\|^2$ be the function we want to minimize. The Gauss-Newton method is based on a linear approximation of $\phi(.)$ in a neighborhood of the current estimate. We denote by $\boldsymbol{\phi}_x(.) = \phi(\exp_x(.))$ the expression of $\phi(.)$ in a normal coordinate system at x. Its Taylor expansion around the origin of this chart is given by:

$$\boldsymbol{\phi}_x(\boldsymbol{v}) = \boldsymbol{\phi}_x(0) + J_\phi(x).\boldsymbol{v} + O(\|\boldsymbol{v}\|^2),$$

where $J_\phi(x) = \left.\frac{\partial\phi(y)}{\partial y}\right|_{y=x}$. By keeping only the linear part we get the following approximation:

$$\frac{1}{2}\|\boldsymbol{\phi}_x(\boldsymbol{v})\|^2 \simeq \frac{1}{2}\|\phi(x)\|^2 + \boldsymbol{v}^T.J_\phi(x)^T.\phi(x) + \frac{1}{2}\boldsymbol{v}^T.J_\phi(x)^T.J_\phi(x).\boldsymbol{v}.$$

The Gauss-Newton step minimizes this approximation:

$$\boldsymbol{v}_{gn} = \arg\min_{\boldsymbol{v}}\left[\boldsymbol{v}^T.J_\phi(x)^T.\phi(x) + \frac{1}{2}\boldsymbol{v}^T.J_\phi(x)^T.J_\phi(x).\boldsymbol{v}\right].$$

It is well known that if $J_\phi(x)$ has full rank, this has a unique minimizer which is the solution of $J_\phi(x)^T.J_\phi(x).\boldsymbol{v} = -J_\phi(x)^T.\phi(x)$.

Our optimization routine now follows the geodesic starting from the current estimate $x^{(old)}$ and whose tangent vector is $\boldsymbol{v}_{gn}$. We thus get the following update equation:

$$x^{(new)} = \exp_{x^{(old)}}(\lambda\boldsymbol{v}_{gn}) = x^{(old)} \circ \exp_{\mathrm{Id}}(\lambda DL^{-1}_{x^{(old)}}(\mathrm{Id}).\boldsymbol{v}_{gn}), \tag{5.12}$$

The classical Gauss-Newton routine uses $\lambda = 1$ at all steps, but a line search could also be used.

We have shown how to adapt the Gauss-Newton procedure for a non-linear least squares problem on a Lie group. A very similar approach would also provide extensions of other classical non-linear least squares optimizers (such as the Levenberg-Marquardt method or Powell's dog leg method) for Lie groups.

This method is applied to solve (5.11), where the Lie group we use is the Cartesian product of N rigid body transformation groups. The Jacobian $J_\phi([r_1, \ldots, r_N])$ can easily be computed by seeing that:

$$\frac{\partial \log_{\mathrm{Id}}(r_i^{-1} \circ r_j \circ r_{j,i}^{(\mathrm{obs})})}{\partial r_i} = DR_{r_j \circ r_{j,i}^{(\mathrm{obs})}}(r_i^{-1}).DInv(r_i)$$
$$\frac{\partial \log_{\mathrm{Id}}(r_i^{-1} \circ r_j \circ r_{j,i}^{(\mathrm{obs})})}{\partial r_j} = DR_{r_{j,i}^{(\mathrm{obs})}}(r_i^{-1} \circ r_j).DL_{r_i^{-1}}(r_j),$$

where $DInv(r) = \left.\frac{\partial s^{-1}}{\partial s}\right|_{s=r}$.

Within this general framework, several improvements can easily be added. To perform the rigid registration between the distortion corrected input frames, we chose to optimize the squared correlation coefficient. It is then straightforward to weight the terms of the cost function (5.11) by this confidence measure. A well known problem of pure least squares approach is the sensitivity of the solution to outliers in the observations. Many solutions have been proposed to address the presence of outliers [Rousseeuw 87]. The most common ones rely on using only a subset of the observations (e.g. least trimmed squares, reweighted least squares) or on a minimization of the sum of a function of the residuals (e.g. M-estimators). To be able to use the efficient least-squares optimizer presented above, the easiest is to use the first approach or to solve the M-estimator problem using iteratively reweighted least squares.

In our particular setting we chose the simple reweighted least squares approach:

$$[r_1^*, \ldots, r_N^*] = \arg \min_{[r_1,\ldots,r_N]} \frac{1}{2} \sum_{(i,j)\in\Theta} w_{j,i}.\rho_{j,i}.\mu_e^2(e_{j,i}^{(\mathrm{obs})})$$
$$w_{j,i} = \begin{cases} 1 & \text{if } \mu_e^2(.) \leq \gamma \\ 0 & \text{otherwise} \end{cases},$$

where $\rho_{j,i}$ is the correlation coefficient between the registered distortion corrected frames i and j, and γ is the 95% quantile of the χ^2 distribution with 3 degrees of freedom. If enough observations are available, our procedure also includes an estimation of the covariance matrix Σ_{ee}.

5.5 Compensating for the Frame Distortions

An interesting point of scanning imaging devices is that the output image is not a representation of a given instant, but a juxtaposition of points acquired at different times [Savoire 04]. Consequently, if the flexible microprobe moves with respect to the imaged tissue, what we observe is not a frozen picture of the tissue, but a skewed image of it. Each scan line indeed relates to a different instant, and the flexible microprobe moved between each scan line.

5.5.1 Influence of Relative Motion

The scanning of the laser can be decomposed into a fast horizontal sinusoidal component and a slow linear uniform vertical component. Horizontally, the imaging is done only on the central part of the trajectory, where the spot velocity is maximal and nearly constant. Since in this part, the spot horizontal velocity V_x (> 5 m/s) is several orders of magnitude higher than the spot vertical velocity V_y (~ 2 mm/s), we assume that the horizontal spot velocity is infinite.

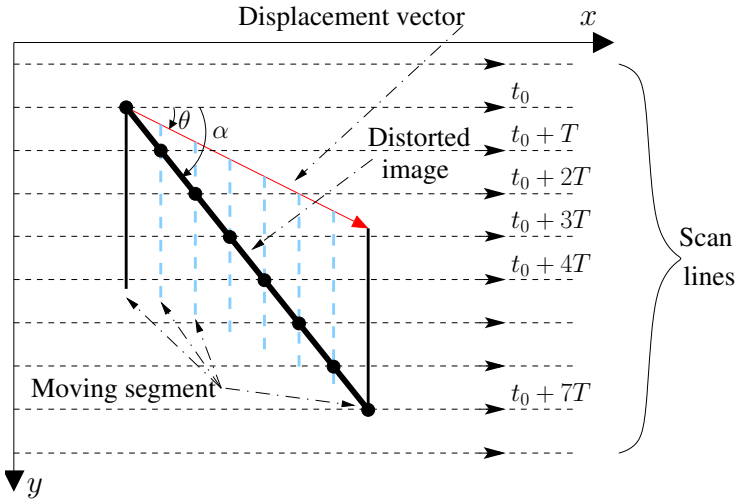


Figure 5.7: Imaging of a moving vertical segment by a scanning laser. The segment has a translation movement from the upper left corner to the lower right corner of the image. The segment is first intersected by a scan line at instant t_0 (black disks represent imaged points). The following scan lines image the segment at different positions (dotted segments). The resulting shape is the slanting segment of angle α.

Let us consider a standard 2D+t volume, without scanning each acquired frame would correspond to a single time instant t_0 and thus to a 2D horizontal slice of the volume. With scanning a point of ordinate y corresponds to the time $t_0 + \frac{y}{V_y}$. The process of image formation therefore comes down to imaging an oblique plane of the volume. Figure 5.7 presents what will be observed when imaging a vertical segment moving with respect to the flexible microprobe.

5.5.2 Motion Distortions Model

Without any further assumption, these motion distortions can have a very general form. This can for example be seen in some famous photographs by Henri Lartigue or Robert Doisneau shot with a slit-scan camera. In our particular case, the main relative motion is due to the gliding of the flexible microprobe along the tissue and some residual movements can be produced by the deformation of the soft tissue.

We again use a hierarchical approach and ignore the effect of the tissue deformation. We thus face the problem of recovering the gliding motion of the flexible

microprobe. This motion will typically be smooth and will mainly be composed of a translation part because of an important torsion resistance of the flexible microprobe. Note that even if consecutive frames can only be slightly rotated, more time distant frames can have a large rotational difference. To be able to robustly recover the gliding motion, we need to further constrain it. We will therefore assume that during the time period T_{scan} taken by the laser to scan a complete input frame, the flexible microprobe is only animated with a translational motion with constant velocity vector $\tilde{\eta} = [\tilde{\eta}^x, \tilde{\eta}^y]$.

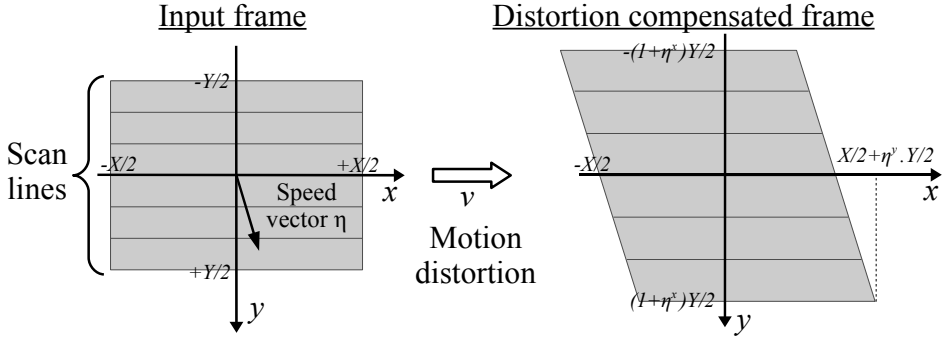


Figure 5.8: Schema of the motion distortion for a constant translational velocity.

Let $p = [x, y]$ be a point in the coordinate system related to the input frame. As shown in Section 5.5.1, each laser scan line, indexed by the ordinate y of a point, corresponds to a different time instant $t(y) = t(0) + \frac{y}{V_y}$, where we recall that V_y is the vertical velocity of the laser scanning process. The point p will therefore be shifted by $(t(y) - t(0)) . \tilde{\eta} = \frac{y}{V_y}\tilde{\eta}$ with respect to the center of the coordinate system. To simplify the notations, we will from now on use the normalized velocity $\eta = \frac{1}{V_y}\tilde{\eta}$. We now have a way to map a point p of an input frame to a point p_d in the corresponding distortion compensated frame by using the following linear transformation:

$$p_d = \begin{bmatrix} 1 & \eta^x \\ 0 & 1 + \eta^y \end{bmatrix} p = M(\eta).p. \tag{5.13}$$

This is the explicit expression for the motion distortion transformations, i.e., $v_k(p) = M(\eta_k).p$ for a given frame k.

From this distortion model, we can derive the transformation model mapping an input frame k to another frame j. With the assumptions that the non-rigid deformations can be ignored, we have $f_j = r_j \circ v_j$, $f_k = r_k \circ v_k$ and $f_{j,k} = f_j^{-1} \circ f_k$, where we recall that f_n denotes a global frame-to-reference mapping. This implies that:

$$f_{j,k} = v_j^{-1} \circ r_j^{-1} \circ r_k \circ v_k = v_j^{-1} \circ r_{j,k} \circ v_k. \tag{5.14}$$

In the local-to-global positioning scheme presented in Section 5.4, we mentioned that we needed to perform rigid registrations of distortion compensated frames. By using (5.14), we see that these registrations can be performed without the need to

explicitly create the motion compensated frames. We simply look for the best rigid body transformation $r_{j,k}$ while keeping v_k and v_j fixed.

5.5.3 Velocity Computation

In order to recover the motion distortions, one could try to register the frames using the complete transformation model (5.14). However, this would imply to ignore the relationship between positions and velocity and would thus not be robust. We therefore chose to compute the velocities using the displacements information only.

Using finite difference equations, we can relate the global positioning and the velocity η. Since our motion model ignores the rotational velocity, we can focus on the displacement of the center of the flexible microprobe. Let $r = [\theta, \tau]$ be the rigid mapping between two consecutive distortion compensated input frames. The center $[0, 0]$ of the first distortion compensated frame is mapped through r to the point τ in the second motion distortion compensated frame. The displacement of the center of the flexible microprobe during the corresponding inter-frame time period T_{frame} is thus given by τ. By using a forward difference equation to relate the speed vector and the displacement vector, we get:

$$\eta = \frac{1}{V_y} . \frac{1}{T_{frame}} . \tau. \tag{5.15}$$

Now let us assume that, with the current estimates of the rigid mapping $r^{(old)}$ and velocity $\eta^{(old)}$, the two consecutive input frames are correctly aligned. By using (5.14), we see that the center $[0, 0]$ of the first (uncompensated) input frame is mapped to the point $q = M\left(\eta^{(old)}\right)^{-1} . \tau^{(old)}$ in the (uncompensated) second input frame. Let $r^{(new)}$ and $\eta^{(new)}$ be the updated rigid mapping and velocity. To keep a correct alignment, the center of the first (uncompensated) input frame should still be mapped to the point q in the (uncompensated) second input frame. We therefore need to have:

$$M\left(\eta^{(new)}\right)^{-1} . \tau^{(new)} = M\left(\eta^{(old)}\right)^{-1} . \tau^{(old)}.$$

On the other hand, according to (5.15), we should have

$$\eta^{(new)} = \frac{1}{V_y} . \frac{1}{T_{frame}} . \tau^{(new)}.$$

This system is solved to get updated estimations. A similar procedure is used with the backward difference equation and the different estimations are averaged to get the final update.

As shown in Fig. 5.5, our algorithm iterates between a global positioning scheme during which the velocities are assumed to be correctly estimated and the motion distortion compensation routine. The final step should be a global positioning one because a typical user will be sensitive to some slight misalignment but not to a slight inconsistency in the motion distortion.

5.5.4 Soft Tissue Deformations

Once a globally consistent, motion compensated, mosaic has been constructed, a refined multi-image registration can be performed by iteratively registering each input frame to the mosaic and updating the mosaic. The mosaicing problem can be written as an optimization problem over the unknown underlying image I and the unknown transformations $[f_1, \ldots, f_N]$ of the following multi-image criterion, $\boldsymbol{S}(f_1, \ldots, f_N, I) = \sum_{n=1}^{N} S(I_n, I \circ f_n)$, where $S(I_a, I_b)$ is a usual pairwise similarity criterion between the two images I_a and I_b. With this framework our mosaic refinement procedure can be seen as an alternate optimization scheme.

We again use a hierarchical approach and divide the fine frame-to-mosaic registrations into two loops of increasing model complexity. First we refine the global linear mappings. Then, to compensate for the small non-rigid deformations, we have adapted the demons algorithm [Thirion 98] to our special case of irregularly sampled input frames. The general demons scheme is modified as follows:

- A fine reference grid is used to make a sparse grid Γ from the scattered point set representing the centers of the fibers composing the flexible microprobe. All pixels of Γ are selected in the input frame k to be *demons*.
- The non-rigid deformations b_k is modeled by a list of elementary displacements (one per fiber or demon).
- To get a regular displacement *field*, the sparse displacement field is smoothed at each iteration by using the scattered data approximation method that will be presented in Section 5.6 (with a large smoothing factor).
- The optical flow is computed for all *demons*. Its computation requires the gradient of the reference image which in our case is a non-uniformly sampled input frame. To get an approximation of this gradient, we reconstruct the non-uniformly sampled input frame on a regular grid using the scattered data approximation method that will be presented in Section 5.6. We then use the gradient of this reconstruction.

We mainly used this scheme because of its efficiency and its adaptation to our particular data, but other schemes could also be used. The residual deformation fields $[b_1, \ldots, b_N]$ could be modeled by using B-splines tensor products on a predefined grid such as in [Rueckert 99]. The framework can also easily be extended to use any other non-rigid registration methods using landmarks-based schemes or more accurate deformation models such as in [Cachier 03].

5.6 Efficient Scattered Data Approximation

The iterative mosaic refinement scheme presented in Section 5.5.4 requires a new mosaic construction at each iteration. Furthermore, the adapted *demons* algorithm needs a method for smoothing deformation fields that are defined on a sparse grid,

together with a method being able to construct a regularly sampled image from an irregularly sampled input.

These goals can be achieved with a single method for scattered data approximation provided that it allows us to control the degree of smoothness of the reconstruction. Since we want to register the input frames with the mosaic, controlling the smoothness of the reconstruction is also important. We indeed need a mosaic that is smooth enough for the registration process not to be trapped in a local minimum but detailed enough for the registration result to be accurate.

As appears above, the scattered data approximation method will be used many times. It is therefore necessary to use a very efficient algorithm.

The usual algorithms for scattered data interpolation or approximation, such as triangulation based methods, Kriging methods, radial basis functions interpolations, B-Spline approximations or moving least squares methods (see e.g. [Amidror 02, Lee 97, Lodha 99] and references therein) do not simultaneously meet the requirements of efficiency, and control over the smoothness of the approximation. In the next section, we develop a main contribution which is an efficient scattered data fitting algorithm that allows a control over the smoothness of the reconstruction.

5.6.1 Discrete Shepard's Like Method

Let $\{(p_k, i_k) \in \Omega \times \mathbb{R}\}$ be the set of sampling points and their associated signal. Our goal is to get an approximation of the underlying function on a regular grid Γ defined in Ω. The main idea is to use a method close to Shepard's interpolation. The value associated with a point in Γ is a weighted average of the nearby sampled values,

$$\hat{I}(p) = \sum_k w_k(p) i_k = \sum_k \frac{h_k(p)}{\sum_l h_l(p)} i_k. \tag{5.16}$$

The usual choice is to take weights that are the inverse of the distance, $h_k(p) = \mathrm{dist}(p, p_k)^{-1}$. In such a case we get a true interpolation [Amidror 02]. An approximation is obtained if a bounded weighting function $h_k(p)$ is chosen. We choose a Gaussian weight $h_k(p) = G(p - p_k) \propto \exp(-||p - p_k||^2/2\sigma_a^2)$ and (5.16) can thus be rewritten as

$$\hat{I}(p) = \frac{\sum_k i_k G(p - p_k)}{\sum_k G(p - p_k)} = \frac{[G \star \sum_k i_k \delta_{p_k}](p)}{[G \star \sum_k \delta_{p_k}](p)}, \tag{5.17}$$

where δ_{p_k} is a Dirac distribution centered at p_k.

Finding the sampling points that are close enough to a given point is a time consuming task. Moreover, the positions of the sampling points are only known up to a certain accuracy and the *resolution* of the reconstruction is always limited by the spacing of the chosen grid Γ. Our method will therefore convert the point cloud to a list of pixels in the reconstruction grid Γ by using a nearest neighbor rule. This mapping is the key to the very efficient method presented here.

Algorithm 10 (Discrete Shepard's like method)

- Let $\{(p_k, i_k) \in \Omega \times \mathbb{R}\}$ be the set of sampling points and their associated signal. Let Γ be a chosen reconstruction grid.

- Create two uniformly sampled images N and D defined on Γ and initialized to zero.
- For all point indexes k:
  - Map the sampling point p_k to the closest pixel p_k^s of the chosen grid Γ.
  - Update N: $N(p_k^s) \leftarrow N(p_k^s) + i_k$
  - Update D: $D(p_k^s) \leftarrow D(p_k^s) + 1$
- Smooth N and D with a recursive Gaussian filter [Deriche 93].
- Create the final reconstruction $R = \frac{N}{D}$ ■

This scattered data approximation technique requires only two Gaussian filtering and one division and is thus very efficient. The smoothness is controlled by the variance of the Gaussian kernel.

5.6.2 Mosaic Construction

The scattered data approximation can further be tailored for the problem of mosaic reconstruction. Once an estimate $\hat{f}_n$ of the frame-to-reference mapping f_n is available, we get a point cloud composed of all transformed sampling points from all the input frames

$$\{(p_k, i_k)\} = \{(\hat{f}_n(p), I_n(p)) | p \in \Lambda_n, n \in [0, \ldots, N]\}, \tag{5.18}$$

where Λ_n is the set of sampling points in the input frame n.

A common approach, when constructing mosaic images, is to minimize the seam artifacts by using feathering or alpha blending [Uyttendaele 01]. With this approach, the mosaic image is a weighted average on the input images and the weighting factor depends, for example, on the distance to the center of the image. This weighting allows to smooth the transitions between the input images.

Moreover, one could have a confidence measure for each input frame based, for example, on the estimated velocity of the frame. Weighting the input frames with this confidence measure for the reconstruction of the mosaic could help producing a visually more pleasing mosaic image.

Both approaches reduce to weighting the importance of a given mapped input sampling point p_k by some factor ρ_k. This weight can readily be used in (5.17) and we get:

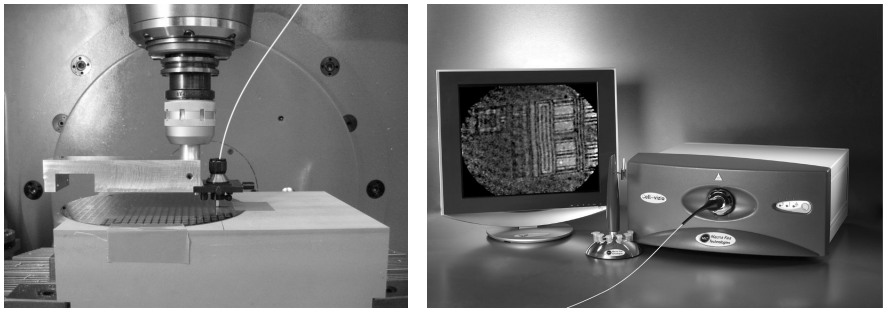
$$\hat{I}(p) = \frac{\sum_k \rho_k i_k G(p - p_k)}{\sum_k \rho_k G(p - p_k)} = \frac{[G \star \sum_k \rho_k i_k \delta_{p_k}](p)}{[G \star \sum_k \rho_k \delta_{p_k}](p)}.$$

Adapting the algorithm is thus straightforward. We only need to change the N and D update steps by $N(p_k^s) \leftarrow N(p_k^s) + \rho_k i_k$ and $D(p_k^s) \leftarrow D(p_k^s) + \rho_k$. Finally, there is often no need to extrapolate the data too much or to reconstruct zones that have a very small confidence. This can easily be achieved by setting the value of the mosaic to an arbitrary (but fixed) value when the total weight factor $D(p)$ is below a predefined threshold.

5.7 Results

5.7.1 Experimental Evaluation

The experimental evaluation of our approach was carried out on a reflectance fibered confocal microscope from Mauna Kea Technologies shown in Fig. 5.9b. For the particular flexible microprobe we used throughout these experiments, the field-of-view is 220×200 µm.



(a) The Deckel-Maho milling machine holding the flexible microprobe.

(b) The Cellvizio, fibered confocal microscope.

Figure 5.9: Experimental system used to control the motion of the flexible microprobe with respect to the imaged rigid object (a silicon wafer).

In order to validate the global positioning and motion distortion compensation framework, image sequences of a rigid object were acquired. The object needed to have structures that could be seen with the reflectance fibered confocal microscope. For the mosaicing to be of interest, we also needed to see shapes whose size were larger than the field-of-view of our imaging device. We therefore chose to image a silicon wafer.

A fair evaluation can only be made by comparing the output of the algorithm with independent information. Apart from simulated data, a ground truth of the imaged region is very difficult to get. Even with a standard microscope having a comparable resolution but a greater FOV it is not easy to see on the wafer whether the exact same region is being imaged or not. In addition to the mosaic, our algorithm estimates the motion of the flexible microprobe. The following evaluations use a computer numerical control (CNC) milling machine, shown in Fig. 5.9a, to hold the flexible microprobe and prescribe a motion with respect to the silicon wafer. The accuracy of the CNC robot is of the order of magnitude of the apparent fiber inter-core distance $d_{ic} = \frac{3.3}{G} = 1.3$ µm ($G \simeq 2.5$ is the magnification of the flexible optical microprobe). This apparent fiber inter-core distance can be thought of as the resolution of the system. We have thus been able to compare the prescribed motion with the recovered one.

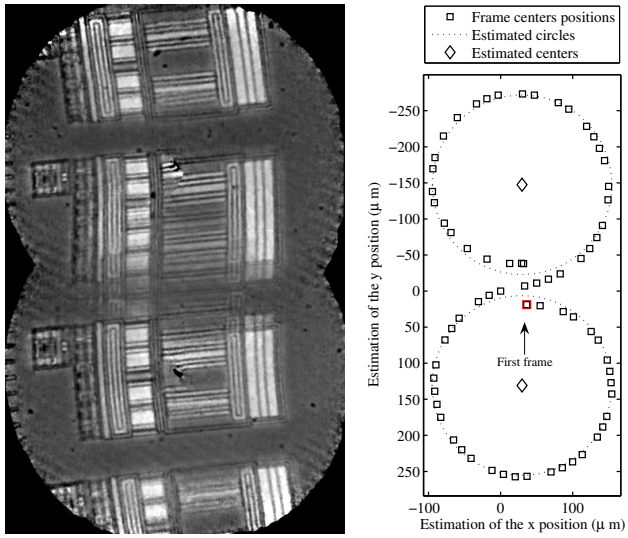
Translational Motion

In the first experiment, the milling machine was programmed to perform two consecutive circles with a radius of 125 μm. The first circle is a clockwise one whereas the second one is a counterclockwise one. The final shape thus looks like an "8." The motion starts and ends at the center of this shape. The milling machine was programmed to keep a constant tangential velocity vector during the experiment. The radius of the circles were chosen so that in the middle of the circles a small blind zone remains. We therefore have both regions where many frames overlap (the center of the "8"), and regions where fewer frames do (the top and bottom of the "8").

As a point of comparison, we compute an initial mosaic by registering the pairs of consecutive frames under a rigid motion assumption. An initial global positioning is computed by composing these local rigid body transformations. No frame distortion compensation is carried out. Fig. 5.10a shows this initial mosaic. As shown in this figure and on the reconstructed path in Fig. 5.10b, geometrical distortions appear. The estimated motion is far from the prescribed "8" and the mosaic has a wavy aspect. Furthermore, the global inconsistency of the estimated transformations is seen on the mosaic (the input frames are not correctly aligned which is especially visible in the middle of the "8"), and on the estimated trajectory (the estimation of the first and last frame centers are far away).

In Fig. 5.11a, we applied our framework for global positioning and motion distortion compensation. The gain is clear both in terms of geometry of the reconstruction and visual quality. Fig. 5.11b shows the estimated motion. The best fitted circle are shown for each half trajectory. The radii of the estimated circles are $R_1 = 123.5$ μm and $R_2 = 123.7$ μm which is quite close the 125 μm expected. The remaining difference can be explained by the uncertainty we have on the horizontal and vertical magnifications G_x and G_y of the flexible microprobe. An unmodeled discrepancy between these magnifications, i.e., $G_x \neq G_y$, could also explain the oscillations in the error plot of Fig. 5.12.

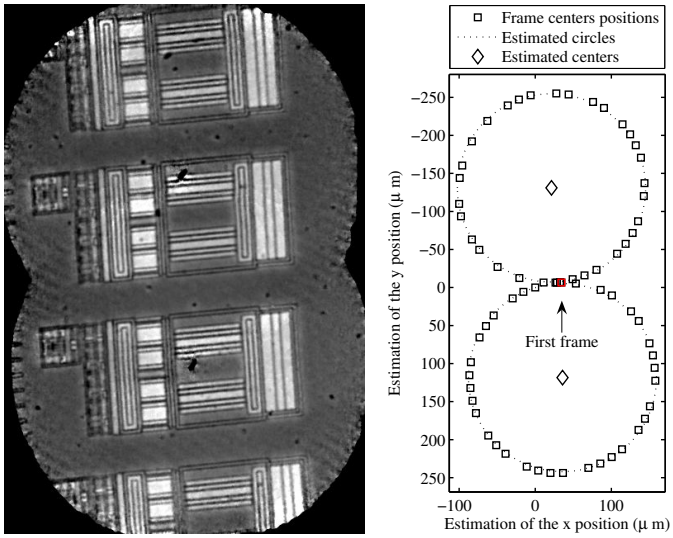
In order to further evaluate the quality of our reconstruction we used a standard microscope to acquire images of the silicon wafer on a similar zone (being able to image the very same spot is a very difficult task due to the redundancy of the wafer pattern). Fig. 5.13 shows this image. A comparison is made between the standard microscope, the reflectance fibered confocal microscope and our reconstruction from the FCM by showing a zoom on a particular structure. Note the enhancement in visual quality we get on the mosaic image with respect to the input signal we use. Because of the importance of the noise level on the input frames and of a small residual aliasing effect, our method is even capable of producing mosaics that have a better resolution that the input frames. This result is crucial since we do not even need to perform computationally expensive super-resolution algorithms as in [Zomet 00]. The gain in resolution is clear on all parts of the mosaic as shown in Fig. 5.14.



(a) Initial output mosaic of the wafer.

(b) Estimated trajectory.

Figure 5.10: Mosaicing results using sequential rigid registrations only for a video sequence during which the milling machine performed two opposite circles of equal diameters. FOV: 465×725 µm. Note the global inconsistency of the alignments (due to accumulated registration errors) and the geometric distortions (due to uncompensated motion artifacts).



(a) Globally consistent mosaic of the wafer.

(b) Estimated trajectory.

Figure 5.11: Mosaicing results using global frame positioning and motion distortion compensation on the same input sequence as in Fig. 5.10. FOV: 474×696 µm.

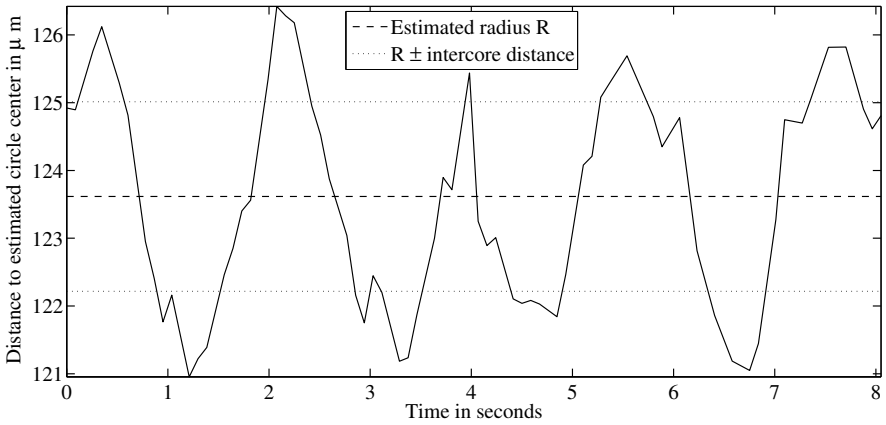
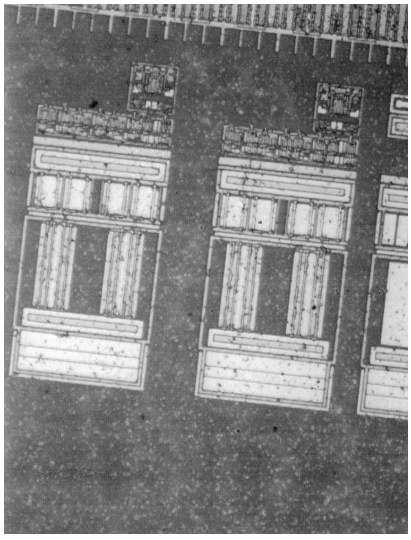
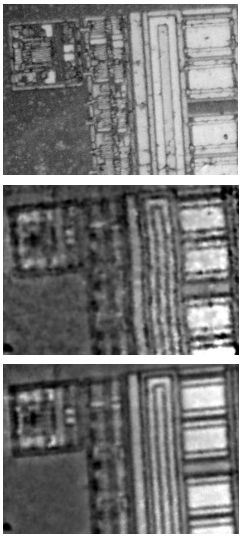


Figure 5.12: Distance of the estimated frame centers to the center of the circle that has been fitted to the estimated trajectory. The plot also shows the estimated radius and a band of width two times the fiber inter-core distance which provides an idea of the accuracy of the estimation.



(a) Image of a similar zone of the silicon wafer acquired with a standard microscope.



(b) Zoom on a portion of the wafer. Top to bottom: Image acquired with a standard microscope; a typical input frame; the reconstructed mosaic.

Figure 5.13: Comparison of our mosaic with a typical input frames and with an image acquired with a standard microscope.

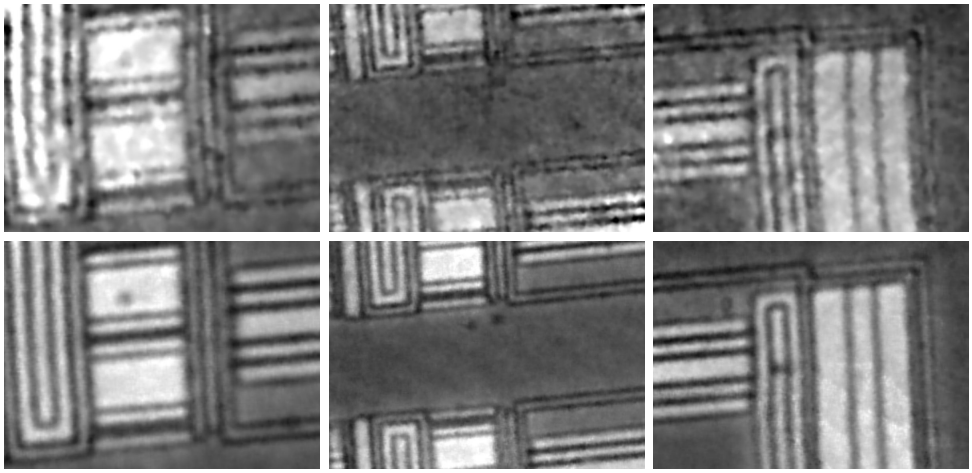


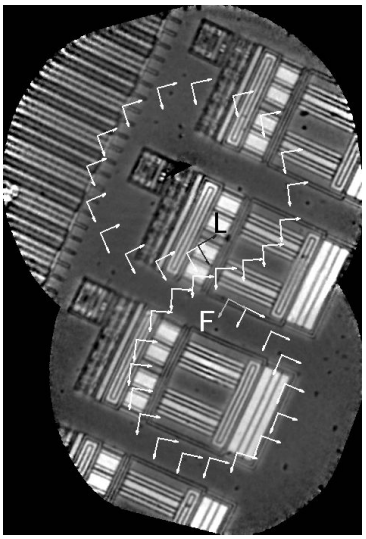
Figure 5.14: Zoom on a portion of the wafer. The first line show some input frames and the second show the corresponding reconstruction. Note the achieved gain in both noise level and resolution.

General Motion

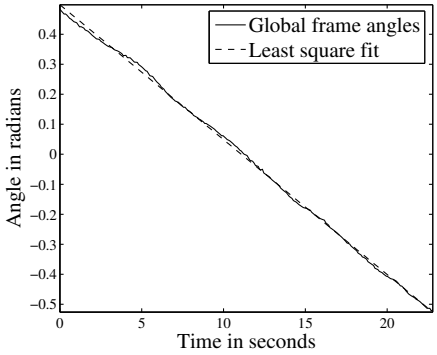
In the second experiment, the milling machine was again programmed to perform two circles with radius 125 μm. In addition to this motion, the table holding the silicon wafer was programmed to perform a rotation of angle $-\frac{\pi}{3}$ around a fixed axis. Both motions have been synchronized to start and end at the same time instant. The rotational velocity of the table was programmed to remain constant. Once again, the milling machine was programmed to keep a constant tangential velocity vector during the experiment. Because of the custom made part needed for the milling machine to hold the flexible microprobe, it was not possible to program the center of rotation of the table to be aligned with the center of the "8" motion imposed by the milling machine. We have however been able to roughly position it there.

Fig. 5.15a shows the mosaic reconstructed using our global positioning and motion distortions compensation algorithms. We have superimposed the estimated trajectory of the flexible microprobe together with two axes showing the estimated orientation of the flexible microprobe. For obvious reasons of clarity, only a subset of the estimated positions are shown. Note that the first and last frame (denoted by F and L on Fig. 5.15a) are not in the same position. This is due to the rotation center of the table not being aligned with the center of the first frame, i.e., the center of the "8" motion. Fig. 5.15b shows the estimated angular orientation of the frame together with the best fitted line. The total angle of $-\frac{\pi}{3}$ corresponds to a true angular velocity of 4.60×10^{-2} rad/s. Using this least-squares fit of the orientation, the estimated angular velocity is 4.50×10^{-2} rad/s which is once again very close to the ground truth.

The power of our hierarchical scheme appears in Fig. 5.16 where different recon-

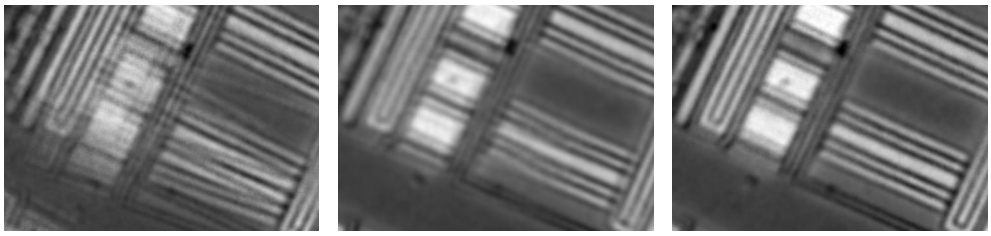


(a) Mosaic using global frame positioning.



(b) Estimated angle and fitted line.

Figure 5.15: Mosaicing results using global positioning and motion distortion compensation for a video sequence during which the milling machine performed two opposite circles of equal diameters while, in the meantime, the table performed a rotation of angle $-\frac{\pi}{3}$ with constant angle velocity. The center of rotation of the table was relatively close to the center of the "8" motion imposed by the milling machine but remains unknown. FOV: 514×758 µm.



(a) Mosaic using sequential rigid registrations only

(b) Mosaic using sequential rigid registrations and motion distortion compensation

(c) Mosaic using global frame positioning and motion distortion compensation

Figure 5.16: Zoom on a portion of the reconstructed mosaic using the same input sequence as in Fig. 5.15.

structions are compared. The first reconstruction uses sequential rigid registration only, global inconsistency is obvious. The second shows the reconstruction using motion distortion compensation but no global alignment. Residual global misalignment implies a blurred result. On the other hand, the mosaic using both global alignment and motion distortion compensation is crisp.

5.7.2 *In Vivo* Studies

Fibered confocal microscopy is designed to visualize cellular structures in living animals. In Fig. 5.1, we presented different types of tissue imaged *in vivo* with the Cellvizio. As a proof of concept, we applied our mosaicing algorithm to the sequences corresponding to the images shown in Fig. 5.1. We will first discuss the results obtained for colon imaging in detail. Then we briefly present the results obtained with the other sequences.

Colon Imaging

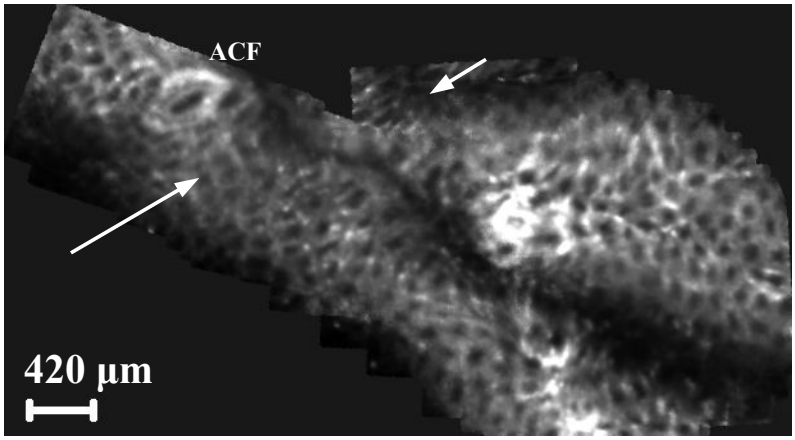
In the field of colon cancer research, the development of methods to enable reliable and early detection of tumors is a major goal. In the colon, changes in crypt morphology are known to be early indicators of cancer development. The crypts that undergo these morphological changes are referred to as Aberrant Crypt Foci (ACF) and they can develop into cancerous lesions.

In laboratory rodents, ACF can either be induced by colon-specific carcinogens or through transgenic mutations. As in humans, ACF in mice are known to be reliable biomarkers for colon cancer and are used to study initiation, promotion and chemoprevention of colorectal cancer. Currently, ACF are routinely imaged, detected and counted under a dissecting microscope following staining with methylene blue. The procedure is to sacrifice the animal, excise the whole colon and open it flat.

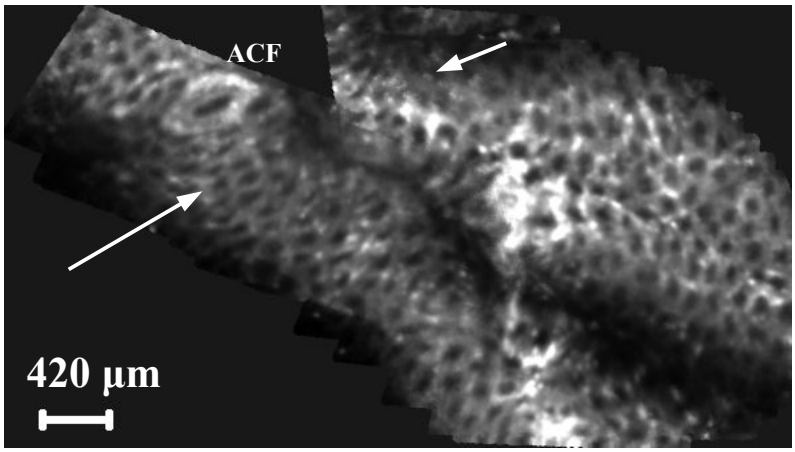
Compared to this method, fluorescence FCM enables the operator to see the lesions in real time and to make an almost immediate evaluation without sacrificing the animal. This offers the possibility of studying groups of individual animals over extended periods with the benefits of reduced inter-animal statistical variation and reduced number of animals used per experiment [Cavé 05]. However, in many cases the limited field-of-view restricts the confidence that the operator has in the ACF counting. By offering an extended field-of-view, mosaicing techniques can be an answer to this restriction.

The effectiveness and relevance of the proposed algorithm for this study are shown on a sequence that has been acquired *in vivo* on a mouse colon stained with acriflavine at 0.001%. The mouse was treated with azoxymethane (AOM) to induce a colon cancer. The input sequence is composed of fifty frames, each with a field-of-view of 425 µm by 303 µm. As shown in Fig. 5.17c, our algorithm allows for a simultaneous visualization of normal and aberrant crypts.

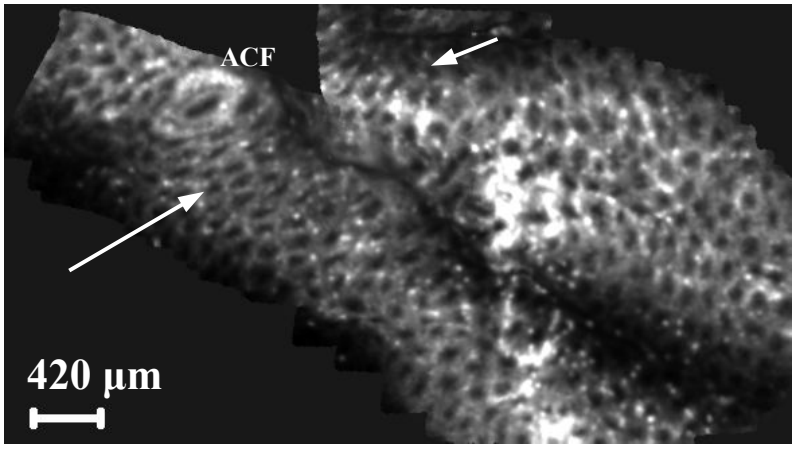
Fig. 5.17b shows the improvement we make with respect to a simple mosaic



(a) Using pairwise rigid registrations of consecutive frames.

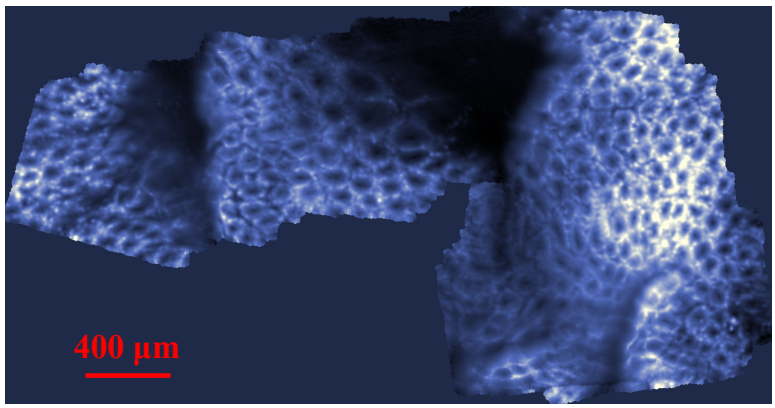


(b) Using global positioning and motion distortion compensation.

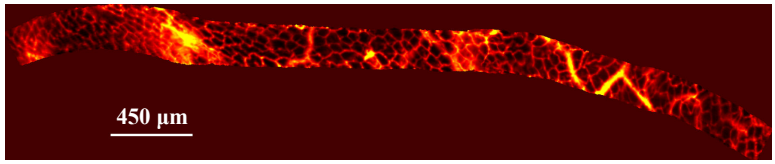


(c) Using our complete algorithm (including non-rigid registration).

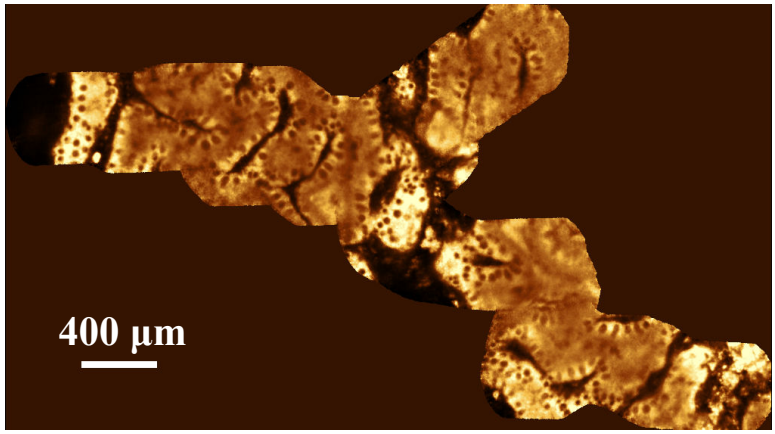
Figure 5.17: Mosaics of 50 *in vivo* mouse colon images after instillation of acriflavine (Courtesy of D. Vignjevic, S. Robine and D. Louvard, Institut Curie, Paris, France). The arrows point to zones of the mosaic where the visual gain is particularly appealing.



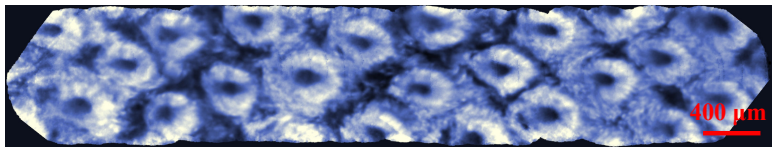
(a) *Ex vivo* mouse colon imaging after instillation of acriflavine (226 input frames). Courtesy of D. Vignjevic, S. Robine and D. Louvard, Institut Curie, Paris, France.



(b) *In vivo* Mouse colon vascularization after injection of FITC-Dextran high MW (300 input frames). Courtesy of M. Booth, MGH, Boston, MA.



(c) *Ex vivo* reflectance imaging of human colon (1500 input frames).



(d) *Ex vivo* imaging of human colon with methylene blue (51 input frames). Courtesy of P. Validire, Institut Mutualiste Montsouris, Paris, France.

Figure 5.18: Pseudo-color mosaics of the colon.

construction, shown in Fig. 5.17a, by using the global positioning and motion distortion compensation. The imaged tissue is very soft and non-linear deformations also occur. Fig. 5.17c illustrates the gain we obtain by taking into account those non-rigid deformations. Some details are lost if we only use rigid body transformations and reappear on our final mosaic. The global frame positioning mosaicing took approximately 3 min on a 2GHz P4 and 15 min if the non-rigid deformations are compensated.

Our method has also been successfully applied to many other types of sequences acquired in both mouse and human colon as shown in Fig. 5.18. Fig. 5.18a provides another example using a sequence that has been acquired on a mouse colon stained with acriflavine. In Fig. 5.18b, the vascularization of the mouse colon was imaged after injection of FITC-Dextran high MW. We also provide results on sequences that have been acquired *ex vivo* in the human colon. Fig. 5.18c uses reflectance FCM whereas Fig. 5.18d uses a sequence acquired with a 660 nm fluorescence FCM after methylene blue staining.

Other Examples

The Cellvizio offers a new way to image and characterize many types of tissue. In many cases, mosaicing can help move beyond the limitations of FCM by offering an extended field-of-view. We provide some insight of this by showing the result of our algorithm on the remaining sample sequences shown in Fig. 5.1.

Fig. 5.19a shows a mosaic constructed from 21 input frames, each with a FOV of 417 µm by 297 µm. On this figure, we see mouse tumoral angiogenesis. The need for *in vivo* imaging is urgent in this field. It can indeed help assess the efficiency of angiogenesis therapy [McDonald 03]. Mosaicing techniques can further help computing objective quantitative measurements.

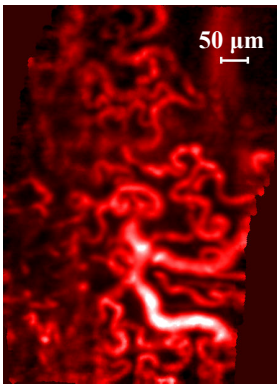
The result shown in Fig. 5.19b is of much clinical interest since it proves that obtaining a microscopic images of human lung tissue without any staining is feasible. Our mosaicing algorithm pushes this interest one step further by showing multiple alveolar structures on a single image.

The mosaic in Fig. 5.19c, arising from 31 input frames, shows the tubular architecture of the kidney. In this setting, mosaicing could help producing objective statistical shape measurements.

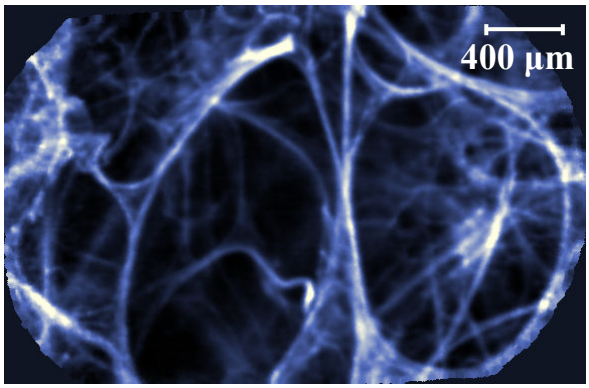
Fig. 5.19d shows the ability of the Cellvizio to image nervous tissue down to the dendritic endings and shows how mosaicing can help seeing many of those dendritic endings simultaneously. 70 input frames all with a FOV of 397 µm by 283 µm were used to produce the mosaic.

Fig. 5.20 clearly shows that even in the *in vivo* case, we are able to get a real gain both in terms of noise level and resolution. The input sequence was composed of 101 images from the human mouth mucosa that have a FOV of 150 µm by 125 µm.

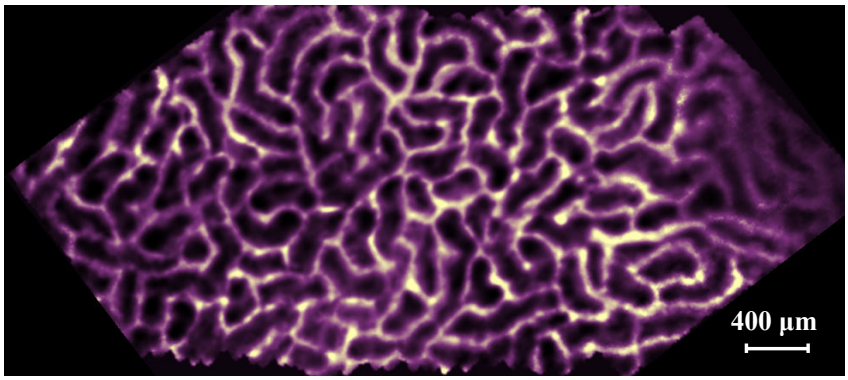
The results shown in Fig. 5.17, Fig. 5.19 and Fig. 5.20 prove the feasibility of mosaicing for *in vivo* soft tissue microscopy.



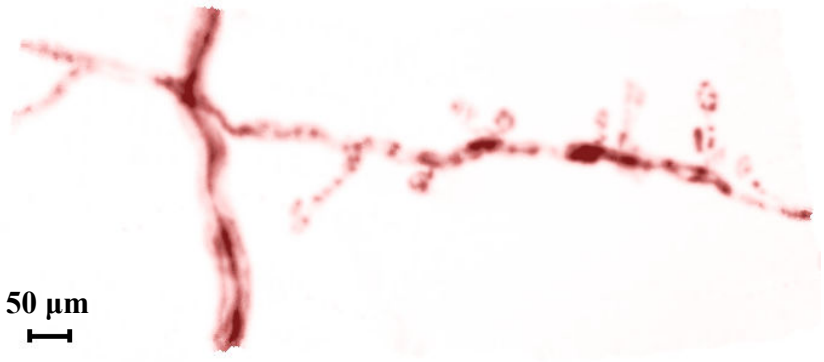
(a) *In vivo* tumoral angiogenesis in mouse with FITC-Dextran high MW (21 input frames).



(b) *Ex vivo* autofluorescence imaging in human lung (15 input frames).

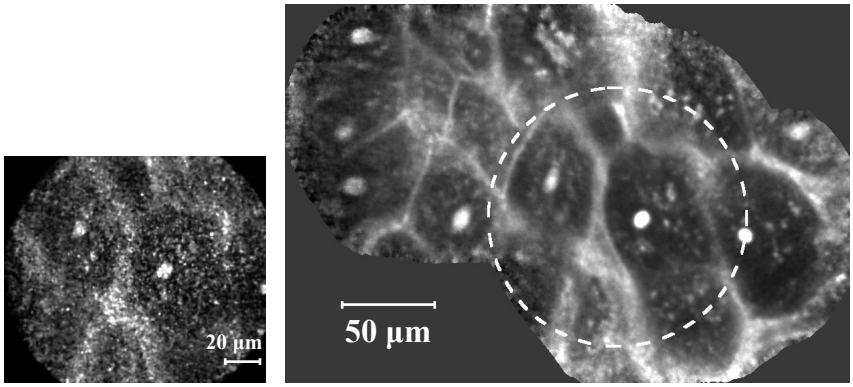


(c) Microcirculation of the peritubular capillaries of a live mouse kidney with FITC-Dextran high MW (31 input frames).



(d) Dendritic receptors in a live Thy1-YFP mouse (70 input frames).

Figure 5.19: Pseudo-color mosaics using different types of images acquired with the Cellvizio (Courtesy notes appear in Fig. 5.1).



(a) First frame of the input sequence (cf. Fig. 5.1c).

(b) Final mosaic (101 input frames). The dashed circle corresponds to the first input frame shown in Fig. 5.20a.

Figure 5.20: Reflectance imaging of human mouth mucosa.

5.8 Conclusions

The problem of video mosaicing for *in vivo* soft tissue fibered confocal microscopy has been explored in this chapter. A fully automatic robust hierarchical approach was proposed. Rigorous tools for estimation problems on Lie groups were used to develop a robust algorithm to recover consistent global alignment from local pairwise registration results. A model of the relationship between the motion and the motion distortion was developed and used to robustly compensate for the motion distortions arising when using a laser scanning device. An efficient scattered data fitting technique was proposed for the construction of the mosaic. This tool was also used to adapt the *demons* algorithm for non-rigid registration with irregularly sampled images.

In terms of algorithmics, future work should aim at making the mosaicing process more efficient. This could, for instance, be based on using less accurate (yet globally consistent) frame-to-reference transformations and compensate for the inaccuracy by using reconstruction methods that are less sensitive to small misregistrations [Burt 83, Levin 04, Su 04]. We should also work on providing the user with a confidence measure on the reconstructed mosaic.

The results shown on various types of images acquired with a fibered confocal microscope are promising and encourage the application of the proposed method for qualitative and quantitative studies on the mosaics. An strong technical validation of our method was performed thanks a computerized numerical control milling machine. The next step, which we address in the next chapter, is to evaluate the usefulness of our algorithm within a real clinical setting.

Appendix: Lie Group Tools for 2D Rigid Transformations

In this section, we provide the explicit expressions of the tools defined in Section 5.3 for the specific case of interest, the 2d rigid body transformations $\mathcal{R}$.

A rigid body transformation r is composed of a rotation of angle θ_r (expressed in $[0, 2\pi] \mod (2\pi)$) followed by a translation $\tau_r = (\tau_r^x, \tau_r^y)$. This set of parameters provides a chart or local coordinate system. The composition and inversion are easily expressed in this coordinate system. Let r and s be two rigid body transformations represented by (θ_r, τ_r) and (θ_s, τ_s), we have:

$$\begin{aligned} s \circ r &: (\theta_s + \theta_r, R_{\theta_s}.\tau_r + \tau_s) \\ r^{-1} &: (-\theta_r, -R_{-\theta_r}.\tau_r), \end{aligned}$$

where $R_\alpha = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix}$ is a rotation matrix. In this coordinate system, the differential maps previously presented are given by:

$$DL_r(s) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta_r) & -\sin(\theta_r) \\ 0 & \sin(\theta_r) & \cos(\theta_r) \end{bmatrix}$$

$$DR_r(s) = \begin{bmatrix} 1 & 0 & 0 \\ -\sin(\theta_r)\tau_s^x - \cos(\theta_r)\tau_s^y & 1 & 0 \\ \cos(\theta_r)\tau_s^x - \sin(\theta_r)\tau_s^y & 0 & 1 \end{bmatrix}$$

$$DInv(r) = \begin{bmatrix} -1 & 0 & 0 \\ \sin(\theta_r)\tau_r^x - \cos(\theta_r)\tau_r^y & -\cos(\theta_r) & -\sin(\theta_r) \\ \cos(\theta_r)\tau_r^x + \sin(\theta_r)\tau_r^y & \sin(\theta_r) & -\cos(\theta_r) \end{bmatrix}.$$

We can notice here that the left composition differential map $DL_r(s)$ does not depend on s.

We now look at the left-invariant Riemannian metric $G(x)$ we use on this Lie group. Let $\boldsymbol{u}_r = (\mathrm{d}\theta_r, \mathrm{d}\tau_r)$ and $\boldsymbol{u}_s = (\mathrm{d}\theta_s, \mathrm{d}\tau_s)$ be elements of the Id-tangent space $T_{\mathrm{Id}}(\mathcal{R})$. We choose the canonical dot product on $T_{\mathrm{Id}}(\mathcal{R})$:

$$\langle \boldsymbol{u}_r \,|\, \boldsymbol{u}_s \rangle_{\mathrm{Id}} = \mathrm{d}\theta_r \mathrm{d}\theta_s + \mathrm{d}\tau_r^T.\mathrm{d}\tau_s \quad \Leftrightarrow \quad G(\mathrm{Id}) = I_3,$$

According to (5.3), the left invariant metric is thus given, at all points, by the matrix $G(x) = DL_r^{-T}.I_3.DL_r^{-1} = I_3$. The Riemannian metric is, in this case, of the simplest possible form and does not depend on the tangent space. This particular Lie group is thus flat and geodesics are straight line. The Riemannian logarithm at the identity is therefore simply given by the parameters expressed in the chart we chose:

$$\log_{\mathrm{Id}}(r) = [\theta_r, \tau_r^x, \tau_r^y]. \tag{5.19}$$

Another common choice, would be to take $G(\mathrm{Id}) = \mathrm{diag}([\lambda^2, 1, 1])$ in order to weight the influence of the angles with respect to the translation, but using the Mahalanobis distance defined in Section 5.3.3 makes this weighting unnecessary.

Chapter 6

Multicenter Clinical Trial

Table of Contents

6.1 Motivation: From the Bench to a Benchmarked Bedside

As is exemplified by some parts of this thesis, research in biomedical image analysis is often driven by a technology-oriented point of view or a theoretical-development-oriented point of view. Biomedical image analysis research is, however, by essence an application-oriented field. Therefore, it seems necessary that with new developments, we also answer the basic question of its potential benefit in terms of biological research or public health.

This chapter aims at determining the actual contribution of our mosaicing tool to the interpretation process of Cellvizio data acquired during gastrointestinal endoscopy. This work is a continuation of our previous efforts in terms of clinical validation presented in [Becker 07, Perchant 06, Thiberville 07, Vercauteren 07a].

In [Meining 07b] for example, Cellvizio was used during a colonoscopy of a 75-year-old man having an 8-month history of diarrhea. Colonic crypts were slightly distorted but not destroyed. Mosaicing allowed clinicians to detect an unusual distance between crypts. This was somewhat difficult to notice directly on the Cellvizio video sequence as only a few crypts can be seen in the field of view. This unusual distance led the clinician to diagnose lymphocytic colitis from the mosaic image. Standard histopathology then confirmed this presumptive diagnosis. This kind of study indeed shows some potential benefits of mosaicing for Cellvizio data. Such work is, however, based on some specifically selected cases and does not prove the scalability of the tools. In essence, it can only be considered as a proof of concept.

Our goal here is to measure the value of our tools on a well-defined medical problem, using standard real-life clinical procedures. In this chapter, we chose to focus on the surveillance of a pre-cancerous condition known as Barrett's esophagus. The strongest evidence one can get to support a clinical hypothesis is provided by a systematic review of a randomized multicenter clinical trial. We have thus integrated our mosaicing tool within such a trial whose main goal was to measure the performance of Cellvizio with respect to the current gold standard for Barrett's esophagus. This trial allows us to get a fair quantitative evaluation of our mosaicing tools for a specific medical application.

Advanced image processing schemes are often time-consuming or require a rather specific setup unfamiliar to clinicians. As such their use is often difficult to integrate within a routine clinical workflow. Therefore, our aim, when we incorporated the mosaicing into the clinical trial, was to integrate all facets of clinical constraints to propose a complete mosaicing solution usable by medical investigators.

The remainder of this chapter is organized as follows. In Section 6.2, we present the motivation behind the Barrett's esophagus multicenter clinical trial and introduce its protocol. In Section 6.3, we show how we designed a complete mosaicing solution that meets clinical constraints. As this mosaicing solution does not provide clinicians with real-time information, it might be difficult for them to envision how they could take full advantage of mosaic images. We thus propose in Section 6.4 a live mosaicing scheme that provides the investigators with direct visual feedback to help them optimize their Cellvizio acquisition. Finally, the outcome of the clinical trial appears in Section 6.5.

6.2 A Multicenter Clinical Trial for the Diagnosis of Barrett's Esophagus

Barrett's esophagus refers to an abnormal change in the cells of the lower end of the esophagus. It is considered to be a pre-malignant condition and is associated with an increased risk of esophageal cancer. As shown in Fig. 6.1, Barrett's esophagus is visible grossly through an endoscope, but the tissue must be examined at the microscopic level to confirm the diagnosis and determine the malignancy of the cells.

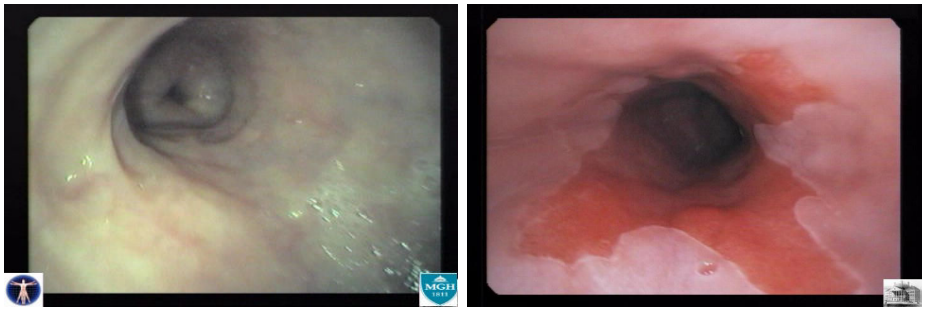


Figure 6.1: Endoscopic images of the esophagus. Left: healthy esophagus. Right: Barrett's epithelium is recognizable by its salmon color whereas the normal mucosa has a pearly white appearance. Images taken from the DAVE project [Kelsey 04, Kelsey 05].

It is widely accepted that the current gold standard for the surveillance of Barrett's esophagus, which is based on somewhat random biopsies [Sampliner 02], is, at the same time, not specific enough, not reproducible enough and uncomfortable for the patient [Sharma 07].

We have seen in the previous chapters that fibered confocal microscopy allows the physician to get a microscopic view of the tissues during endoscopy. In this case, we are thus not limited by the number of biopsies and can expect a diagnosis that is more specific, more reproducible and more comfortable for the patient. The main goal of the multicenter clinical trial on Barrett's esophagus is to assess the performance of fibered confocal microscopy with respect to the current gold standard. The trial is sponsored by Mauna Kea Technologies. The investigators are PD. Dr. Alexander Meining, Technical University Munich (lead investigator), Pr. Dr. Thomas Rösch Charite University, Berlin and Pr. Dr. Stephan Miehlke, Dresden University of Technology.

An additional goal of the trial is to compare the diagnosis based on only the mosaics, which have been constructed from a Cellvizio video sequence, with respect to both the diagnosis based on only the Cellvizio video sequences and the one based on the gold standard. This will serve to validate our mosaicing algorithm from an applicative point of view.

6.2.1 Basics About Barrett's Esophagus

The esophagus is a muscular tube that carries food and saliva from the mouth to the stomach. Sometimes a small amount of material comes back up from the stomach to the esophagus. Such occasional reflux is considered to be normal. However, when it happens frequently, it may produce burning symptoms that are often referred to as heartburn. It is considered to be a medical condition known as gastroesophageal reflux disease (GERD).

GERD can provoke an inflammation of the mucosa of the esophagus (esophagitis). This in turn can sometimes lead to Barrett's esophagus (BE) which is a condi-

tion, named after Numan R. Barrett (1903–1979), in which the esophagus changes so that some of its lining is replaced by a type of tissue similar to that normally found in the intestine. This process is called intestinal metaplasia. The origin of BE probably involves multipotential undifferentiated cells.

Barrett's esophagus itself does not cause any particular symptoms. It is, however, important because the intestinal metaplasia that occurs in BE confers an increased cancer risk of the adenocarcinoma type. It is in fact the most common precursor of esophageal cancer. BE has a poor prognosis[1], mainly because individuals that come for consultation already have late-stage disease. Once cancer is diagnosed, patients have a median survival time of less than one year. Less than 10 percent of patients survive for longer than five years despite combined chemotherapy and surgery. It is thus important to detect Barrett's esophagus as early as possible.

The rate of esophageal adenocarcinoma is increasing in the Western world. Esophagitis secondary to GERD is the most common medical condition in Western countries. Among the 30 percent of adults who have heartburn at least once a month, a third have endoscopic evidence of esophagitis. In 10 percent of patients with esophagitis, the condition progresses to Barrett metaplasia. Approximately 0.5–2.0 percent of adults in the Western world have Barrett metaplasia. As many as 11,000 people in the US will be diagnosed with esophageal cancer in the next year. Estimates of the incidence of Barrett's esophagus range between 800,000 and two million cases.

6.2.2 Current Medical Practice

Diagnosing Barrett's esophagus is difficult because it often does not exhibit specific symptoms. When heartburn or acid regurgitation is the dominant symptom, the specificity is sufficiently high to diagnose GERD. If no additional indication for further evaluation exists, patients may be confidently treated for GERD without undergoing confirmatory tests. An endoscopic examination with biopsy is however required to confirm the diagnosis because Barrett's esophagus can only be diagnosed at the microscopic level.

For other diseases, it is often the case that advanced diagnosis need only be performed on high-risk populations. For Barrett adenocarcinoma, the problem is that conventional clinical risk factors are neither sensitive nor specific enough for the classification of individuals with a high risk. Besides GERD, risk factors for Barrett's esophagus include being a man, being white or Hispanic, and being an older adult. Therefore, surveillance should be required for almost all patients with GERD. This approach is unfortunately neither feasible nor cost effective.

Although controversy exists, many physicians recommend that adult patients who are over the age of 50 and have had GERD symptoms for more than five years be screened for Barrett's esophagus. With the recognition of BE as a pre-malignant lesion, the crucial issue is surveillance. The goal of surveillance is to detect early

[1] prediction of the future course of a disease and chance for recovery

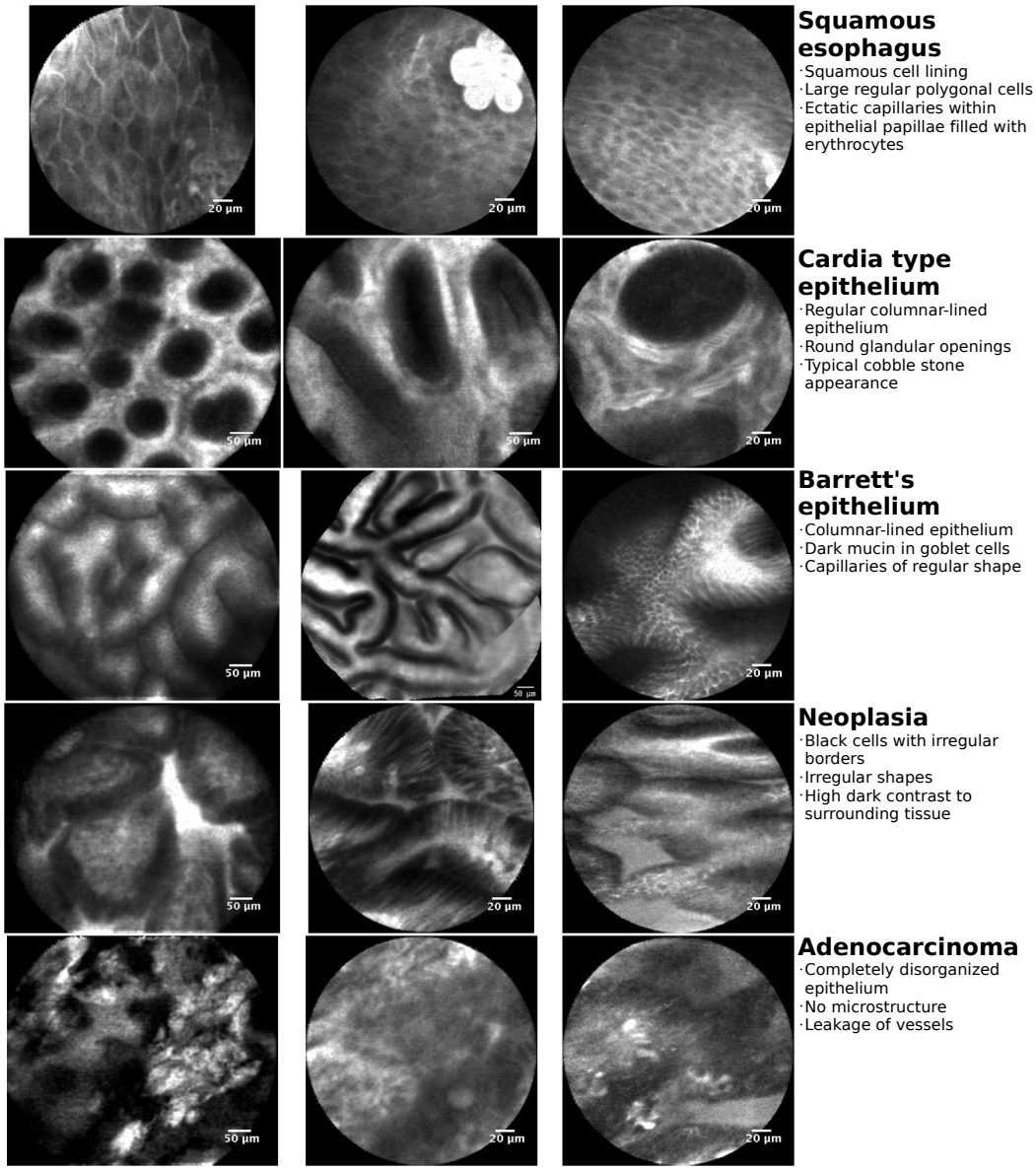


Figure 6.2: *In vivo* fibered confocal microscopy of the esophagus with Cellvizio. A classification of different tissue types can be made from the cellular architecture.

signs of cancer. When people who have Barrett's esophagus develop cancer, the process goes through an intermediate stage in which cancer cells appear in the Barrett's tissue (dysplasia). The process is patchy and cannot be seen directly through the endoscope, so that today multiple biopsies must be taken. Even then, the cancer cells can be missed and dysplasia is not the ideal marker for selecting patients with a high risk for adenocarcinoma. Approximately 40 percent of patients who have high-grade dysplasia without cancer, according to endoscopic biopsy findings, are found to have adenocarcinoma at surgery.

If the patient has esophageal cancer, or BE and high-grade dysplasia, the doctor may recommend a major surgical procedure in which the esophagus is removed completely and the stomach is pulled into the chest (esophagectomy). Although this treatment is effective, it is associated with significant health risks.

The surgical treatment of people with high-grade dysplasia is controversial. Some experts believe that esophagectomy should be used as a measure to protect against cancer. Other experts believe that it is sufficient to schedule screening endoscopies every three to six months and perform an esophagectomy only if cancer develops. Doctors generally do not recommend surgery for people with declining health or for those who are too weak to withstand a major procedure.

6.2.3 Fibered Confocal Microscopy as an Alternative

We have seen that fibered confocal microscopy offers unprecedented characterization capabilities of the GI mucosa in real time during endoscopy. It is not limited by the number of biopsies so that the endoscopist can benefit from a real-time assessment of the malignancy of the tissue in any area. This can potentially improve the specificity, robustness, and patient comfort as well as limit the cost of the current surveillance protocol.

In a feasibility study conducted by PD. Dr. Alexander Meining, fibered confocal microscopy was used to detect malignant and pre-malignant modifications in the gastrointestinal tract. The conclusion of this feasibility study is that Cellvizio yields a 92 percent accuracy in the detection of neoplasia as compared with conventional histopathology. Figure 6.2 provides a feeling of how discriminant Cellvizio images can be with respect to different conditions. The disadvantage of the preliminary study was, however that the acquisition protocol was not optimal so that many images were not of sufficient quality for diagnosis purposes. Early experiments with the optimized protocol used for the clinical trial showed significantly better image quality and reproducibility.

Using a distal scanning fibered confocal microscope, another study presented in [Kiesslich 06] was shown to allow for a high accuracy in the diagnosis of both Barrett metaplasia and Barrett neoplasia. A possible weak point of this work is that the system used by the authors cannot be operated via any standard endoscopes as can Cellvizio. In addition to this, the required quantity of contrast agent (fluorescein) is 10 times as much as the one used in the present trial. This involves a higher risk of side effects. Moreover, a multicenter verification of their data is not available.

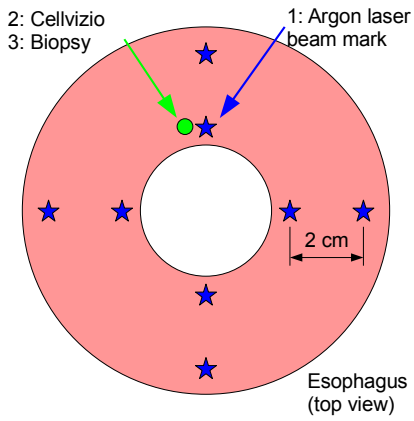


Figure 6.3: Schematic view of the acquisition protocol as seen from the endoscope. The complete four quadrant biopsy pattern is first marked with an argon plasma coagulator. Cellvizio is then used on the left hand side of each mark. Finally, a biopsy is taken at each spot imaged with Cellvizio.

6.2.4 Clinical Trial Protocol

The main objective of the clinical trial sponsored by Mauna Kea Technologies is to compare the current gold-standard practice with fibered confocal microscopy to detect metaplasia or neoplasia in patients suffering from Barrett's esophagus. An additional goal, of essential importance for us, is to compare the performance of diagnosis on the mosaics with respect to the two previous modalities.

Double-blind multicenter clinical trials are considered the most reliable form of scientific evidence for this kind of comparison because they eliminate all forms of spurious causality. For this reason, the protocol that has been established dispatches the patients across three different centers. For each acquisition, the gold-standard diagnosis, the diagnosis based on fibered confocal microscopy and the one based on the mosaics are performed in a double-blinded mode, i.e., independently from each other and without seeing the data from the other modalities.

Gold Standard: Four Quadrant Biopsy For the purpose of detecting neoplasia in the esophagus at an early stage, it is the current practice to take one piece of tissue from each quadrant of the esophagus (left, right, front, and back). This is repeated every 1–2 cm throughout the length of the Barrett's segment. Any endoscopic peculiarities such as elevations or ulcers are examined by a separate, additional biopsy. Therefore, at least $4\times3=12$ biopsies have to be carried through in case of a four-centimeter-long sleeve-shaped Barrett's esophagus. All specimens obtained with a biopsy should be fixed separately. This protocol, known as *four quadrant biopsy*, is considered to be the current best medical practice and is thus used as a gold standard.

Acquisition Protocol In order to fairly compare fibered confocal microscopy with the gold standard, it is necessary to acquire the Cellvizio sequences and the biopsies on the same spot. To do that, the endoscopist starts by marking all the areas where a biopsy is to be executed according to the four quadrant biopsy protocol. The mark is performed by means of an argon plasma coagulator. About ten marks will be made for each patient. In a second step, the patient receives an intravenous (IV) injection of fluorescein. Cellvizio sequences are then acquired directly at the left-hand side of the argon marked areas. Finally, a tissue specimen is obtained, through biopsy, at the same spot where fibered confocal microscopy was used. This procedure is summarized in Fig. 6.3.

Diagnosis Protocol For recording purposes, the endoscopist will document the suspected fibered confocal microscopy findings during the acquisition. After the acquisition, the sample identification names are randomized and renamed according to a given correspondence table in order to guarantee a blinded study. The biopsy specimens are then dispatched to a pathology center for histopathological determination of the results. Histopathology is performed without any knowledge of the anamnestic[2], endoscopic, Cellvizio or other histopathologic data. The interpretation of the results of the fibered confocal microscopy sequences is performed in a blinded way analogously to the histopathologic gold standard by an histopathologist. After randomization of the fibered confocal microscopy sequences, the data is processed by our fully automatic mosaicing pipeline. A new layer of randomization and renaming is performed on the mosaics. The mosaics are then interpreted in a blinded mode by a gastroenterologist. Three months after the completion of the randomization, a new interpretation of the results of the fibered confocal microscopy sequences is performed by a gastroenterologist.

Planning the Number of Cases The planning of the number of cases is carried out for the examination of the hypothesis H1 that the specificity for Barrett metaplasia and Barrett neoplasia are not considerably worse than 95 percent [Kiesslich 06]. For this reason, the planning is to be performed for an unilateral equivalence test of the non-inferiority. A value of 90 percent is considered to be the lowest acceptable limit for a not considerably worse specificity, i.e., a deviation from the value of 95 percent by 5 percent downwards. Since two main target quantities are of interest with regard to the planned study (specificity of Barrett as well as specificity of neoplasia), the critical p-value is adjusted according to the Bonferroni correction[3] so as to keep to a global significance level of $\alpha = 5\%$. A significance level of $\alpha = 2.5\%$ is thus needed for each of the individual examinations. The error of second kind is fixed at 0.2. A prevalence of Barrett metaplasia equal to 67.6 percent and a prevalence of neoplasia equal to 16.9 percent can be assessed on the

[2] relating to the current or previous medical history of a patient

[3] The Bonferroni correction states that if an experimenter is testing n independent hypotheses on a set of data, then the statistical significance level that should be used for each hypothesis separately is $1/n$ times what it would be if only one hypothesis were tested.

basis of the literature available [Kiesslich 06]. It follows that the total number of the observation units to be recruited has to be 738. Since about 10 observation units (biopsies and Cellvizio data) are made from each patient, the consequence is that 74 patient are required.

6.3 From a Mosaicing Algorithm to a Mosaicing Solution

Within the clinical trial protocol presented above, our goal is to provide for each input Cellvizio sequence, a large field-of-view reconstruction of the imaged region and assess the diagnosis performance on these mosaics. As with any large clinical trial, the automated processing of the data poses a number of challenges in terms of image processing but also in terms of logistics. These challenges are typically not addressed by a single algorithm such as the one we presented in Chapter 5.

In order to get a consistent and reproducible mosaicing solution rather than a mosaicing algorithm, several key elements are required. First of all, the data should be acquired in a standardized manner. This point is taken care of by the acquisition protocol of Section 6.2.4. Second, the data processing needs to be reproducible and robust enough to cope with a very large database with a failure rate that should be close to zero. This will typically be addressed by a pipeline of several algorithms rather than a single algorithm, see e.g. [Jones 06, Zijdenbos 02]. Finally, since we deal with such a large amount of clinical data, the software tools that are used, need to manage the computational burden and meet the necessary regulatory requirements. This will normally require the use of distributed computing resources and a strict software version control management.

In this section, we detail the critical elements that we designed to answer these challenges and move from a mosaicing algorithm to a mosaicing solution that is able to cope with a large multicenter clinical trial.

6.3.1 A Dedicated Offline Scene-Splitting Algorithm

Motivation

Given the large amount of data that we want to process, it becomes absolutely necessary to have a fully automatic algorithm that fails in only a very few cases. As appears in Chapter 5, our mosaicing algorithm still has a major shortcoming. It will indeed be robust to some failures in the registration of pairs of images but will dramatically fail if the Cellvizio sequence that is used as an input does not present a smooth motion trajectory of the optical microprobe with respect to the soft tissue.

The initial assumption used by the mosaicing algorithm is indeed that consecutive frames of the input video sequence have a sufficient overlap for registration purposes. From this presumption, an initial estimate of the global positioning of the frames is obtained by using a rigid body registration technique to align the pairs of

consecutive frames and by composing these spatial transformations. Furthermore, our algorithm will be able to compensate for the motion distortions on the input frames up to a first order. This means that we estimate the speed of the optical microprobe with respect to the imaged tissue by assuming it is constant during the time required to acquire one frame. Obviously even if this hypothesis has proved to be reasonable on many sequences, there will always be cases where the acceleration of the optical microprobe cannot be neglected and changes over time.

We could have explicitly used more sophisticated assumptions within the mosaicing algorithm but this would have led to a space of unknowns with many more degrees of freedom than we could have robustly estimated. A more viable approach is to keep our mosaicing algorithm as it is, and embed it within a larger mosaicing solution that encloses a set of pre- and post-processing steps. The main goal being to feed our mosaicing algorithm only with sequences that we know can be processed.

Relation to Shot Change Detection Algorithms

To find good sequences, we need an algorithm that segments an input Cellvizio sequence into a set of smooth-motion sub-sequences usable for further mosaicing. We refer to these sub-sequences as scenes. This problem of scene splitting seems closely related to that of shot change detection which has been an active field of computer vision research for some time now [Gargi 00]. In this setting, even if a rigorous definition of a shot is difficult to setup, it is rather clear that a shot refers to a sequence of frames acquired continuously with one camera. It is then often the case that different shots show very different contents. Many shot change detection algorithms are thus based on measuring variations of visual features such as color histograms or variations of audio features [Boreczky 98]. These approaches are obviously not well suited to our setting where we only have grayscale images, no sound and where a scene change mostly means discontinuity of the motion of the optical probe. A number of algorithms have been proposed that perform motion analysis or optical-flow computation to determine shot changes. These methods provide a better match for our scene-splitting problem. However, each of the individual frames we work on often show little informative content. Pairs of images also often have a relatively small amount of overlap compared to regular video. It is therefore difficult to perform a meaningful block-matching analysis.

Our approach to the problem of scene splitting will thus rely on a method that searches for a global motion between the frames. This program is a rather difficult one for several reasons. First of all, it is difficult to run a complete registration procedure between all the pairs of consecutive frames as the computational requirement would be daunting. And even though we could, the problem of evaluating whether the registration process succeeded or not is far from trivial. In [Bouthemy 99], the authors proposed an interesting shot change detection algorithm. Their scheme is based on a robust image registration that aims at finding the dominant motion between consecutive frames. The quality of the registration is then evaluated by looking at the area of support that accounts for the dominant motion. It is nice to

see that the change decision is not directly related to the similarity metric that is optimized during the registration but rather to an intuitive sort of agreement measure. The main problem of this approach for us lies in the computational burden of the robust registration and on the fact that in our case, we could sometimes find a well accepted dominant motion when the images are mostly composed of noise.

Scene Splitting

It should be noted that our goal is not so to detect all the changes than to find smooth-motion sub-sequences. Our emphasis is, in the end, not to make the largest possible mosaics but rather to be sure that every single mosaic we compute is informative. It is thus not too problematic to over-segment the sequences. Our approach first looks for a global motion between consecutive frames. As we need a conservative scene splitting, we can first use a coarse global registration and then check whether or not to accept the registration results. If the coarse registration is not sufficient, a fine registration would indeed typically get stuck.

The first step of our scene-splitting algorithm is to discard all the frames that have not enough informative content. One Cellvizio sequence will typically be acquired on one given tissue type. If the clinician recorded some data while not being in contact with the tissue, there will also be frames that show only noise content. It is likely that the image intensities will not vary too much across the informative frames. We found that looking at the median and median absolute deviation (MAD) of the intensities of each frame was enough to discard all the noise images. Our scheme is based on a simple low signal outlier rejection:

Algorithm 11 (Noise Frames Detection)

- For each frame I_k, compute the median and MAD of the image intensities: $\alpha_k = \text{med}(I_k)$, $\beta_k = \text{MAD}(I_k)$
- Compute the median and MAD over time of these quantities: $\alpha = \text{med}(\alpha_k)$, $A = \text{MAD}(\alpha_k)$, $\beta = \text{med}(\beta_k)$, $B = \text{MAD}(\beta_k)$
- For each frame k, if $\alpha_k < \alpha - 4A$ or $\beta_k < \beta - 4B$, we consider it as noise. ■

The second step of our algorithm consists of the registration of the pairs of consecutive frames. If the pair encloses a frame that has been tagged by Algorithm 11 as being a noise only frame, we discard this pair and thus do not need to perform a registration for it. Given a regular candidate pair, we are interested in finding a coarse spatial transformation to align the two images. Therefore, we only look for a translation with a pixel accuracy. This step is efficiently computed through a globally optimal normalized correlation matching [Lewis 95]. This scheme is based on a Fourier transform of the data and a spatial normalization to compute, in one pass, the correlation coefficient for each possible translation. Because of some border effects, the normalized correlation matching of [Lewis 95] is however not inverse consistent in the sense that the registration of I_k with I_{k+1} does not always give the opposite of the translation found by the registration of I_{k+1} with I_k.

In the final step of our scene-splitting scheme, it is necessary to decide whether the translation found by the normalized correlation matching can be accepted or not. A natural approach would be to use a set of different features, such as different image similarity metrics, to describe the pair of aligned images and train a classifier to distinguish acceptable pairs. It is however very difficult to get a meaningful training data set since even by visual inspection of the pairwise registration result, a human expert cannot always assess whether the mosaicing algorithm will work or not with this pair of frames among the sequence to be processed. We have thus chosen a rather simple approach that proved to be efficient. We tested a simple threshold on different intuitive similarity metrics such as the mean square error, the correlation coefficient, the mutual information and so on. We chose a set of Cellvizio sequences that were representative of the data we could encounter during the clinical trial. This data was given to the scene-splitting scheme, using different metrics and different thresholds, before passing the sub-sequences to the mosaicing algorithm. After visual inspection of the computed mosaics, we chose the best combination of metric and threshold. As a results, we found that it was first possible to discard the pairs of images for which the normalized correlation matching of [Lewis 95] provided too little overlap between the frames or provided severely inverse inconsistent results, i.e., where the forward and backward translations disagreed with more than ten pixels. Then, the metric we found to be the most effective turned out to be a form of normalized mean square error (NMSE):

$$\frac{\sum_{p\in\Omega_k}\left(I_k(p)-I_{k+1}(p+\tau)\right)^2}{\sqrt{\sum_{p\in\Omega_k}\left(I_k(p)-\bar{I}_k\right)^2\,.\,\sum_{p\in\Omega_k}\left(I_{k+1}(p+\tau)-\bar{I}_{k+1}\right)^2}}$$

where Ω_k is the region of overlap between $I_k(p)$ and $I_{k+1}(p+\tau)$ and where $\bar{I}_k = \sum_{p\in\Omega_k} I_k(p)/|\Omega_k|$ and $\bar{I}_{k+1} = \sum_{p\in\Omega_k} I_{k+1}(p+\tau)/|\Omega_k|$ are average intensities. With the combination of these two rules, we were able to automatically find smooth motion subsequences that can be processed by our mosaicing algorithm. On our test case of 100 sequences, only one subsequence failed to provide a decent mosaic.

6.3.2 Formatting the Output for an Easier Diagnosis

Once an input Cellvizio sequence has been segmented by our scene-splitting algorithm, a possibly large number of smooth-motion subsequences are available for further mosaicing. Some of these sequences can be composed of a very small number of frames or even a single frame. On other parts where the practitioner almost did not move the optical microprobe, we can have a large number of frames for a very little improvement of the field-of-view after the mosaicing. It would neither be informative nor computationally efficient to mosaic every subsequence generated by the scene splitting-scheme. The major drawback of generating too many mosaics is however that it goes somewhat against our goal of providing an efficient and complete representation of the pertinent information in the input sequence. If the clinician is to trade a careful inspection of a video sequence against a careful inspection of a possibly large amount of images, the gain for him would be minimal.

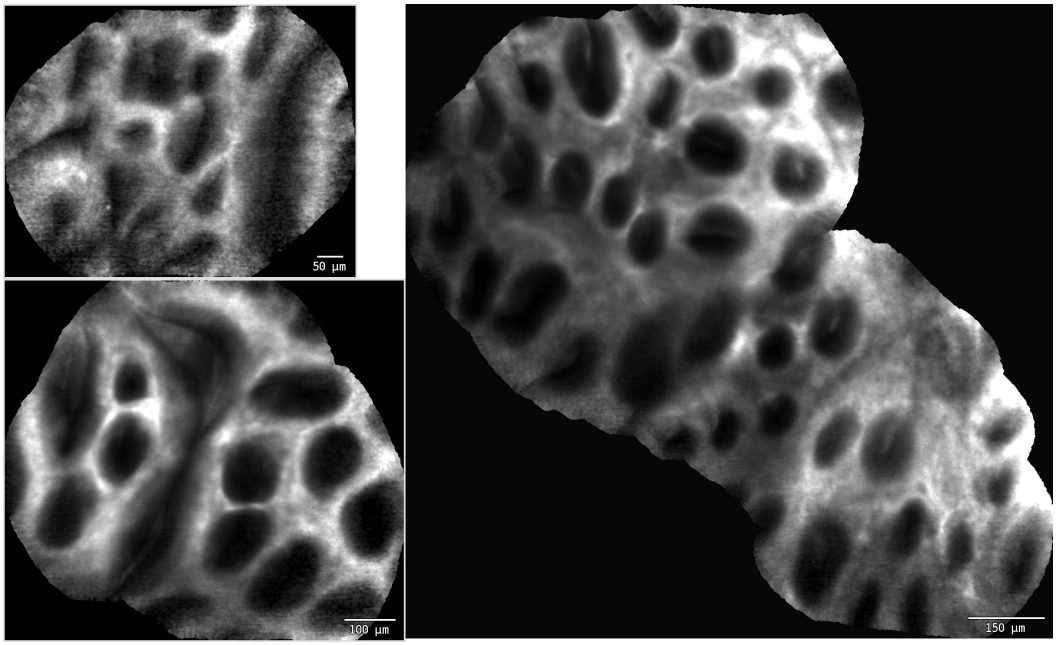


Figure 6.4: One example *montage* output of our mosaicing solution on an acquisition from the clinical trial. The clinician gets the relevant information in one glance.

After a visual examination of the outputs of the scene splitting followed by mosaicing scheme, we realized that most of the important information was naturally contained in the largest mosaics. A discussion with the clinicians led us to only consider the three largest ones. This fact can favorably be used to lower the computational burden of the mosaicing. From the scene-splitting algorithm, we indeed have an estimate of the translation between consecutive frames. This leads to an estimation of the area covered by the mosaic that would be generated from a given input subsequence. We can thus choose the three subsequences that are predicted to provide the largest mosaics and run the mosaicing only on those three scenes.

Compared to making a diagnosis on one image, a careful inspection of three different images is still somewhat tedious for the clinician. We have thus decided to create a *montage* image that simply tiles the three mosaics into one single image. We present one example of such a *montage* in Fig. 6.4. This has the advantage of allowing the clinician to get all the relevant information in one glance. It is thus an important improvement in terms of diagnosis time and cost efficiency.

This type of output format is also very beneficial for quality control issues. As we mentioned previously, our scene-splitting algorithm has largely been tailored by visual inspection of the results on a rather large test sample of 100 input sequences. It was thus necessary to be able to rapidly check for failures of the algorithm, and assess the global quality of the results. These montages allowed us to do it efficiently.

In addition to these montages, we also create a set of supporting files that are mainly devoted to providing verification information in case the user has some

doubts about a mosaic. This set contains movies of the mosaicing process. The consecutive frames are shown sequentially in a reference coordinate system with a position that corresponds to the global spatial transformation used by the mosaicing algorithm.

6.3.3 Distributing the Computation

Because of the size of the clinical trial we work on, the computation time required to process all the data prohibits the use of a single computer running all the jobs. This is all the more true that we have been investigating the use of different parameters on a rather large test sample before actually processing the data for the multicenter clinical trial.

We have therefore implemented a rather simple distributed computing environment based on a data parallelism paradigm. As shown in Fig. 6.5, we use a classical object model. A server has a list of input Cellvizio sequences on which the complete mosaicing pipeline should be run. Several clients provide the computational force. When a client is available, it sends a request for work message to the server. If the server has a non-processed files in his queue, it sends a reply back stating which job the clients needs to perform. To create this computing cluster, we used the standard common object request broker architecture (CORBA).

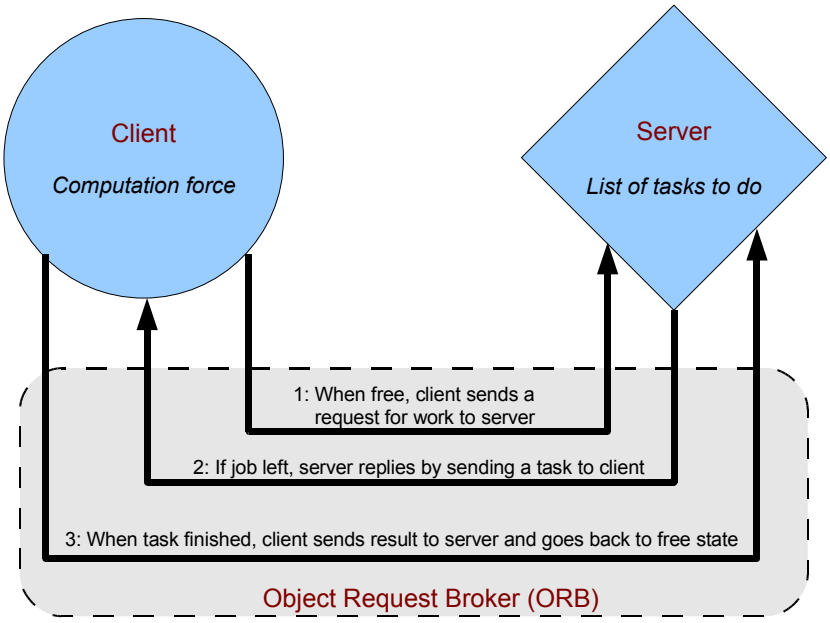


Figure 6.5: Distributed computation model. The server dispatches the files to be processed among a pool of clients.

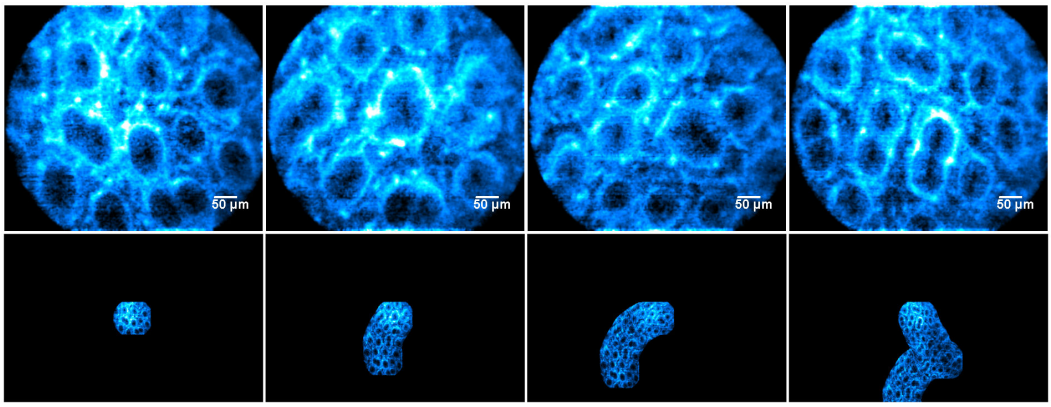


Figure 6.6: The concept of live mosaicing on a sequence of lymphocytic colitis. As time goes by (left to right), the current image (top row) is roughly register and stitched to the current mosaic (bottom row). We therefore have a growing mosaic. Courtesy of PD. Dr. A. Meining, Klinikum rechts der Isar, Munich.

6.4 Real-time User Feedback to Improve the Acquisition

6.4.1 Motivation

In the previous chapters, we have argued that fibered confocal microscopy provided a potential tool for *optical biopsies*. With the aid of our mosaicing software, we have shown that it was possible to get at the same time a microscopic resolution and a wide field-of-view without having to increase the size of the optical microprobe and thus the invasiveness.

What makes the interest of Cellvizio is not only its capability of acquiring microscopic images of tissues *in vivo* but mainly the fact that it allows to do so in real time with a direct visualization by the clinician. It is exactly this that makes it possible to unify, within a single procedure, the disease suspicion made during endoscopy, the actual diagnosis and the treatment.

The mosaics we create with our post-processing scheme are very useful for a careful inspection of the Cellvizio data after the endoscopy and for inclusion in the patient record. An ideal situation would however be to have a mosaic constructed in real time, on the fly during endoscopy as shown in Fig. 6.6. Such a tool would also help the investigator evaluate whether the area intended for examination was adequately targeted. This work was also presented in [Vercauteren 08b].

The computational burden imposed by running the full mosaicing pipeline makes it however impossible to get such a detailed mosaic reconstruction in real time. In [Loewke 07], the authors proposed to use a robot to hold the probe and a set of sensors to get an estimate of the global positioning of the frames. While such an external information allows to ease the computational cost, using a robot will definitely not fit into the current endoscopic procedure. It will also be cumbersome and more invasive as sensor needs to be placed at the tip of the optical microprobe.

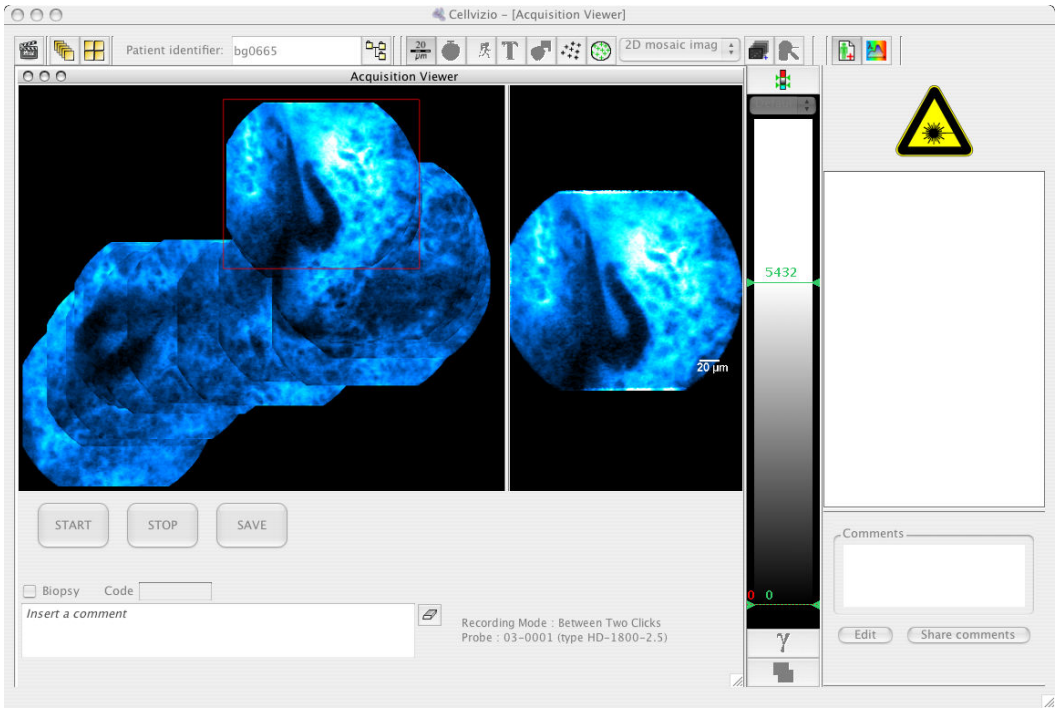


Figure 6.7: The graphical user interface of the live mosaicing. The clinician can see on the same screen, the regular acquisition view on the right and the simple mosaic, stitched on-the-fly, on the left. Courtesy of PD. Dr. A. Meining, Klinikum rechts der Isar, Munich.

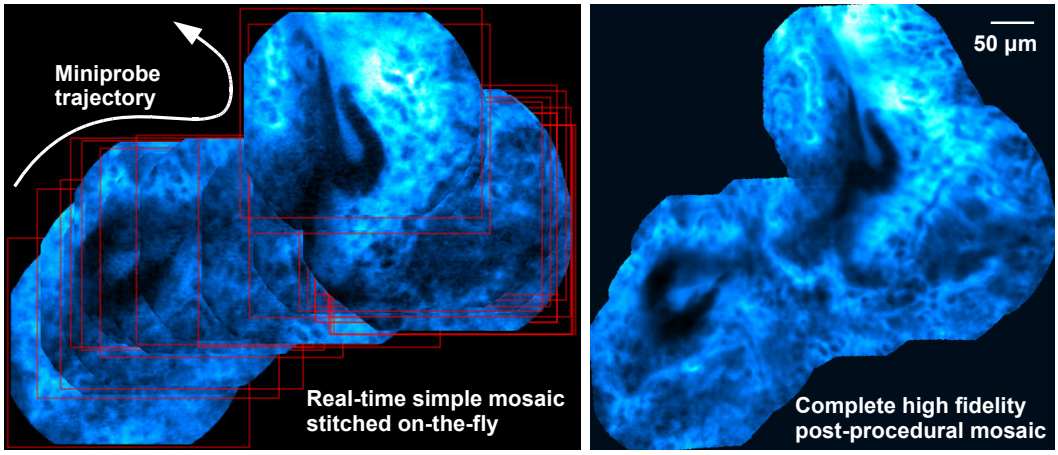


Figure 6.8: Comparison of the real-time mosaic and the post-processing mosaic on a Cellvizio sequence of a severe ulcerative colitis. Courtesy of PD. Dr. A. Meining, Klinikum rechts der Isar, Munich.

Our approach to real-time mosaicing will thus focus on keeping the ease of use of Cellvizio at least at its current state. This implies that we have to resort to using a purely algorithmic solution to the problem of real-time mosaicing. To let this be feasible, we will fallback to less accurate models than the one presented in Chapter 5 and less precise reconstruction schemes.

In addition to the *real-time field-of-view enhancement* target, a major goal of our live mosaicing tool lies in the training of the practitioner. Thanks to the direct visual feedback that the live mosaicing provides, clinicians can assess in real time whether their acquisition is smooth and of good image quality. As shown in Fig. 6.7, the live mosaicing will indeed provide nice-looking mosaics as long as the motion of the optical microprobe is smooth and as long as the image quality is sufficient to register pairs of consecutive images. When it is not the case, the clinician will either see no mosaic being constructed or a rather random stitching of the input images. This is of major interest for the Cellvizio system because without such feedback, it is often difficult for the user to assess the quality of his acquisition. Early qualitative evaluation has shown that the live mosaicing tool has the effect of reducing the learning curve needed to harness Cellvizio. It can even improves the quality of the acquisition for experienced users.

6.4.2 Real-time Algorithm

In Section 6.3.1, we have seen that the fast normalized correlation matching algorithm of [Lewis 95] allowed the registration results to be accurate and robust enough to evaluate the smoothness of the motion of the optical microprobe with respect to the imaged tissue. As a matter of fact, this algorithm also allows us to get a decent alignment of the images in real time. The main idea of the fast normalized correlation, is to evaluate, in one pass, the correlation coefficient between the fixed image F and the translated moving image $M \circ \tau$ for every translation τ with integer components.

This algorithm has been designed for template matching. As such it is theoretically correct when the support of $M \circ \tau$ is included in the support of F. It is not completely accurate to use it when F and M have the same size. In practice however it works also well in this case and shows manageable border effects. Here is a very brief overview of this scheme. The similarity criterion can be written as:

$$\mathrm{Sim}\,(F, M \circ \tau)\, \frac{\sum_p (F(p) - \bar{F})(M(p+\tau) - \bar{M})}{\sqrt{\sum_p (F(p) - \bar{F})^2 \sum_p (M(p+\tau) - \bar{M})^2}} = \gamma(-\tau), \tag{6.1}$$

where $\bar{F}$ is the mean of F and $\bar{M}$ is the mean of M. Let us look at the numerator of (6.1). Let $F'(p) = F(p) - \bar{F}$, $M'(p) = M(p) - \bar{M}$ and $M'_{rev}(p) = M'(-p)$. We see that

$$\gamma^{num}(\tau) = \sum_p F'(p) M'_{rev}(\tau - p) \tag{6.2}$$

is the convolution of the normalized fixed image F' with the reversed normalized moving image M'_{rev}. This can efficiently be computed with the Fourier transform

$\mathcal{F}$:

$$\gamma^{num}(\tau) = \mathcal{F}^{-1}\big(\mathcal{F}(F')\,\mathcal{F}(M'_{rev})\big) = \mathcal{F}^{-1}\big(\mathcal{F}(F')\,\mathcal{F}^{*}(M')\big),$$

where we used the convolution theorem and the fact that for a real signal, time reversal of the signal is accomplished through complex conjugate of the Fourier transform. Furthermore, as shown in [Lewis 95], running sums that compute the integral of the image intensity and the integral of the squared image intensity are used to compute the denominator. Once the full correlation coefficient map has been computed, we only need to find its maximum to get the optimal translation.

Besides its fast computation time and global optimality properties, this algorithm has the very nice property of not requiring any gradient-descent like loop. This is a very important property for real-time algorithms that needs to terminate their computations before the deadline of the next event. If a gradient-descent like scheme were to be used, the only way we could enforce the worst-case execution time would be to limit the number of iterations. This implies that it would become possible not to reach convergence.

Given that using the fast normalized correlation matching, we have a way to register in real time the consecutive frames of the acquired Cellvizio sequence, what we now need to do is to display a mosaic based on these alignments. It would be possible to use a mosaic reconstruction scheme such as the one presented in Chapter 5. However, this would be computationally expensive and the computation time would depend on the size of the imaged area. This implies a non-deterministic and growing computation time that is incompatible with real-time processing. Therefore, we chose to use a simple *dead leaves model* where the current frame is simply overlaid on top of the previous frames at a position dictated by the registration. In addition to requiring very little computational time, this approach has the advantage on not being too sensitive to registration errors. As we do not mix the information of several images, a small misregistration will not lead to a blurred image but rather to an image where seams are still visible at the edge of the images, as shown in Fig. 6.7. If an undetected gross registration error appears, the good thing is that we still visualize correctly the information of the current frame.

Thanks to this visualization mode, large alignment errors can be tolerated. However, these errors can be tiring for the clinicians since the images can be placed at somewhat random locations. This will cause a kind of visual flicker. It would thus be advantageous to detect gross registration errors and wipe-out the history when such a discontinuity is detected. Again, the scheme we use needs to run in real time and should thus use as little processing as necessary. From the fast normalized correlation matching, we trivially have access to the correlation coefficient between the images for the optimum translation. We found that, for the real-time algorithm, a simple threshold on this correlation coefficient was sufficient to provide decent visual comfort to the clinician.

Table 6.1: A qualitative evaluation of the usefulness of the live mosaicing during GI endoscopy. Three different investigators have been using Cellvizio for some time and have then been given the opportunity to test live mosaicing. They answered the following questions after at least several tens of Cellvizio procedures with live mosaicing. The answers written in the table are direct quotes from the investigators.

	A	B	C
How often do you use the live mosaicing mode with respect to the classical acquisition mode?	Every patient	Now 100%	Every patient
When using the live mosaicing mode, how often do you find the real-time mosaic to be informative?	80%	100%	75%
Are there some protocols for which the live mosaicing does not bring any added value?	No	No	I think it is always an advantage. It provides an idea, how stable you are and how far you slip with the probe. It also gives you an idea on how to cover more than just one spot in case the area is very stable.
When using the live mosaicing mode, how often do you look at the mosaic window with respect to the movie only window?	Stable picture: no need to look at mosaicing. Movement: look at mosaicing window only.	80–90%	50%

6.4.3 Evaluation of Live Mosaicing

As shown in Fig. 6.8, even in simple mosaicing problems, the output of the live mosaicing naturally does not matches the output of the post-processing mosaicing algorithm in terms of image quality, image details and signal to noise ratio. It however still does a pretty good job for a real-time algorithm.

Once again, this type of algorithm is very difficult to validate as we do not have access to a ground truth data that we could use to compare to. In Chapter 5, we used the fact that our post-processing mosaicing algorithm also provides an estimation of the position of the optical microprobe. We compared the trajectory found by our algorithm with the path imposed by a high precision computer numerical control robot. We could use the same approach to evaluate the real-time algorithm but it would not be that meaningful. What we actually need is indeed to provide feedback to the clinician on the stability of the motion, and show him a rough mosaic. This is not directly related to how accurately the motion can be estimated.

In order to judge the clinical relevance of our live mosaicing tool, we have thus decided to perform a qualitative evaluation based on clinical expertise. The three investigators involved in the clinical trial had been working with Cellvizio for some time before we gave them the live mosaicing software shown in Fig. 6.7. After a short training period, the investigators were given the opportunity to use the live mosaicing or not. After a few tens of patients each, we asked them to rate the usefulness of our tool based on a few questions that are listed in Table 6.1. It can be seen from the results that all the investigators chose to keep using the live mosaicing during the acquisition. This fact only should support the clinical relevance of the scheme. With a subjective perspective, we also found that the data we received from the investigators became better suited for the post-processing mosaicing solution.

6.5 Clinical Trial Outcome

This section is still work in progress. The acquisition of the data by the clinical investigators is planned to be finished in early 2008. Even if partial data has already been acquired and evaluated by histopathology, it is not possible to provide a partial measure of the clinical trial outcome. This would indeed require us to infringe the anonymization and randomization of the data before the end of the clinical trial. Obviously this would be unacceptable.

To get an early qualitative feeling of the performance of our mosaicing solution, we present in Fig. 6.9 and Fig. 6.10 the result of our tool on 24 random Cellvizio sequences taken from the sequences we have already received and processed. The final quantitative results are expected for mid-2008.

6.6 Discussion

The importance of validation of biomedical image processing methods is now well recognized [Jannin 02]. It brings to light the important characteristics of algorithmic systems for the clinicians and helps them evaluate their potential benefits and shortcomings. Without a thorough clinical evaluation of a given biomedical image processing scheme, it is impossible to estimate the value of a new development in terms of clinical value or economic repercussions.

In this chapter, we have focused on the critical public health question of early cancer detection for patients suspected of having Barrett's esophagus. We proposed to enhance the potential benefits of using fibered confocal microscopy for Barrett's esophagus screening by using adapted video mosaicing techniques. Cellvizio may eventually allow the clinician to make a real-time evaluation of the malignancy of the tissues. It should also improve the specificity and sensitivity of the diagnosis which is currently based on a somewhat random biopsy protocol. The mosaicing tools we proposed should help the clinician go one step further by giving him more confidence in his findings and by allowing him to see all the relevant information of the Cellvizio data in one single glance.

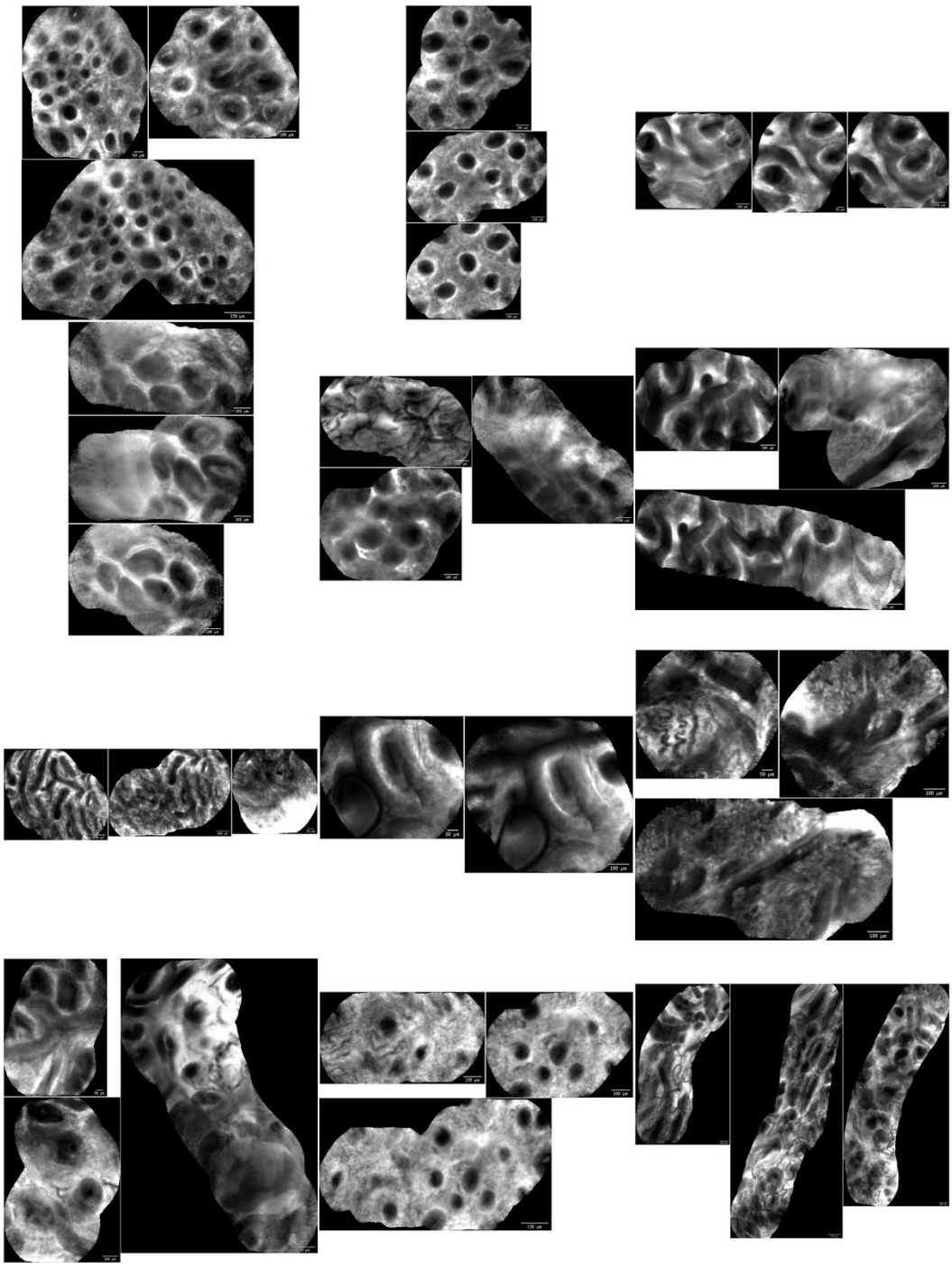


Figure 6.9: Mosaicing results on 12 (out of more than 700) random sequences from the clinical trial. Note that if the complete sequence can be splitted in less than three scenes, the montage will obviously include less than three mosaics.

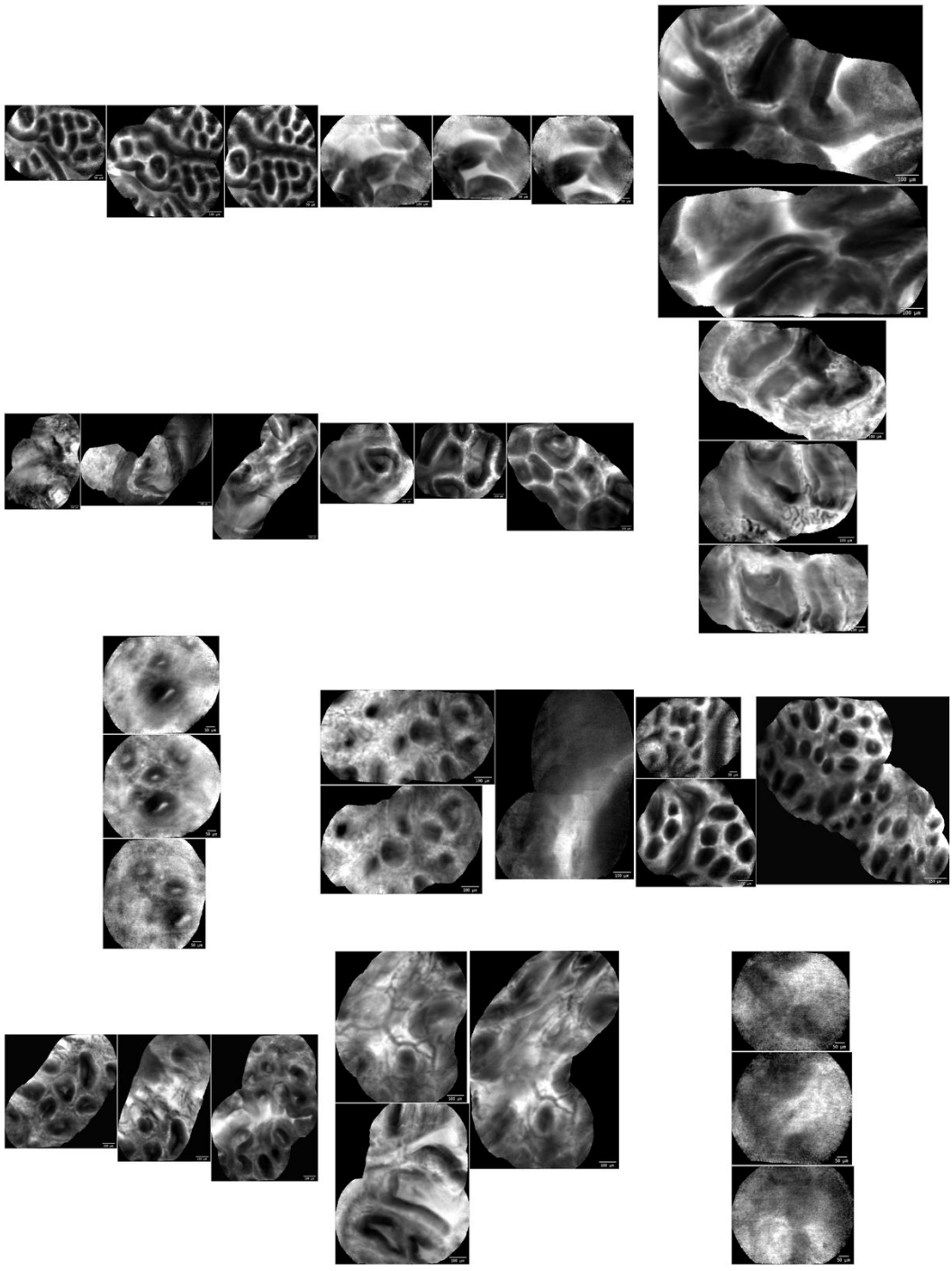


Figure 6.10: Mosaicing results on 12 other (out of more than 700) random sequences from the clinical trial.

We have also seen in this chapter that moving a given algorithm from bench to bedside is far from being straightforward. A great deal of effort was put on engineering a complete mosaicing solution that could fit in the current routine medical protocol. We have developed a scene-splitting algorithm that finds the smooth-motion parts of a Cellvizio sequence. A simple distributed computing architecture has been built to deal with the computational requirements. A collaboration with the clinicians led us to design a user-friendly output format for our results. Finally, a live mosaicing scheme was proposed to provide direct visual feedback to the investigators and help them improve their acquisitions. Qualitative evaluations showed that our tools have proved to meet the clinicians demands.

Chapter 7

Conclusions

Table of Contents

We have seen in this thesis that fibered confocal microscopy provides the ability to image biological tissues *in vivo* and *in situ* with a spatio-temporal resolution which is sufficient to observe life at the cellular level as it happens. FCM, and especially Cellvizio by Mauna Kea Technologies, is thus a potential tool for clinicians and biologists to conduct dynamic optical biopsies.

Our goal throughout the manuscript has been to show how advanced image processing and image registration schemes could help move beyond some hardware limitations of the imaging device. We will now briefly review the contributions we presented in the course of this manuscript and will conclude by proposing some possible extensions of our work.

7.1 Contributions

7.1.1 Methodological Contributions

As with many theses in the field of biomedical image analysis, the most important contributions in terms of research are the ones that can be transferred to solve other problems. In this section, we focus on the theoretical aspects of the work we presented in this manuscript.

Our first significant methodological contribution, presented in Chapter 3 and [Vercauteren 07b], illustrates the importance of moving the findings in one given field of study to another. In this chapter, we indeed showed that the ESM framework of [Benhimane 04, Malis 04] could advantageously be transposed from vision-based

robot control to biomedical image registration. The emphasis was set to provide a sound mathematical treatment of the ESM framework with as much insight as we could to help us generalize it. We showed that the key to this scheme was to take into account the specificity of the image registration problem whereas, in other works, registration is often treated as a generic optimization case.

One important thing to notice is that some information about the optimal image alignment between two images can be gathered without knowing what the optimal alignment is. Indeed, especially in the mono-modal case, given the optimal alignment, we know that the warped moving image should be very close to the fixed image. From this information, we can compute the derivative of the similarity criterion at the optimum and boost the convergence rate of image registration schemes. In Chapter 3, this information was used for linear image registration while in Chapter 4, it allowed us to provide interesting theoretical roots to an efficient variant of Thirion's demons algorithm [Vercauteren 07b]. To the best of our knowledge, this is the first theoretical explanation of this symmetric forces variant.

We have highlighted the fact that spatial transformation spaces can also provide additional information to the optimization schemes. In most registration cases, we do not deal with classical vector spaces but rather with Lie groups. We showed throughout this thesis that the geometric structure of these Lie groups could be used in a very principled manner and that the additional computational burden was far from being problematic. We have seen that optimization on a Lie group could be achieved by using an unconstrained optimizer on the Lie algebra and projecting the update step back to the Lie group through the exponential map ($s \leftarrow s \circ \exp(\boldsymbol{u})$). This is to contrast with a classical additive update rule ($s \leftarrow s + \boldsymbol{u}$) that requires a constrained optimizer to respect the geometry of the Lie group. Following the ESM framework, such a geometric optimization was used in Chapter 3 and [Vercauteren 07b] to take into account the constraints of linear spatial transformation groups. In Chapter 4 and [Vercauteren 07d], we proposed the first use of this geometric optimization framework in the context of non-rigid non-parametric image registration. This led us to our fast diffeomorphic image registration algorithm based on Thirion's demons.

Based on similar ideas and borrowing the Lie group point of view on spatial transformations, we proposed in Chapter 5 and [Vercauteren 05, Vercauteren 06], a robust algorithm that uses local pairwise image registration results to compute a global, groupwise, multi-image alignment. Our formulation uses a Riemannian distance between spatial transformations and a least-square approach to get a globally consistent alignment of the input frames. This is close to a Fréchet mean approach. An independent work showed that our approach also provides interesting results for the problem of 3D ultrasound mosaicing [Wachinger 07].

In our mosaicing algorithm, we pushed the modeling of the spatial transformations one step further. Since fibered confocal microscopy images are distorted by motion artifacts, we needed to estimate the velocity of the optical microprobe. We proposed a hierarchical approach that iterates between using finite differences to estimate velocities based on position information and finding position information

based on registration of motion-compensated input frames [Vercauteren 06]. An additional important contribution in this context is the extensive technical validation we performed. By comparing the path of the optical microprobe with the one imposed by a robot, we have been able to evaluate the accuracy of our schemes with respect to some ground truth data.

7.1.2 Contributions in Terms of Applications

The field of biomedical image analysis is at the same time a technology-driven and an application-oriented research field. As such, applications have played a crucial role in this thesis. They first represent challenges to solve and, once a potential solution is found, the final goal is always to perform an applicative validation.

In Chapter 2, a fair comparison of different image reconstruction schemes allowed us to choose, based on grounded basis, the algorithm that best meets the requirements of fibered confocal microscopy.

If image registration has mostly been treated as a building block for our mosaicing algorithm, we have also shown another important application of it in Chapter 3. The efficient linear image registration scheme of Chapter 3 was coupled with vessel segmentation and kinetic analysis. This led us to a region tracking algorithm that enables the measurement of cell trafficking in microvessels [Perchant 07]. To the best of our knowledge, there had been no previous attempt to perform blood flow velocity measurement on a sequence acquired *in vivo* with global tissue motion.

In Chapter 5 and [Vercauteren 06, Thiberville 07, Becker 07], we presented how the mosaic images we create can help the clinician or biologist to get an efficient and complete representation of a Cellvizio sequence. The actual application of mosaicing was, however, only fully presented in Chapter 6 where we included an evaluation of our algorithm in a multicenter clinical trial. Our goal was to integrate all facets of clinical constraints to propose a complete mosaicing solution usable by medical investigators. This also led us to develop a real-time mosaicing algorithm that serves as a confidence measure and quality feedback for the clinicians. We have seen that, thanks to this visual feedback, the quality of Cellvizio acquisitions improves even for experienced practitioners.

7.1.3 Software Contributions

For the customers of Mauna Kea Technologies, the most important contributions of this thesis are the ones they can use, meaning those that are implemented and are available to clinicians and biologists. We have therefore implemented most of the algorithms presented here within Mauna Kea Technologies proprietary softwares. The real-time image reconstruction scheme we chose and implemented is now used for every single image seen by a Cellvizio user. The post-processing mosaicing algorithm and the associated image registration schemes are planned to be released in 2008 but are already used on a daily basis for the clinical trial of Chapter 6. Similarly, the real-time mosaicing algorithm is being used on a daily basis by the

clinical investigators of the multicenter trial and should be released in early 2008.

Software is not only important for end-users but also for other researchers in our field. As pointed out in [Kovačević 06, Yoo 05], real scholarship in computational sciences should not only include research articles but also software and data to reproduce the results. If such *reproducible research* is often difficult in practice, it is all the more true for biomedical imaging where the data can often not be made accessible. It was not the main goal of this thesis to definitely move towards reproducible research. However, we thought that it would be very interesting to provide access to a reference implementation of our most important methodological contribution. An open-source implementation of our diffeomorphic demons algorithm using the Insight Toolkit can thus be found in [Vercauteren 07c]. Our diffeomorphic demons implementation has also been integrated into MedINRIA, the free medical image navigation and research tool of Asclepios research group, INRIA Sophia Antipolis [Toussaint 07]. It is thus easily available to clinicians of different medical imaging specialties. It is, for example, already used within the European project IST-2004-027749 Health-e-Child.

7.2 Perspectives

Many short term research directions have already been presented in the conclusion of the previous chapters. In this section, we summarize the most important ones and present some potential directions for long-term research.

7.2.1 Incorporating More Prior Knowledge

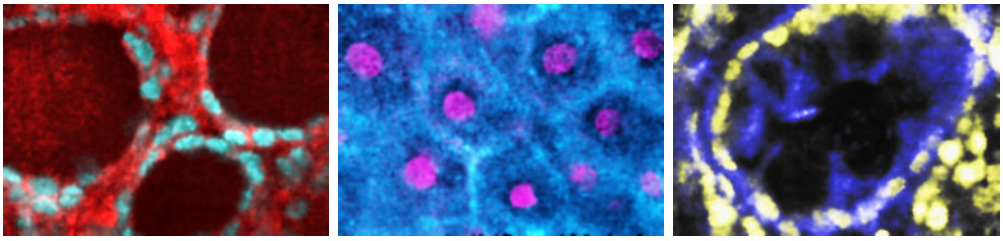
The first direction we propose is to use even more prior knowledge about the physics of acquisition. We have, for example, seen in Chapter 5 that mosaicing allowed us, to some extent, to achieve some super-resolution. However, our mosaic reconstruction scheme does not take advantage of the information resulting from the optics and the imaging device. To get a true super-resolution effect, we would need a reconstruction algorithm that takes into account the point spread function (PSF) of the system. The problem is that each fiber optic in the bundle may have a different PSF. It might thus be worth looking at the frames theory that allows us to model a system with a spatially varying PSF. We refer the interested reader to [Mallat 99] for an overview of the frames theory. This approach would also require an accurate method to estimate the PSF of each fiber. Such a study would also open the way for the deconvolution of Cellvizio images in cases where the optical properties and the sampling distance do not fully meet the sampling theorem criterion. For classical *benchtop* confocal microscopy, deconvolution is already an active field of research, see e.g. [Dey 06]. If the PSF of the system can not be assessed before imaging, one could also rely on blind or semi-blind, i.e. parametric, deconvolution algorithms [Pankajakshan 07].

From the applicative point of view, we have seen that our mosaicing tools can be used for translational research. We have shown mosaics for small animal imaging

and clinical research. An additional step would be to use our tools within an actual clinical routine workflow. Furthermore, it would be interesting to adapt the mosaicing solution we developed for the multicenter clinical trial to each specific application. We could take into account the requirements of the applications, and the prior knowledge we have, to design custom mosaicing solutions.

7.2.2 Introducing Other Modalities

With a growing number of target-specific fluorescent dyes and new bio-engineered mouse models available for the biologists, benchtop confocal microscope manufacturers have developed a new generation of benchtop microscopes. They feature multiple excitation wavelengths and detection channels, and are thus capable of measuring simultaneously the distribution of several fluorescent markers. In [Jean 07], the authors proposed a multispectral fibered confocal microscope that potentially allows for multilabeling studies *in vivo*. Some typical images appear in Fig. 7.1. We have shown that our mosaicing algorithm worked with both fluorescence and reflectance fibered confocal microscopy on images acquired on small animals and on humans. It would still be a great challenge to adapt the algorithms presented in this thesis to the video sequences acquired with a multispectral Cellvizio. Each spectral band could potentially present different motion artifacts as the acquisition of the different bands need not be simultaneous.



(a) Thyroid stained with To-PRO-3 and DiA

(b) Cervix stained with YOYO-1 and DiD

(c) Colon stained with To-PRO-3 and DiA

Figure 7.1: Multilabeled human tissues imaged *ex vivo* using a multicolor fibered confocal microscope prototype developed by Mauna Kea Technologies.

As far as other modalities are concerned, we have mentioned that mosaicing can help bridge the gap between macroscopic imaging devices such as MRI and PET and microscopic modalities such as fibered confocal microscopy. Figure 7.2 shows an early proof of concept of this potentiality. Figure 7.3 displays a clinical setup where the clinician acquires at the same time, in the small bowel, endoscopic images with a double balloon enteroscope and microscopic images with a Cellvizio under X-ray visual control [Miehlke 07]. Future research should aim at developing this potential. We will need to register, or at least colocalize when the imaging scales are still too different, images from very different modalities. A potential solution would be to integrate the Cellvizio and our mosaics within a macroscopic image tracking and computer-aided diagnosis tool such as the one presented in

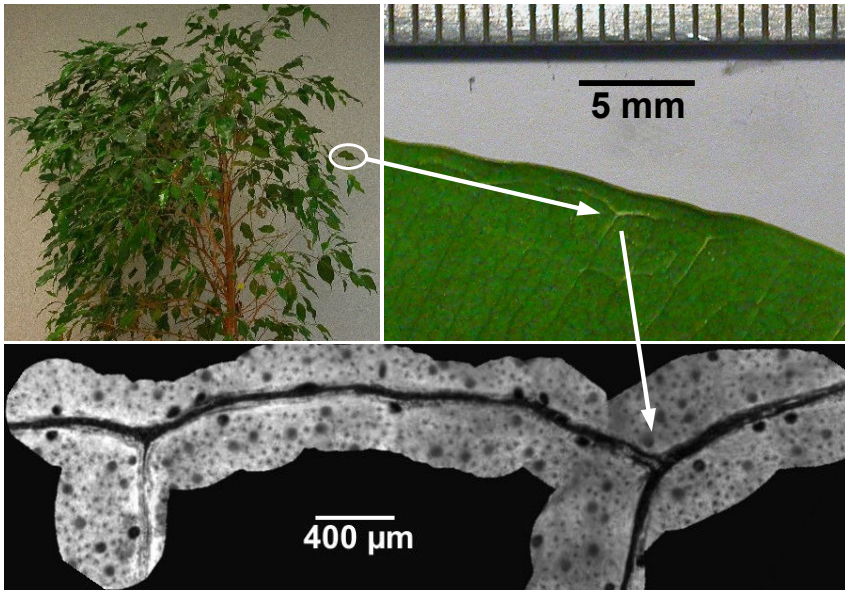


Figure 7.2: Bridging the gap between macroscopic and microscopic modalities: a proof of concept. Top left: *Ficus Benjamina*. Top right: A particular leaf showing a vein triple point. Bottom: Autofluorescence fibered confocal microscopy mosaic.

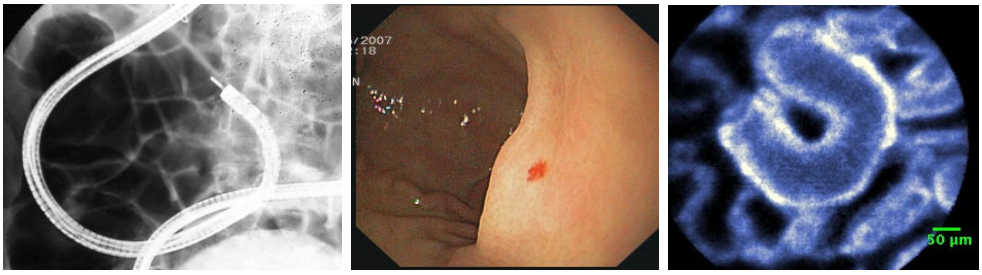


Figure 7.3: Double balloon enteroscopy combined with Cellvizio imaging under X-ray control enables multimodal visualization of the small bowell [Miehlke 07]. Courtesy of Pr. Dr. Stephan Miehlke, Dresden University of Technology.

[Glocker 07]. This problem will also require 3D to 2D image registration algorithms such as the ones presented in [Groher 06].

7.2.3 Registration

From the efficient registration algorithms we presented in this thesis, several possible perspectives emerge. Let us look at the spatial transformations first. For some specific applications such as brain research, incorporating prior knowledge would be beneficial to drive the non-rigid registration algorithm. More advanced regularization terms could, for example, be used in the cost function. One possible extension would be to use our diffeomorphic update rule in combination with efficient schemes to solve the linear equations that arise from the linearization and discretization of Euler-Lagrange equations associated with the image registration

problem [Cahill 07].

We have seen only a small difference between the compositive demons and the diffeomorphic demons. This leads to the conclusion that the current limit of our non-rigid registration algorithm might not really be the update rule but rather the representation of the spatial transformation. It might be advantageous to see how our efficient scheme could be adapted to use a particle representation for diffeomorphisms such as in [Marsland 07].

Another direction of research would be to incorporate multimodal similarity criterion within our scheme. A sensible way of doing it could be to divide the first step of the demons, intensity matching, into an EM approach to multimodal similarity such as in [Zöllei 07]. Another approach would be to directly look for the flow of multimodal similarity metrics such as in [Hermosillo 01, Hermosillo 02]. We have also started working with Boon Thye Thomas Yeo, MIT, on extending our tools to diffusion-weighted images whose registration is still a challenge [Faugeras 07].

7.2.4 Mosaic Analysis

Finally, a natural continuation to our work would be to use our mosaicing algorithm as a preprocessing step to feed advanced quantitative tools. The cell trafficking analysis algorithm we presented in Chapter 3 can be seen as a proof of concept for using registered video sequences as an input to quantitative analysis. Obviously, more work can be done.

An important contribution would be to design some computer-aided diagnosis tools to classify the different types of tissue encountered in a specific disease surveillance protocol such as the Barrett esophagus one we presented in Chapter 3. Some endoscopists have indeed only a basic histopathology background. It is not always easy for them to interpret the microscopic data that Cellvizio provides. Computer-aided diagnosis could help them reduce the learning curve needed to make a confident diagnosis. Potential solutions for such classifiers might rely on visual features extraction, coupled with trained classifiers such as in [Golland 03]. We could also use shape extraction paired with a stochastic shape analysis such as in [Golland 05, Stoyan 96].

As we can see from these different potential extensions, dynamic *in vivo* and *in situ* microscopy uncovers several opportunities for future research. We have highlighted a number of open problems ranging from theoretical to very applied ones. We are convinced that combining these imaging devices with advanced image processing schemes may decisively transform the field of biomedical imaging.

Bibliography

[Aldroubi 01] Akram Aldroubi and Karlheinz Gröchenig. *Nonuniform sampling and reconstruction in shift-invariant spaces*. SIAM Review, volume 43, number 4, pages 585–620, December 2001. Cited on page(s) 27.

[Amidror 02] Isaac Amidror. *Scattered data interpolation methods for electronic imaging systems: A survey*. Journal of Electronic Imaging, volume 11, number 2, pages 157–176, April 2002. Cited on page(s) 26, 96.

[Arigovindan 05] Muthuvel Arigovindan, Michael Sühling, Patrick Hunziker and Michael Unser. *Variational image reconstruction from arbitrarily spaced samples: A fast multiresolution spline solution*. IEEE Transactions on Image Processing, volume 14, number 4, pages 450–460, April 2005. Cited on page(s) 27.

[Arsigny 06a] Vincent Arsigny, Olivier Commowick, Xavier Pennec and Nicholas Ayache. *A Log-Euclidean framework for statistics on diffeomorphisms*. In Rasmus Larsen, Mads Nielsen and Jon Sporring, editors, Proceedings of the 9th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'06), volume 4190 of *Lecture Notes in Computer Science*, pages 924–931. Springer-Verlag, 2006. Cited on page(s) 53, 64, 65, 73, 74.

[Arsigny 06b] Vincent Arsigny, Xavier Pennec and Nicholas Ayache. *Bi-invariant means in Lie groups. application to left-invariant polyaffine transformations*. Research Report 5885, INRIA Sophia-Antipolis, April 2006. Cited on page(s) 40, 85.

[Ashburner 07] John Ashburner. *A fast diffeomorphic image registration algorithm*. Neuroimage, volume 38, number 1, pages 95–113, October 2007. Cited on page(s) 72, 73, 74.

[Aubert-Broche 06] Berengère Aubert-Broche, Mark Griffin, G. Bruce Pike, Alan C. Evans and D. Louis Collins. *Twenty new digital brain phantoms for creation of validation image data bases*. IEEE Transactions on Medical Imaging, volume 25, number 11, pages 1410–1416, November 2006. Cited on page(s) 61, 63, 71.

[Ayache 06] Nicholas Ayache, Tom Vercauteren, Grégoire Malandain, Fabien Oberrietter, Nicolas Savoire and Aymeric Perchant. *Processing and*

mosaicing of fibered confocal images. In MICCAI Workshop on Microscopic Image Analysis with Applications in Biology (MIAAB'06), October 2006. Invited. Cited on page(s) 4, 6, 9.

[Barron 94] John L. Barron, David J. Fleet and Steven S. Beauchemin. *Performance of optical flow techniques*. International Journal of Computer Vision, volume 12, number 1, pages 43–77, February 1994. Cited on page(s) 53.

[Bazin 07] Pierre-Louis Bazin, Lotta Maria Ellingsen and Dzung L. Pham. *Digital homeomorphisms in deformable registration*. In Nico Karssemeijer and Boudewijn P. F. Lelieveldt, editors, Proceedings of Information Processing in Medical Imaging (IPMI'07), volume 4584 of *Lecture Notes in Computer Science*, pages 211–222, Kerkrade, The Netherlands, July 2007. Springer-Verlag. Cited on page(s) 72.

[Becker 07] Valentin Becker, Tom Vercauteren, Claus Hann von Weyern, Christian Prinz, Roland M. Schmid and Alexander Meining. *High resolution miniprobe-based confocal microscopy in combination with video-mosaicing*. Gastrointestinal Endoscopy, volume 66, number 5, pages 1001–1007, November 2007. Cited on page(s) 5, 7, 80, 111, 137.

[Beg 05] M. Faisal Beg, Michael I. Miller, Alain Trouvé and Laurent Younes. *Computing large deformation metric mappings via geodesic flows of diffeomorphisms*. International Journal of Computer Vision, volume 61, number 2, pages 139–157, February 2005. Cited on page(s) 53, 64, 65, 75.

[Benhimane 04] Selim Benhimane and Ezio Malis. *Real-time image-based tracking of planes using efficient second-order minimization*. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots Systems (IROS'04), pages 943–948, Sendai, Japan, September 2004. Cited on page(s) 32, 33, 37, 53, 54, 65, 135.

[Benhimane 06] Selim Benhimane. *Vers une approche unifiée pour le suivi temps réel et l'asservissement visuel*. PhD thesis, Ecole des Mines de Paris, Paris, France, December 2006. Cited on page(s) 32, 38.

[Benhimane 07] Selim Benhimane and Ezio Malis. *Homography-based 2d visual tracking and servoing*. International Journal of Robotics Research, volume 26, number 7, pages 661–676, July 2007. Cited on page(s) 40, 41.

[Bernard 99] Christophe Bernard. *Wavelets and ill-posed problems: Optic flow estimation and scattered data interpolation*. PhD thesis, Ecole Polytechnique, Palaiseau, France, November 1999. Cited on page(s) 27.

[Boissonnat 02] Jean-Daniel Boissonnat and Frédéric Cazals. *Smooth surface reconstruction via natural neighbour interpolation of distance functions*. Computational Geometry, volume 22, number 1–3, pages 185–203, May 2002. Cited on page(s) 26.

[Boreczky 98] John S. Boreczky and Lynn D. Wilcox. *A hidden Markov model framework for video segmentation using audio and image features*. In Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'98), volume 6, pages 3741–3744, May 1998. Cited on page(s) 120.

[Bourg-Heckly 00] Geneviève Bourg-Heckly, Jocelyne Blais, Juan J. Padilla, Olivier Bourdon, Jacques Etienne, Françis Guillemin and Lionel Lafay. *Endoscopic ultraviolet-induced autofluorescence spectroscopy of the esophagus : Tissue characterization and potential for early cancer diagnosis*. Endoscopy, volume 32, number 10, pages 756–765, 2000. Cited on page(s) 14.

[Bourgeais 05] Laurence Bourgeais, Régis C. Lambert, Caroline Custody, Danièle Paupardin-Trisch, Nathalie Leresche and Pierre Vincent. *In vivo calcium imaging in the thalamus*. In Proceedings of the 35th meeting of the Society for Neuroscience (SFN'05), Washington, DC, November 2005. Cited on page(s) 30.

[Bouthemy 99] Patrick Bouthemy, Marc Gelgon and Fabrice Ganansia. *A unified approach to shot change detection and camera motion characterization*. IEEE Transactions on Circuits and Systems for Video Technology, volume 9, number 7, pages 1030–1044, October 1999. Cited on page(s) 120.

[Bro-Nielsen 96] Morten Bro-Nielsen and Claus Gramkov. *Fast fluid registration of medical images*. In Proceedings of the 4th International Conference on Visualization in Biomedical Computing (VBC'96), pages 267–276, September 1996. Cited on page(s) 53, 55.

[Brown 03] Matthew Brown and David G. Lowe. *Recognising panoramas*. In Proceedings of the 9th International Conference on Computer Vision (ICCV'03), pages 1218–1225, October 2003. Cited on page(s) 87.

[Burt 83] Peter J. Burt and Edward H. Adelson. *A multiresolution spline with application to image mosaics*. ACM Transactions on Graphics, volume 2, number 4, pages 217–236, 1983. Cited on page(s) 109.

[Cachier 03] Pascal Cachier, Eric Bardinet, Didier Dormont, Xavier Pennec and Nicholas Ayache. *Iconic feature based nonrigid registration: The PASHA algorithm*. Computer vision and image understanding, volume 89, number 2–3, pages 272–298, February 2003. Cited on page(s) 53, 54, 55, 56, 95.

[Cachier 04] Pascal Cachier and Nicholas Ayache. *Isotropic energies, filters and splines for vectorial regularization.* Journal of Mathematical Imaging and Vision, volume 20, number 3, pages 251–265, May 2004. Cited on page(s) 55.

[Cahill 07] Nathan D. Cahill, J. Alison Noble and David J. Hawkes. *Fourier methods for nonparametric image registration.* In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'07), pages 1–8, June 2007. Cited on page(s) 141.

[Can 04] Ali Can, Charles V. Stewart, Badrinath Roysam and Howard L. Tanenbaum. *A feature-based technique for joint linear estimation of high-order image-to-mosaic transformations: Mosaicing the curved human retina.* IEEE Transactions on Pattern Analysis and Machine Intelligence, volume 24, number 3, pages 412–419, March 2004. Cited on page(s) 87.

[Cavé 05] Charlotte Cavé, Mariette Fourmeaux du Sartel, Anne Osdoit, Benjamin Abrat, Sacha Loiseau, Danijela Vignjevic, Sylvie Robine and Daniel Louvard. *Fibered confocal fluorescence microscopy: a new tool to perform colonoscopy in mice.* In Proceedings of the Digestive Disease Week (DDW'05), Chicago, IL, May 2005. Poster. Cited on page(s) 104.

[Chefd'hotel 02] Christophe Chefd'hotel, Gerardo Hermosillo and Olivier Faugeras. *Flows of diffeomorphisms for multimodal image registration.* In Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'02), pages 753–756, 2002. Cited on page(s) 64.

[Cherrie 02] Jon B. Cherrie, Richard K. Beatson and Garry N. Newsam. *Fast evaluation of radial basis functions: Methods for generalized multiquadrics in* $\mathbb{R}^n$. SIAM Journal on Scientific Computing, volume 23, number 5, pages 1549–1571, 2002. Cited on page(s) 27.

[Christensen 96] Gary E. Christensen, Richard D. Rabitt and Michael I. Miller. *Deformable templates using large deformation kinematics.* IEEE Transactions on Image Processing, volume 5, number 10, pages 1435–1447, October 1996. Cited on page(s) 64.

[Christensen 97] Gary E. Christensen, Sarang C. Joshi and Michael I. Miller. *Volumetric transformation of brain anatomy.* IEEE Transactions on Medical Imaging, volume 16, number 6, pages 864–877, December 1997. Cited on page(s) 53.

[Christensen 06] Gary E. Christensen, Xiujuan Geng, Jon G. Kuhl, Joel Bruss, Thomas J. Grabowski, Imran A. Pirwani, Michael W. Vannier,

John S. Allen and Hanna Damasio. *Introduction to the non-rigid image registration evaluation project (NIREP).* In Josien P.W. Pluim, Boštjan Likar and Frans A. Gerritsen, editors, Proceedings of the 3rd International Workshop on Biomedical Image Registration (WBIR'06), volume 4057/2006 of *Lecture Notes in Computer Science*, pages 128–135, Utrecht, The Netherlands, July 2006. Springer. Cited on page(s) 71.

[Chui 03] Haili Chui, Lawrence Win, Robert Schultz, James S. Duncan and Anand Rangarajan. *A unified non-rigid feature registration method for brain mapping.* Medical Image Analysis, volume 7, number 2, pages 113–130, June 2003. Cited on page(s) 88.

[Davis 98] James Davis. *Mosaics of scenes with moving objects.* In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'98), pages 354–360, 1998. Cited on page(s) 80, 87, 88.

[Deriche 93] Rachid Deriche. *Recursively implementing the Gaussian and its derivatives.* Research Report 1893, INRIA Sophia-Antipolis, 1993. Cited on page(s) 97.

[Dey 06] Nicolas Dey, Laure Blanc-Féraud, Christophe Zimmer, Pascal Roux, Zvi Kam, Jean-Christophe Olivo-Marin and Josiane Zerubia. *Richardson-Lucy algorithm with total variation regularization for 3D confocal microscope deconvolution.* Microscopy Research and Technique, volume 69, number 4, pages 260–266, April 2006. Cited on page(s) 138.

[Dickensheets 96] David L. Dickensheets and Gordon S. Kino. *Micromachined scanning confocal optical microscope.* Optics Letter, volume 21, number 10, pages 764–766, May 1996. Cited on page(s) 19.

[do Carmo 92] Manfredo Perdigao do Carmo. Riemannian geometry. Birkhauser, 1992. Cited on page(s) 83, 84.

[Duchon 77] Jean Duchon. *Splines minimizing rotation-invariant semi-norms in sobolev spaces.* In Walter Schempp and Karl Zeller, editors, Constructive Theory of Functions of Several Variables, volume 571 of *Lecture Notes in Mathematics*, pages 85–100, Oberwolfach, April 1977. Cited on page(s) 27.

[Elter 06] Matthias Elter, Stephan Rupp and Christian Winter. *Physically motivated reconstruction of fiberscopic images.* In Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06), pages 599–602, Hong Kong, August 2006. Cited on page(s) 22.

[Farnebäck 06] Gunnar Farnebäck and Carl-Fredrik Westin. *Affine and deformable registration based on polynomial expansion.* In Rasmus Larsen, Mads Nielsen and Jon Sporring, editors, Proceedings of the 9th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'06), volume 4191 of *Lecture Notes in Computer Science*, pages 857–864. Springer-Verlag, 2006. Cited on page(s) 32.

[Fasshauer 06] Greg Fasshauer. *Meshfree methods.* In Michael Rieth and Wolfram Schommers, editors, Handbook of Theoretical and Computational Nanotechnology, pages 33–97. American Scientific Publishers, 2006. Cited on page(s) 27.

[Faugeras 07] Olivier Faugeras, Christophe Lenglet, Théodore Papadopoulo and Rachid Deriche. *Non rigid registration of diffusion tensor images.* Research Report 6104, INRIA Sophia-Antipolis, January 2007. Cited on page(s) 141.

[Fletcher 04a] P. Thomas Fletcher and Sarang C. Joshi. *Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors.* In ECCV Workshops CVAMIA and MMBIA, pages 87–98, 2004. Cited on page(s) 83.

[Fletcher 04b] P. Thomas Fletcher, Conglin Lu, Stephen M. Pizer and Sarang Joshi. *Principal geodesic analysis for the study of nonlinear statistics of shape.* IEEE Transactions on Medical Imaging, volume 23, number 8, pages 995–1005, 2004. Cited on page(s) 83.

[Flusberg 05] Benjamin A. Flusberg, Eric D. Cocker, Wibool Piyawattanametha, Juergen C. Jung, Eunice L. M. Cheung and Schnitzer Mark J. *Fiber-optic fluorescence imaging.* Nature Methods, volume 2, number 12, pages 941–950, December 2005. Cited on page(s) 19.

[Gargi 00] Ullas Gargi, Rangachar Kasturi and Susan H. Strayer. *Performance characterization of video-shot-change detection methods.* IEEE Transactions on Circuits and Systems for Video Technology, volume 10, number 1, pages 1–13, February 2000. Cited on page(s) 120.

[Glocker 07] Ben Glocker, Sonja Buhmann, Chlodwig Kirchhoff, Thomas Mussack, Maximilian Reiser and Nassir Navab. *Towards a computer aided diagnosis system for colon motility dysfunctions.* In Maryellen L. Giger and Nico Karssemeijer, editors, Proceedings of the SPIE, Medical Imaging 2007: Computer-Aided Diagnosis, volume 6514, San Diego, CA, March 2007. Cited on page(s) 140.

[Golland 03] Polina Golland and Bruce Fischl. *Permutation tests for classification: Towards statistical significance in image-based studies*. In Christopher J. Taylor and J. Alison Noble, editors, Proceedings of Information Processing in Medical Imaging (IPMI'03), volume 2732 of *Lecture Notes in Computer Science*, pages 330–341, Ambleside, UK, July 2003. Springer-Verlag. Cited on page(s) 141.

[Golland 05] Polina Golland, W. Eric L. Grimson, Martha E. Shenton and Ron Kikinis. *Detection and analysis of statistical differences in anatomical shape*. Medical Image Analysis, volume 9, number 1, pages 68–86, February 2005. Cited on page(s) 141.

[Goshtasby 03] Ardeshir Goshtasby, Lawrence Staib, Colin Studholme and Demetri Terzopoulos. *Nonrigid image registration: Guest editors' introduction*. Computer vision and image understanding, volume 89, number 2–3, pages 109–113, February 2003. Special Issue on Nonrigid Registration. Cited on page(s) 51, 52.

[Groher 06] Martin Groher, Nicolas Padoy, Tobias F. Jakobs and Nassir Navab. *New CTA protocol and 2D-3D registration method for liver catheterization*. In Rasmus Larsen, Mads Nielsen and Jon Sporring, editors, Proceedings of the 9th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'06), volume 4191 of *Lecture Notes in Computer Science*, pages 873–881. Springer-Verlag, 2006. Cited on page(s) 140.

[Helgason 01] Sigurdur Helgason. Differential geometry, Lie groups, and symmetric spaces. American Mathematical Society, 2001. Cited on page(s) 34, 36, 83.

[Hellier 01] Pierre Hellier, Christian Barillot, Etienne Mémin and Patrick Pérez. *Hierarchical estimation of a dense deformation field for 3d robust registration*. IEEE Transactions on Medical Imaging, volume 20, number 5, pages 388–402, May 2001. Cited on page(s) 61.

[Helmchen 02] Fritjof Helmchen. *Miniaturization of fluorescence microscopes using fibre optics*. Experimental Physiology, volume 87, number 6, pages 737–745, November 2002. Cited on page(s) 19.

[Hermosillo 01] Gerardo Hermosillo and Olivier Faugeras. *Dense image matching with global and local statistical criteria: A variational approach*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'01), pages 73–78, 2001. Cited on page(s) 141.

[Hermosillo 02] Gerardo Hermosillo, Christophe Chefd'Hotel and Olivier Faugeras. *Variational methods for multimodal image matching*. International

Journal of Computer Vision, volume 50, number 3, pages 329–343, December 2002. Cited on page(s) 141.

[Higham 05] Nicholas J. Higham. *The scaling and squaring method for the matrix exponential revisited*. SIAM Journal on Matrix Analysis and Applications, volume 26, number 4, pages 1179–1193, 2005. Cited on page(s) 39.

[Hill 01] Derek L. G. Hill, Philipp G. Batchelor, Mark Holden and David J. Hawkes. *Medical image registration*. Physics in Medicine and Biology, volume 46, number 1, pages 1–45, 2001. Cited on page(s) 51, 52.

[Hoffman 06] Arthur Hoffman, Martin Goetz, Michael Vieth, Peter R. Galle, Markus F. Neurath and Ralf Kiesslich. *Confocal laser endomicroscopy: Technical status and current indications*. Endoscopy, volume 38, number 12, pages 1275–1283, December 2006. Cited on page(s) 19.

[Hummel 49] Paul Matthew Hummel and Charles Louis Seebeck Jr. *A generalization of taylor's expansion*. The American Mathematical Monthly, volume 56, number 4, pages 243–247, April 1949. Cited on page(s) 37.

[Ibáñez 05] Luis Ibáñez, Will Schroeder, Lydia Ng and Josh Cates. *The ITK software guide*. Kitware, Inc., 2 edition, 2005. Cited on page(s) 52, 56, 58, 60, 68.

[Inoue 05] Haruhiro Inoue, Shin ei Kudo and Akira Shiokawa. *Technology insight: Laser-scanning confocal microscopy and endocytoscopy for cellular observation of the gastrointestinal tract*. Nature Clinical Practice: Gastroenterology & Hepatology, volume 2, number 1, pages 31–37, January 2005. Cited on page(s) 19.

[Irani 95] Michal Irani, P. Anandan and Steve Hsu. *Mosaic based representations of video sequences and their applications*. In Proceedings of the 5th International Conference on Computer Vision (ICCV'95), pages 605–611, June 1995. Cited on page(s) 80.

[Jannin 02] Pierre Jannin, J. Michael Fitzpatrick, David J. Hawkes, Xavier Pennec, Ramin Shahidl and Michael W. Vannier. *Validation of medical image processing in image-guided therapy*. IEEE Transactions on Medical Imaging, volume 21, number 12, pages 1445–1449, December 2002. Guest editorial. Cited on page(s) 130.

[Jean 07] Florence Jean, Geneviève Bourg-Heckly and Bertrand Viellerobe. *Fibered confocal spectroscopy and multicolor imaging system for in*

vivo fluorescence analysis. Optics Express, volume 15, number 7, pages 4008–4017, April 2007. Cited on page(s) 139.

[Jones 06] Thouis R. Jones, Anne E. Carpenter, David M. Sabatini and Polina Golland. *Methods for high-content, high-throughput image-based cell screening*. In MICCAI Workshop on Microscopic Image Analysis with Applications in Biology (MIAAB'06), October 2006. Cited on page(s) 119.

[Joshi 00] Sarang C. Joshi and Michael I. Miller. *Landmark matching via large deformation diffeomorphisms*. IEEE Transactions on Image Processing, volume 9, number 8, pages 1357–1370, August 2000. Cited on page(s) 64.

[Jung 04] Juergen C. Jung, Amit D. Mehta, Emre Aksay, Raymond Stepnoski and Mark J. Schnitzer. *In vivo mammalian brain imaging using one- and two-photon fluorescence microendoscopy*. Journal of Neurophysiology, volume 92, number 5, pages 3121–3133, November 2004. Cited on page(s) 10.

[Kelsey 04] Peter B. Kelsey. *Esophagus - Barrett's esophagus*. The DAVE Project, May 2004. Cited on page(s) 113.

[Kelsey 05] Peter B. Kelsey. *Stomach - Normal upper endoscopy*. The DAVE Project, January 2005. Cited on page(s) 113.

[Kiesslich 06] Ralf Kiesslich, Liebwin Gossner, Martin Goetz, Alexandra Dahlmann, Michael Vieth, Manfred Stolte, Arthur Hoffman, Michael Jung, Bernard Nafe, Peter R. Galle and Markus F. Neurath. *In vivo histology of Barrett's esophagus and associated neoplasia by confocal laser endomicroscopy*. Clinical Gastroenterology and Hepathology, volume 4, number 8, pages 979–987, August 2006. Cited on page(s) 116, 118, 119.

[Kovačević 06] Jelena Kovačević. *From the editor-in-chief*. IEEE Transactions on Image Processing, volume 15, number 12, pages 3625–3626, December 2006. On Reproducible Research. Cited on page(s) 138.

[Krissian 00] Karl Krissian, Grégoire Malandain, Nicholas Ayache, Régis Vaillant and Yves Trousset. *Model-based detection of tubular structures in 3D images*. Computer vision and image understanding, volume 80, number 2, pages 130–171, November 2000. Cited on page(s) 45.

[Kudo 01] Shin-Ei Kudo, Carlos A. Rubio, Cláudio Rolim Teixeira, Hiroshi Kashida and Etsuko Kogure. *Pit pattern in colorectal neoplasia: Endoscopic magnifying view*. Endoscopy, volume 33, pages 367–373, 2001. Cited on page(s) 14.

[Laemmel 04] Elisabeth Laemmel, Magalie Genet, Georges Le Goualher, Aymeric Perchant, Jean-François Le Gargasson and Eric Vicaut. *Fibered confocal fluorescence microscopy (Cell-viZio™) facilitates extended imaging in the field of microcirculation.* Journal of Vascular Research, volume 41, number 5, pages 400–411, 2004. Cited on page(s) 13, 44.

[Lane 00] Pierre M. Lane, Andrew L. P. Dlugan, Rebecca Richards-Kortum and Calum E. MacAulay. *Fiber-optic confocal microscopy using a spatial light modulator.* Optics Letter, volume 25, number 24, pages 1780–1782, December 2000. Cited on page(s) 19.

[Le Goualher 04] Georges Le Goualher, Aymeric Perchant, Magalie Genet, Charlotte Cavé, Bertrand Viellerobe, Frédéric Berier, Benjamin Abrat and Nicholas Ayache. *Towards optical biopsies with an integrated fibered confocal fluorescence microscope.* In Christian Barillot, David R. Haynor and Pierre Hellier, editors, Proceedings of the 7th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'04), volume 3217 of *Lecture Notes in Computer Science*, pages 761–768. Springer-Verlag, 2004. Cited on page(s) 10, 22, 24, 25.

[Lee 97] Seungyong Lee, George Wolberg and Sung Yong Shin. *Scattered data interpolation with multilevel B-splines.* IEEE Transactions on Visualization and Computer Graphics, volume 3, number 3, pages 228–244, 1997. Cited on page(s) 27, 96.

[Lee 05] Pei Yean Lee and John B. Moore. *Gauss-Newton-on-manifold for pose estimation.* Journal of Industrial and Management Optimization, volume 1, number 4, pages 565–587, November 2005. Cited on page(s) 33.

[Levin 04] Anat Levin, Assaf Zomet, Shmuel Peleg and Yair Weiss. *Seamless image stitching in the gradient domain.* In Proceedings of the 8th European Conference on Computer Vision (ECCV'04), pages 377–389, 2004. Cited on page(s) 80, 109.

[Lewis 95] John P. Lewis. *Fast template matching.* In Proceedings of the International Conference on Vision Interface (VI'95), pages 120–123, 1995. Cited on page(s) 43, 121, 122, 127, 128.

[Lin 06] Ken Y. Lin, Marco A. Maricevich, Aymeric Perchant, Sacha Loiseau, Ralph Weissleder and Umar Mahmood. *Novel imaging method and morphometric analyses of microvasculature in live mice using a fiber-optic confocal laser microprobe.* In Proceedings of the Radiological Society of North America (RSNA'06), Chicago, Il, USA, 2006. Cited on page(s) 45.

[Lodha 99] Suresh K. Lodha and Richard Franke. *Scattered data techniques for surfaces.* In Dagstuhl '97, Scientific Visualization, pages 181–222, Washington, DC, USA, 1999. IEEE Computer Society. Cited on page(s) 26, 96.

[Loewke 07] Kevin E. Loewke, David B. Camarillo, Christopher A. Jobst and J. Kenneth Salisbury. *Real-time image mosaicing for medical applications.* In Proceedings of Medicine Meets Virtual Reality (MMVR'07), pages 304–309, February 2007. Cited on page(s) 125.

[Madsen 99] Kaj Madsen, Hans Bruun Nielsen and Ole Tingleff. *Methods for non-linear least squares problems.* Lecture Notes, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, 1999. Cited on page(s) 35.

[Mahony 02] Robert Mahony and Jonathan H. Manton. *The geometry of the Newton method on non-compact Lie-groups.* Journal of Global Optimization, volume 23, number 3, pages 309–327, August 2002. Cited on page(s) 33, 34, 65.

[Maintz 98] J. B. Antoine Maintz and Max A. Viergever. *A survey of medical image registration.* Medical Image Analysis, volume 2, number 1, pages 1–36, March 1998. Cited on page(s) 32.

[Malis 04] Ezio Malis. *Improving vision-based control using efficient second-order minimization techniques.* In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'04), April 2004. Cited on page(s) 32, 53, 54, 135.

[Mallat 99] Stéphane Mallat. A wavelet tour of signal processing. Academic Press, September 1999. Cited on page(s) 138.

[Marsland 04] Stephen Marsland and Carole Twining. *Constructing diffeomorphic representations for the groupwise analysis of non-rigid registrations of medical images.* IEEE Transactions on Medical Imaging, volume 23, number 8, pages 1006–1020, 2004. Cited on page(s) 53, 64, 75.

[Marsland 07] Stephen Marsland and Robert I. McLachlan. *A Hamiltonian particle method for diffeomorphic image registration.* In Nico Karssemeijer and Boudewijn P. F. Lelieveldt, editors, Proceedings of Information Processing in Medical Imaging (IPMI'07), volume 4584 of *Lecture Notes in Computer Science*, pages 396–407, Kerkrade, The Netherlands, July 2007. Springer-Verlag. Cited on page(s) 141.

[McDonald 03] Donald M. McDonald and Peter L. Choyke. *Imaging of angiogenesis: From microscope to clinic.* Nature Medicine, volume 9, number 6, pages 713–725, June 2003. Cited on page(s) 13, 107.

[Meining 07a] Alexander Meining, Monther Bajbouj, Stefan Delius and Christian Prinz. *Confocal laser scanning microscopy for in vivo histopathology of the gastrointestinal tract.* Arab Journal of Gastroenterology, volume 8, number 1, pages 1–4, March 2007. Cited on page(s) 15.

[Meining 07b] Alexander Meining, Valentin Schwendy Susanne Becker, Roland M. Schmid and Christian Prinz. *In vivo histopathology of lymphocytic colitis.* Gastrointestinal Endoscopy, volume 66, number 2, pages 398–400, August 2007. Cited on page(s) 80, 112.

[Miehlke 07] Stephan Miehlke, Daniela Aust, Ahmed Madisch and Andrea Morgner. *Miniprobe confocal laser microscopy during double-balloon enteroscopy – a feasability study.* In Proceedings of the 15th United European Gastroenterology Week (UEGW'07), October 2007. Cited on page(s) 139, 140.

[Miller 98] Michael I. Miller, Sarang C. Joshi and Gary E. Christensen. *Large deformation fluid diffeomorphisms for landmark and image matching.* In Arthur Toga, editor, Brain Warping, pages 115–131. Elsevier, 1998. Cited on page(s) 64, 65.

[Modersitzki 04] Jan Modersitzki. Numerical methods for image registration. Oxford University Press, 2004. Cited on page(s) 32, 53, 55, 56.

[Murakami 03] Kenzi Murakami, Akiko Murata, Takeshi Suga, Hideya Kitagawa, Yoshitaka Kamiya, Mitsunori Kubo, Kazuya Matsumoto, Hiroshi Miyajima and Masahiro Katashiro. *A miniature confocal optical microscope with MEMs gimbal scanner.* In Proceedings of the 12th International Conference on TRANSDUCERS, Solid-State Sensors, Actuators and Microsystems (TRANSDUCERS'03), volume 1, pages 587–590, 2003. Cited on page(s) 19.

[Noblet 05] Vincent Noblet, Christian Heinrich, Fabrice Heitz and Jean-Paul Armspach. *3-D deformable image registration: A topology preservation scheme based on hierarchical deformation models and interval analysis optimization.* IEEE Transactions on Image Processing, volume 14, number 5, pages 553–566, May 2005. Cited on page(s) 64.

[Oh 06] Wang-Yuhl Oh, Brett E. Bouma, Nicusor Iftimia, Ronit Yelin and Guillermo J. Tearney. *Spectrally-modulated full-field optical coherence microscopy for ultrahigh-resolution endoscopic imaging.* Optics Express, volume 14, number 19, pages 8675–8684, September 2006. Cited on page(s) 10.

[Ourselin 00] Sébastien Ourselin, Alexis Roche, Sylvain Prima and Nicholas Ayache. *Block matching: A general framework to improve robustness*

of rigid registration of medical images. In Proceedings of the 3rd International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'00), volume 1935 of *Lecture Notes in Computer Science*, pages 557–566. Springer-Verlag, 2000. Cited on page(s) 88.

[Pankajakshan 07] Praveen Pankajakshan, Bo Zhang, Laure Blanc-Feraud, Zvi Kam, Jean-Christophe Olivo-Marin and Josiane Zerubia. *Parametric blind deconvolution for confocal laser scanning microscopy*. In Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS'07), pages 6531–6534, Lyon, France, August 2007. Cited on page(s) 138.

[Papoulis 02] Athanasios Papoulis and S. Unnikrishna Pillai. Probability, random variables and stochastic processes. McGraw-Hill, 4th edition, 2002. Cited on page(s) 83, 86.

[Peleg 00] Shmuel Peleg, Benny Rousso, Alex Rav-Acha and Assaf Zomet. *Mosaicing on adaptive manifolds*. IEEE Transactions on Pattern Analysis and Machine Intelligence, volume 22, number 10, pages 1144–1154, October 2000. Cited on page(s) 80.

[Pennec 98] Xavier Pennec and Nicholas Ayache. *Uniform distribution, distance and expectation problems for geometric features processing*. Journal of Mathematical Imaging and Vision, volume 9, number 1, pages 49–67, July 1998. Cited on page(s) 86.

[Pennec 99] Xavier Pennec, Pascal Cachier and Nicholas Ayache. *Understanding the* demon's algorithm*: 3D non-rigid registration by gradient descent*. In Proceedings of the 2nd International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'99), volume 1679 of *Lecture Notes in Computer Science*, pages 597–605. Springer-Verlag, 1999. Cited on page(s) 52, 53, 56, 58, 60, 68.

[Pennec 06a] Xavier Pennec. *Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements*. Journal of Mathematical Imaging and Vision, volume 25, number 1, pages 127–154, July 2006. Cited on page(s) 86.

[Pennec 06b] Xavier Pennec. *Statistical computing on manifolds for computational anatomy*. Habilitation à diriger des recherches, Université Nice Sophia-Antipolis, December 2006. Cited on page(s) 52.

[Pennec 06c] Xavier Pennec, Pierre Fillard and Nicholas Ayache. *A Riemannian framework for tensor computing*. International Journal of Computer

Vision, volume 66, number 1, pages 41–66, January 2006. Cited on page(s) 83.

[Perchant 05] Aymeric Perchant, Georges Le Goualher and Frédéric Berrier. *Method for processing an image acquired through a guide consisting of a plurality of optical fibers*. US Patent number US2005207668, Mauna Kea Technologies, Published on 22 September 2005. Cited on page(s) 10, 22, 24.

[Perchant 06] Aymeric Perchant, Tom Vercauteren, Charlotte Cavé and Sacha Loiseau. *Mosaicing of confocal microscopic in vivo soft tissue video sequences improves multimodality approach*. In Proceedings of the Radiological Society of North America (RSNA'06), Chicago, Illinois, December 2006. Abstract only. Cited on page(s) 5, 7, 111.

[Perchant 07] Aymeric Perchant, Tom Vercauteren, Fabien Oberrietter, Nicolas Savoire and Nicholas Ayache. *Region tracking algorithms on laser scanning devices applied to cell traffic analysis*. In Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'07), pages 260–263, Arlington, USA, April 2007. Cited on page(s) 4, 6, 32, 42, 44, 137.

[Qiao 07] Feng Qiao, Tinsu Pan, John W. Clark Jr and Osama R. Mawlawi. *Region of interest motion compensation for PET image reconstruction*. Physics in Medicine and Biology, volume 52, number 10, pages 2675–2689, May 2007. Cited on page(s) 52.

[Rogelj 06] Peter Rogelj and Stanislav Kovačič. *Symmetric image registration*. Medical Image Analysis, volume 10, number 3, pages 484–493, June 2006. Cited on page(s) 52, 54, 61.

[Rousseeuw 87] Peter J. Rousseeuw and Annick M. Leroy. Robust regression and outlier detection. Wiley, 1987. Cited on page(s) 91.

[Rueckert 99] Daniel Rueckert, Luke I. Sonoda, Carmel Hayes, Derek L. G. Hill, Martin O. Leach and David J. Hawkes. *Nonrigid registration using free-form deformations: Application to breast MR images*. IEEE Transactions on Medical Imaging, volume 18, number 8, pages 712–721, 1999. Cited on page(s) 95.

[Rueckert 06] Daniel Rueckert, Paul Aljabar, Rolf A. Heckemann, Joseph V. Hajnal and Alexander Hammers. *Diffeomorphic registration using B-splines*. In Rasmus Larsen, Mads Nielsen and Jon Sporring, editors, Proceedings of the 9th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'06), volume 4191 of *Lecture Notes in Computer Science*, pages 702–709. Springer-Verlag, 2006. Cited on page(s) 64.

[Sabharwal 99] Yashvinder S. Sabharwal, Andrew R. Rouse, LaTanya Donaldson, Mark F. Hopkins and Arthur F. Gmitro. *Slit-scanning confocal microendoscope for high-resolution in vivo imaging.* Applied Optics, volume 38, number 34, pages 7133–7144, December 1999. Cited on page(s) 19.

[Sampliner 02] Richard E. Sampliner and The Practice Parameters Committee of the American College of Gastroenterology. *Updated guidelines for the diagnosis, surveillance, and therapy of Barrett's esophagus.* American Journal of Gastroenterology, volume 97, number 8, pages 1888–1895, August 2002. Cited on page(s) 113.

[Sato 97] Yoshinobu Sato, Jian Chen, Reza A. Zoroofi, Noboru Harada, Shinichi Tamura and Takeshi Shiga. *Automatic extraction and measurement of leukocyte motion in microvessels using spatiotemporal image analysis.* IEEE Transactions on Biomedical Engineering, volume 44, number 4, pages 225–236, April 1997. Cited on page(s) 42, 44.

[Savoire 04] Nicolas Savoire, Georges Le Goualher, Aymeric Perchant, François Lacombe, Grégoire Malandain and Nicholas Ayache. *Measuring blood cells velocity in microvessels from a single image: Application to* in vivo *and* in situ *confocal microscopy.* In Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'07), pages 456–459, April 2004. Cited on page(s) 42, 91.

[Sawhney 98] Harpreet S. Sawhney, Steve Hsu and R. Kumar. *Robust video mosaicing through topology inference and local to global alignment.* In Proceedings of the 5th European Conference on Computer Vision (ECCV'98), volume 2, pages 103–119, 1998. Cited on page(s) 87, 88.

[Sharma 03] Prateek Sharma, Allan P. Weston, Margarita Topalovski, Rachel Cherian, Achyut Bhattacharyya and Richard E. Sampliner. *Magnification chromoendoscopy for the detection of intestinal metaplasia and dysplasia in Barrett's oesophagus.* Gut, volume 52, number 1, pages 24–27, January 2003. Cited on page(s) 14.

[Sharma 07] Prateek Sharma, Sachin Wani and Ajay Bansal. *The quest for intestinal metaplasia - Is it worth the effort?* American Journal of Gastroenterology, volume 102, number 6, pages 1162–1165, June 2007. Cited on page(s) 113.

[Sokolov 03] Konstantin Sokolov, Jesse Aaron, Betsy Hsu, Dawn Nida, Ann Gillenwater, Michele Follen, Calum MacAulay, Karen Adler-Storthz,

Brian Korgel, Michael Descour, Renata Pasqualini, Wadih Arap, Wan Lam and Rebecca Richards-Kortum. *Optical systems for in vivo molecular imaging of cancer*. Technology in Cancer Research & Treatment, volume 2, number 6, pages 491–504, December 2003. Cited on page(s) 19.

[Stoyan 96] Dietrich Stoyan, Wilfrid S. Kendall and Joseph Mecke. Stochastic geometry and its applications. Wiley, 2nd edition, July 1996. Cited on page(s) 141.

[Strohmer 97] Thomas Strohmer. *Computationally attractive reconstruction of bandlimited images from irregular samples*. IEEE Transactions on Image Processing, volume 6, number 4, pages 540–548, April 1997. Cited on page(s) 27.

[Su 04] Ming-Shing Su, Wen-Liang Hwang and Kuo-Young Cheng. *Analysis on multiresolution mosaic images*. IEEE Transactions on Image Processing, volume 13, number 7, pages 952–959, 2004. Cited on page(s) 109.

[Sung 02] Kung-Bin Sung, Chen Liang, Michael Descour, Tom Collier, Michele Follen and Rebecca Richards-Kortum. *Fiber-optic confocal reflectance microscope with miniature objective for in vivo imaging of human tissues*. IEEE Transactions on Biomedical Engineering, volume 49, number 10, pages 1168–1172, October 2002. Cited on page(s) 19.

[Thiberville 07] Luc Thiberville, Sophie Moreno-Swirc, Tom Vercauteren, Eric Peltier, Charlotte Cavé and Geneviève Bourg Heckly. In vivo *imaging of the bronchial wall microstructure using fibered confocal fluorescence microscopy*. American Journal of Respiratory and Critical Care Medicine, volume 175, number 1, pages 22–31, January 2007. Chosen for the cover of the AJRCCM paper issue. Cited on page(s) 5, 7, 15, 111, 137.

[Thirion 98] Jean-Philippe Thirion. *Image matching as a diffusion process: An analogy with Maxwell's demons*. Medical Image Analysis, volume 2, number 3, pages 243–260, 1998. Cited on page(s) 52, 53, 54, 55, 58, 60, 68, 95.

[Toussaint 07] Nicolas Toussaint, Jean-Christophe Souplet and Pierre Fillard. *MedINRIA: Medical image navigation and research tool by INRIA*. In Proceedings of MICCAI'07 Workshop on Interaction in Medical Image Analysis and Visualization, Brisbane, Australia, November 2007. Cited on page(s) 75, 138.

[Trouvé 98] Alain Trouvé. *Diffeomorphisms groups and pattern matching in image analysis.* International Journal of Computer Vision, volume 28, number 3, pages 213–221, July 1998. Cited on page(s) 64, 65.

[Urschler 07] Martin Urschler, Stefan Kluckner and Horst Bischof. *A framework for comparison and evaluation of nonlinear intra-subject image registration algorithms.* Insight Journal – ISC/NA-MIC Workshop on Open Science at MICCAI, November 2007. Available online with code source at `http://hdl.handle.net/1926/561`. Cited on page(s) 71, 75.

[Uyttendaele 01] Matthew Uyttendaele, Ashley Eden and Richard Szeliski. *Eliminating ghosting and exposure artifacts in image mosaics.* In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'01), pages 509–516, 2001. Cited on page(s) 97.

[Vaillant 04] Marc Vaillant, Michael I. Miller, Laurent Younes and Alain Trouvé. *Statistics on diffeomorphisms via tangent space representations.* Neuroimage, volume 23, number S1, pages 161–169, 2004. Cited on page(s) 64.

[Vázquez 05] Carlos Vázquez, Eric Dubois and Janusz Konrad. *Reconstruction of nonuniformly sampled images in spline spaces.* IEEE Transactions on Image Processing, volume 14, number 6, pages 713–725, June 2005. Cited on page(s) 27.

[Vercauteren 05] Tom Vercauteren, Aymeric Perchant, Xavier Pennec and Nicholas Ayache. *Mosaicing of confocal microscopic* in vivo *soft tissue video sequences.* In James S. Duncan and Guido Gerig, editors, Proceedings of the 8th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'05), volume 3749 of *Lecture Notes in Computer Science*, pages 753–760. Springer-Verlag, 2005. Cited on page(s) 5, 7, 78, 136.

[Vercauteren 06] Tom Vercauteren, Aymeric Perchant, Grégoire Malandain, Xavier Pennec and Nicholas Ayache. *Robust mosaicing with correction of motion distortions and tissue deformation for* in vivo *fibered microscopy.* Medical Image Analysis, volume 10, number 5, pages 673–692, 2006. Annual MedIA/MICCAI Best Paper Award 2006. Cited on page(s) 5, 7, 28, 29, 41, 42, 78, 80, 136, 137.

[Vercauteren 07a] Tom Vercauteren, Anne Osdoit, Aymeric Perchant and Sacha Loiseau. *Mosaicing of confocal microscopic video sequences: Larger field of view and still higher resolution!* In Proceedings of the Digestive Disease Week (DDW'07), page AB352, Washington, DC, May 2007. Abstract only. Cited on page(s) 5, 8, 111.

[Vercauteren 07b] Tom Vercauteren, Xavier Pennec, Ezio Malis, Aymeric Perchant and Nicholas Ayache. *Insight into efficient image registration techniques and the demons algorithm.* In Nico Karssemeijer and Boudewijn P. F. Lelieveldt, editors, Proceedings of Information Processing in Medical Imaging (IPMI'07), volume 4584 of *Lecture Notes in Computer Science*, pages 495–506, Kerkrade, The Netherlands, July 2007. Springer-Verlag. Cited on page(s) 4, 5, 6, 32, 52, 53, 58, 135, 136.

[Vercauteren 07c] Tom Vercauteren, Xavier Pennec, Aymeric Perchant and Nicholas Ayache. *Diffeomorphic demons using ITK's finite difference solver hierarchy.* Insight Journal – ISC/NA-MIC Workshop on Open Science at MICCAI, November 2007. Available online with source code at `http://hdl.handle.net/1926/510`. Cited on page(s) 6, 52, 71, 75, 138.

[Vercauteren 07d] Tom Vercauteren, Xavier Pennec, Aymeric Perchant and Nicholas Ayache. *Non-parametric diffeomorphic image registration with the demons algorithm.* In Nicholas Ayache, Sébastien Ourselin and Anthony J. Maeder, editors, Proceedings of the 10th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'07), volume 4792 of *Lecture Notes in Computer Science*, pages 319–326, Brisbane, Australia, October 2007. Springer-Verlag. Cited on page(s) 5, 6, 52, 53, 68, 136.

[Vercauteren 07e] Tom Vercauteren, Aymeric Perchant and Nicholas Ayache. *In vivo microscopy for real-time structural and functional cellular imaging.* ERCIM News, volume 69, pages 34–35, April 2007. Cited on page(s) 4, 6, 9.

[Vercauteren 08a] Tom Vercauteren, Nicholas Ayache, Nicolas Savoire, Grégoire Malandain and Aymeric Perchant. *Processing of* in vivo *fibered confocal microscopy video sequences.* In Jens Rittscher, Raghu Machiraju and Stephen T. C. Wong, editors, Microscopic Image Analysis for Life Science Applications. Artech House, 2008. Approximatively 30 pages. In press. Cited on page(s) 4, 6, 9.

[Vercauteren 08b] Tom Vercauteren, Alexander Meining, François Lacombe and Aymeric Perchant. *Real time autonomous video image registration for endomicroscopy: Fighting the compromises.* In Jose-Angel Conchello, Carol J. Cogswell and Tony Wilson, editors, Proc. SPIE BIOS - Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing XV, San Jose, CA, USA, January 2008. SPIE. Full-paper accepted on abstract. Cited on page(s) 7, 125.

[Vincent 06] Pierre Vincent, Uwe Maskos, Igor Charvet, Laurence Bourgeais, Luc Stoppini, Nathalie Leresche, Jean-Pierre Changeux, Régis Lambert, Paolo Meda and Danièle Paupardin-Tritsch. *Live imaging of neural structure and function by fibred fluorescence microscopy.* EMBO report, volume 7, number 11, pages 1154–1161, November 2006. Cited on page(s) 11, 30.

[Wachinger 07] Christian Wachinger, Wolfgang Wein and Nassir Navab. *Three-dimensional ultrasound mosaicing.* In Nicholas Ayache, Sébastien Ourselin and Anthony J. Maeder, editors, Proceedings of the 10th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'07), volume 4792 of *Lecture Notes in Computer Science*, Brisbane, Australia, October 2007. Springer-Verlag. Cited on page(s) 136.

[Wang 03] Thomas D. Wang, Michael J. Mandella, Christopher H. Contag and Gordon S. Kino. *Dual-axis confocal microscope for high-resolution in vivo imaging.* Optics Letter, volume 28, number 6, pages 414–416, March 2003. Cited on page(s) 19.

[Wang 05] He Wang, Lei Dong, Jennifer O'Daniel, Radhe Mohan, Adam S. Garden, K. Kian Ang, Deborah A. Kuban, Mark Bonnen, Joe Y. Chang and Rex Cheung. *Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy.* Physics in Medicine and Biology, volume 50, number 12, pages 2887–2905, 2005. Cited on page(s) 52, 54, 61.

[Wilson 90] Tony Wilson. Confocal microscopy. Academic Press, October 1990. Cited on page(s) 17, 18.

[Winter 06] Christian Winter, Stephan Rupp, Matthias Elter, Christian Münzenmayer, Heinz Gerhäuser and Thomas Wittenberg. *Automatic adaptive enhancement for images obtained with fiberscopic endoscopes.* IEEE Transactions on Biomedical Engineering, volume 53, number 10, pages 2035–2046, October 2006. Cited on page(s) 22.

[Wüstner 03] Michael Wüstner. *A connected Lie group equals the square of the exponential image.* Journal of Lie Theory, volume 13, number 1, pages 307–309, 2003. Cited on page(s) 74.

[Xue 06] Zhong Xue, Dinggang Shen, Bilge Karacali, Joshua Stern, David Rottenberg and Christos Davatzikos. *Simulating deformations of MR brain images for validation of atlas-based segmentation and registration algorithms.* Neuroimage, volume 33, number 3, pages 855–866, November 2006. Cited on page(s) 64.

[Yang 05] Lisong Yang, Aaron Mac Raighne, Eithne M. McCabe, L. Andrea Dunbar and Toralf Scharf. *Confocal microscopy using variable-focal-length microlenses and an optical fiber bundle.* Applied Optics, volume 44, number 28, pages 5928–5936, October 2005. Cited on page(s) 19.

[Yoo 05] Terry S. Yoo and Dimitris N. Metaxas. *Open science - Combining open data and open source software: Medical image analysis with the Insight Toolkit.* Medical Image Analysis, volume 9, number 6, pages 503–506, December 2005. Guest editorial. Cited on page(s) 138.

[Zijdenbos 02] Alex P. Zijdenbos, Reza Forghani and Alan C. Evans. *Automatic "pipeline" analysis of 3-D MRI data for clinical trials: Application to multiple sclerosis.* IEEE Transactions on Medical Imaging, volume 21, number 10, pages 1280–1291, October 2002. Cited on page(s) 119.

[Zöllei 07] Lilla Zöllei, Mark Jenksinon, Samson Timoner and William M. Wells. *A marginalized MAP approach and EM optimization for pair-wise registration.* In Nico Karssemeijer and Boudewijn P. F. Lelieveldt, editors, Proceedings of Information Processing in Medical Imaging (IPMI'07), volume 4584 of *Lecture Notes in Computer Science*, pages 662–674, Kerkrade, The Netherlands, July 2007. Springer-Verlag. Cited on page(s) 141.

[Zomet 00] Assaf Zomet and Shmuel Peleg. *Efficient super-resolution and applications to mosaics.* In Proceedings of the 15th International Conference on Pattern Recognition (ICPR'00), volume 1, pages 3–7, September 2000. Cited on page(s) 99.