



HAL
open science

Variational Methods for Multimodal Image Matching

Gerardo Hermosillo Valadez

► **To cite this version:**

Gerardo Hermosillo Valadez. Variational Methods for Multimodal Image Matching. Human-Computer Interaction [cs.HC]. Université Nice Sophia Antipolis, 2002. English. NNT : . tel-00457459

HAL Id: tel-00457459

<https://theses.hal.science/tel-00457459>

Submitted on 17 Feb 2010

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

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

THÈSE

présentée à

UNIVERSITÉ DE NICE - SOPHIA ANTIPOLIS

pour obtenir le titre de

DOCTEUR EN SCIENCES

École Doctorale Sciences et Technologies de l'Information et de
la Communication

Spécialité

Image & Vision

soutenue par

Gerardo Hermosillo Valadez

le 3 mai 2002

Titre

**Variational Methods for Multimodal
Image Matching**

Méthodes Variationnelles pour le Recalage Multimodal

Directeur de thèse : Olivier Faugeras

Jury

Président

Nicholas Ayache

Rapporteurs

Luis Alvarez
Joachim Weickert
Laurent Younes

Examineurs

Michel Barlaud
Luc Robert

Acknowledgments

This thesis was funded by the Mexican National Council for Science and Technology (CONACYT) through the scholarship offered in conjunction with the French Society for the Exportation of Educational Resources (SFERE). I am grateful to Luis Alvarez, Joachim Weickert and Laurent Younes for accepting to be reviewers of the thesis dissertation. Their comments have being highly encouraging to me. I also thank Michel Barlaud and Luc Robert for accepting to be examiners of the thesis and Nicholas Ayaache, who was the chairman of the committee and whose kind suggestions helped improve the manuscript. I thank very specially my thesis advisor Olivier Faugeras, for welcoming me as part of the ROBOTVIS team and for the many hours of fruitful and exciting research discussions. I have learned a lot from his example during these years and my thesis has largely benefited from his deep insight and rigorous work. I would like to acknowledge numerous friends and colleagues with whom I have partaken of many aspects of life over the last few years. I can not possibly mention them all but they will recognize themselves. Each of them brings warm memories to my mind. Perhaps more special thanks are due to José Gomes, Jacques Bride and Chirstophe Chefd'Hotel, with whom I worked more closely. I am thankful for their help and what they have taught me. I must also mention the people that first welcomed me four years ago, of which Nour, Imad, Diane, Nikos and Robert are just a few. They all occupy special places in my memory. Thanks go also to our friends Jacques, Fred, Marie-Cécile, Pierre, David, and many others who have shared with us most of our social activities during our stay in France. I want to express my deep thankfulness to my wife Alejandra, who has helped me a lot with her strength and courage. Her love and unwavering support have been indispensable pillars to my work, and I truly share this accomplishment with her. Throughout my life, my parents Jorge and Socorro have helped me in every possible way, frequently not without self-sacrifice. I dedicate this achievement to them in gratefulness for their unconditional love.

Gerardo Hermosillo, May 2002.

Abstract

During the past few years, the use of the theory of partial differential equations has provided a solid formal approach to image processing and analysis research, and has yielded provably well-posed algorithms within a set of clearly defined hypotheses. These algorithms are the state-of-the-art in a large number of application fields such as image de-noising, segmentation and matching. At the same time, the combination of stochastic and variational approaches has led to powerful algorithms which may also be described in terms of partial differential equations. This is the approach followed in the present work, which studies the problem of dense matching between two images using statistical dissimilarity criteria. Two classes of algorithms are considered, corresponding to these criteria being calculated globally for the entire image, or locally within corresponding regions. In each case, three dissimilarity criteria are studied, defined as the opposite of the following similarity measures: mutual information (well adapted to a general statistical dependence between the grey-level intensities), correlation ratio (adapted to a functional dependence), and cross correlation (adapted to an affine dependence). The minimization of the sum of the dissimilarity term and a regularization term defines, through the associated Euler-Lagrange equations, a set of coupled functional evolution equations. Particular emphasis is put in establishing the conditions under which these evolution equations are well posed, i.e. they have a unique solution. It is shown that the proposed algorithms satisfy these conditions for two classes of linear regularization terms, including one which encourages discontinuities of the solution at the contours of the reference image. The discretization and the numerical implementation of the matching algorithms is discussed in detail and their performance is illustrated through several real and synthetic examples, both with 2D and 3D images. As these examples show, the described algorithms are of interest in applications which do not necessarily involve sensors of multiple modalities. They are also of special interest to the medical imaging community, where data fusion between different imaging sensors often requires correcting for nonlinear distortions.

Resumé

Depuis quelques années, l'utilisation des équations aux dérivées partielles a pourvu la recherche en traitement d'images d'une approche formelle solide, et a abouti à des algorithmes dont on peut montrer le caractère bien posé, étant donné un ensemble d'hypothèses clairement définies. Ces algorithmes forment l'état de l'art dans beaucoup de domaines d'application tels que le débruitage, la segmentation et la mise en correspondance. En parallèle à ceci, des approches combinant des principes variationnels et stochastiques ont amené à de puissants algorithmes qui peuvent aussi être décrits en termes d'équations aux dérivées partielles. C'est l'approche suivie dans ce travail, où est étudié le problème de mise en correspondance dense entre deux images, en utilisant des critères statistiques de dissemblance. Deux classes d'algorithmes sont considérées, selon que ces critères soient calculés globalement pour toute l'image, ou localement entre des régions correspondantes. Dans chaque cas, trois critères de dissemblance sont étudiés, définis comme l'opposé des critères de ressemblance suivants: information mutuelle (bien adaptée à une dépendance statistique très générale entre les niveaux de gris), rapport de corrélation (adapté à une dépendance fonctionnelle), et corrélation croisée (adaptée à une dépendance affine). La minimisation de la somme du terme de dissemblance et un terme de régularisation définit, à travers les équations d'Euler-Lagrange, un système d'équations fonctionnelles d'évolution. Nous étudions les conditions sous lesquelles ces équations d'évolution sont bien posées, c'est-à-dire ont une solution unique et montrons que les algorithmes proposés satisfont ces conditions pour deux classes d'opérateurs linéaires régularisants, dont une est conçue pour encourager des variations rapides de la solution le long des contours de l'image de référence. La performance de ces algorithmes est illustrée à travers plusieurs exemples synthétiques et réels, aussi bien sur des images 2D que 3D. Comme le montrent ces exemples, les algorithmes décrits sont applicables à des problèmes qui ne font pas nécessairement intervenir des capteurs de modalités différentes. Ils sont aussi spécialement intéressants pour la communauté de l'imagerie médicale, où le problème de fusionner des données provenant de différentes modalités d'imagerie nécessite souvent de corriger des distorsions non-linéaires.

Contents

Méthodes Variationnelles pour le Recalage Multimodal	15
Introduction	23
Contributions	24
Document Layout	26
I A Generic Image Matching Problem	29
1 Overview	31
1.1 Definition of Images	31
1.2 Image Matching	31
1.3 Multimodality and Statistical Similarity Criteria	33
1.4 Dense Matching and the Variational Framework	36
2 Study of the Abstract Matching Flow	41
2.1 Definitions and Notations	41
2.2 Basic Properties	42
2.3 Semigroups of Linear Operators	44
2.4 Solutions of the Abstract Matching Flow	47
2.4.1 Mild and Strong Solutions	48
2.4.2 Classical Solution	48
3 Regularization Operators	53
3.1 Functional Spaces	53
3.2 Notations	54
3.3 Image Driven Anisotropic Diffusion	54
3.4 The Linearized Elasticity Operator	57
3.5 Existence of Minimizers	59

II	Study of Statistical Similarity Measures	63
4	Definition of the Statistical Measures	65
4.1	Global Criteria	66
4.2	Local Criteria	68
4.3	Continuity of \mathbf{MI}^g and \mathbf{MI}^l	71
5	The Euler-Lagrange Equations	75
5.1	Global Criteria	75
5.1.1	Mutual Information	75
5.1.2	Correlation Ratio	78
5.1.3	Cross Correlation	79
5.2	Local Criteria	80
5.2.1	Mutual Information	81
5.2.2	Correlatio Ratio	83
5.2.3	Cross Correlation	83
5.3	Summary	84
6	Properties of the Matching Terms	87
6.1	Preliminary Results	87
6.2	Global Criteria	89
6.2.1	Mutual Information	89
6.2.2	Correlation Ratio	97
6.2.3	Cross Correlation	103
6.3	Local Criteria	105
6.3.1	Mutual Information	106
6.3.2	Correlation Ratio	110
6.3.3	Cross Correlation	116
III	Implementation Aspects	121
7	Numerical Schemes	123
7.1	Regularization Operators	123
7.1.1	The Linearized Elasticity Operator	124
7.1.2	The Nagel-Enkelmann Operator	126
7.2	Dissimilarity Terms	127
7.3	Approximate Implementations of $F_{\mathbf{MI}}^g(\mathbf{h})$ and $F_{\mathbf{CR}}^l(\mathbf{h})$	131
7.3.1	Mutual Information	131
7.3.2	Correlation Ratio	132
7.3.3	Parallel Implementation	133

8	Determining Parameters	137
8.1	Determining the Smoothing Parameter	138
9	Experimental Results	143
9.1	Classification	143
9.2	Description of the Experiments	143
	Appendices	163
A	Other Applications	163
A.1	Entropy Minimization for Image Segmentation	163
A.2	Diffeomorphic Matching	164
B	Library Description	171
B.1	General Remarks	171
B.2	C++ Listing: Global Matching Functions	172
B.2.1	Mutual Information	172
B.2.2	Correlation Ratio	174
B.2.3	Cross Correlation	177
B.3	C++ Listing: Local Matching Functions	178
B.3.1	Mutual Information	179
B.3.2	Correlation Ratio	181
B.3.3	Cross Correlation	183
B.4	C++ Listing: 2D Matching Flow	185
B.5	C++ Listing: Main Program and Multiscale Handling	187
	Bibliography	189

List of Figures

1.1	Examples of different image modalities	32
1.2	Monomodal matching examples	34
1.3	Nonrigid “multimodal” matching examples	35
1.4	Synthetic sensors and the support of their joint intensity distributions	37
1.5	Schematic joint intensity distribution	38
2.1	The complex plane and the sector Δ of Definition 2.2	46
2.2	The complex plane and the sectors Δ_δ and Σ_δ defined in Theorem 2.7	46
2.3	The complex plane and the sectors Σ_δ and Δ_δ defined in Theorem 2.13.	51
4.1	Local joint intensity distribution.	69
5.1	Interpretation of $\nabla \mathcal{J}_{\text{MI}^g}(\mathbf{h})$	77
5.2	Interpretation of $\nabla \mathcal{J}_{\text{CR}^g}(\mathbf{h})$	79
5.3	Interpretation of $\nabla \mathcal{J}_{\text{CC}^g}(\mathbf{h})$	81
7.1	Execution flow for the master processor	134
7.2	Execution flow for each slave processor in the parallel implementation of the matching flow.	135
8.1	Density estimation example 1	140
8.2	Density estimation example 2	141
9.1	Behavior of the two regularization operators	145
9.2	Displacement fields with linearized elasticity and anisotropic diffusion	146
9.3	Linearized elasticity with ξ close to $\frac{1}{2}$ and close to one	146
9.4	Determinant of $D(\text{Id} + \mathbf{h}^*)$	147
9.5	Proton density image matching against T2-weighted MRI.	148
9.6	Deformation field recovered in the experiment of figure 9.5.	149
9.7	Components of the deformation field	149
9.8	Determinant of the Jacobian of the deformation	150
9.9	Matching with local mutual information and correlation ratio	151
9.10	Realigned image and its superposition with the reference image	151
9.11	Components of the applied deformation field	152

9.12	Components of the recovered deformation field	152
9.13	Determinant of the Jacobian for the deformation	153
9.14	Global mutual information with fMRI data	154
9.15	Matching of T2-weighted anatomical MRI against EPI functional MRI	155
9.16	Matching of anatomical vs diffusion-tensor-related MRI	156
9.17	Stereo matching using global mutual information.	157
9.18	Deformed and reference images	158
9.19	Components of the obtained deformation field	158
9.20	Determinant of the Jacobian for the obtained deformation	159
9.21	Human template matching. Reference (left) and target (right) images.	160
9.22	Deformed template	160
9.23	Some corresponding points	161
9.24	Components of the obtained deformation	161
9.25	Determinant of the Jacobian for the obtained deformation	162
A.1	Segmentation example 1	165
A.2	Segmentation example 2	165
A.3	Segmentation example 3	166
A.4	Diffeomorphic matching using local mutual information	167
A.5	Diffeomorphic matching using local cross correlation: first example .	168
A.6	Diffeomorphic matching using local cross correlation: second example	168
A.7	Diffeomorphic matching using local cross correlation: third example .	169

Méthodes Variationnelles pour le Recalage Multimodal

Introduction

Cette thèse porte sur le problème de la mise en correspondance dense entre deux images, et en particulier lorsqu'une comparaison directe des intensités s'avère impossible. Résoudre automatiquement ce problème est une étape fondamentale dans l'exploitation et l'étude du contenu des images. Par exemple, c'est un pré-requis essentiel dans plusieurs problèmes de vision par ordinateur tels que l'étalonnage de caméras et la reconstruction 3D à partir de (au moins) deux vues d'une scène. Le problème peut être vu génériquement comme celui de la fusion de données, c'est-à-dire celui de la mise en correspondance d'informations provenant de plusieurs sources. Quand les sources sont d'une nature complémentaire, elles partagent par définition très peu d'information commune, et il s'avère donc difficile de fusionner leur sorties respectives. Ceci est un problème très courant dans l'analyse des images médicales, où l'on est souvent confronté à des multiples modalités d'imagerie (Tomographie par rayons X, Résonance Magnétique Nucléaire, Emission de Positrons, etc.). Dans ce contexte, le problème est souvent appelé recalage multimodal. D'autres situations où une comparaison directe des intensités devient inutile apparaissent en vision par ordinateur. Ainsi, mettre en correspondance des structures similaires sous des conditions d'illumination variantes ou lorsque les objets ont des propriétés de réflectance ou de diffusion différents (albédos différents) sont deux exemples où les méthodes de recalage multimodal peuvent également s'appliquer.

Nous proposons une approche variationnelle pour le recalage multimodal non-rigide. Les techniques décrites reposent sur le calcul de mesures statistiques de dissemblance entre les intensités de régions correspondantes. Deux familles d'algorithmes sont considérées, correspondant au calcul global ou local de ces critères. L'approche suivie est celle d'une modélisation continue du problème et le calcul de la première variation des critères statistiques. L'existence et l'unicité d'une solution aux flots de minimisation est démontrée pour les trois critères étudiés (dans leur version locale et globale) ainsi que pour deux familles d'opérateurs différentiels de régularisation.

Plan du manuscrit

Le document est divisé en trois parties. La première partie (chapitres 1 à 3) est consacrée à la description des concepts essentiels mis en jeu dans l'appariement de deux images en utilisant des critères statistiques de dissemblance, et donne une vue d'ensemble de l'approche proposée. Les conditions nécessaires à l'existence et unicité de la solution du problème de minimisation sont établies et deux opérateurs de régularisation sont étudiés en montrant qu'ils satisfont les propriétés requises. La seule partie des algorithmes qui n'est pas traitée est celle qui concerne le terme de mise en correspondance, issue du critère de dissemblance. C'est là l'objet de la deuxième partie (chapitres 4 à 6), qui étudie en détail ce terme des équations fonctionnelles,

en calculant tout d'abord la première variation des six critères de dissemblance et en établissant ensuite leurs bonnes propriétés pour le caractère bien posé du processus de minimisation. Finalement, la troisième partie (chapitres 7 à 9) décrit en détail la discrétisation et l'implémentation numérique des algorithmes qui résultent de ces équations, et présente des résultats expérimentaux avec des déformations synthétiques et réelles, mettant en jeu des images 2D ou 3D. Un résumé de chaque chapitre est donné à continuation.

PARTIE I: Un Problème Générique de Mise en Correspondance Dense

CHAPITRE 1

Ce chapitre introduit les concepts de base mis en jeu dans la mise en correspondance de deux images. Il commence par une introduction de la théorie de l'espace d'échelle (scale-space) et donne la définition d'image adoptée dans la suite. Il continue en définissant le problème de mise en correspondance en fonction du type de transformation recherchée et décrit ensuite les critères statistiques de ressemblance en général. Il termine en décrivant le formalisme du calcul de variations et résume l'approche suivie dans cette thèse en donnant la forme générale des équations d'évolution qui gouvernent le processus de minimisation.

CHAPITRE 2

Ce chapitre est consacré à l'étude du problème de minimisation introduit au chapitre 1, dans le cadre abstrait de l'analyse fonctionnelle. L'existence et l'unicité de plusieurs types (faible, forte, classique) de solutions au problème d'évolution est démontrée en supposant un terme de mise en correspondance Lipschitz-continu et des opérateurs de régularisation générant des semi-groupes (continus, analytiques) de contractions.

CHAPITRE 3

Ce chapitre étudie la partie régularisation des algorithmes de mise en correspondance. Deux familles différentes d'opérateurs linéaires sont considérées, dont une conçue pour encourager les discontinuités du champ de déplacement le long des contours de l'image de référence. Il est montré que ces opérateurs génèrent des semi-groupes uniformément continus et analytiques de contractions et satisfont donc les conditions nécessaires établies au chapitre 2. Après une discussion sur les espaces fonctionnels considérés, des preuves générales d'existence de fonctions minimisantes pour les fonctionnelles d'énergie proposées sont décrites.

PARTIE II: Étude des Mesures Statistiques de Similarité

CHAPITRE 4

Ce chapitre introduit les deux classes de termes de dissemblance considérées, appelées globales et locales. Leur définition est donnée en termes d'estimations non-paramétriques de la densité jointe des intensités à partir soit des images dans leur ensemble (globales) soit de régions correspondantes autour de chaque point (locales). Dans chaque cas, trois mesures de similarité sont définies: corrélation croisée, rapport de corrélation et information mutuelle.

CHAPITRE 5

Dans ce chapitre, les équations d'Euler-Lagrange associées sont dérivées pour les six mesures de dissemblance. Étant donnée la forme complexe de ces fonctionnelles, un calcul explicite de leur dérivée de Gâteaux est nécessaire pour le calcul des équations d'Euler-Lagrange.

CHAPITRE 6

Ce chapitre est consacré à montrer que les gradients calculés au chapitre 5 définissent des fonctions satisfaisant les conditions de continuité nécessaires à l'existence et unicité de la solution du problème de minimisation, telles qu'elles sont établies au chapitre 2.

PARTIE III: Aspects Numériques

CHAPITRE 7

Ce chapitre décrit les schémas numériques utilisés pour discrétiser les équations d'évolution continues, ainsi que pour l'interpolation des images et leurs gradients. Des schémas en temps explicites et implicites sont considérés. Le schéma d'implémentation de l'estimation de densité basé sur du filtrage récursif est décrit en détail.

CHAPITRE 8

Ce chapitre étudie l'estimation des différents paramètres des algorithmes, et en particulier du paramètre de lissage dans l'estimation de la densité jointe d'intensités.

CHAPITRE 9

Ce chapitre présente des résultats expérimentaux pour tous les algorithmes décrits en utilisant des données aussi bien synthétiques que réelles. Les exemples incluent des images 2D pour des applications en vision par ordinateur, et des images 3D provenant de différentes modalités d'imagerie médicale.

Conclusion

Dans cette thèse, nous étudions le problème de mise en correspondance dense entre deux images, en utilisant des critères statistiques de dissemblance. Deux classes d'algorithmes sont considérées, selon que ces critères soient calculés globalement pour toute l'image, ou localement entre des régions correspondantes. Dans chaque cas, trois critères de dissemblance sont étudiés, définis comme l'opposé des critères de ressemblance suivants: information mutuelle (bien adaptée à une dépendance statistique très générale entre les niveaux de gris), rapport de corrélation (adapté à une dépendance fonctionnelle), et corrélation croisée (adaptée à une dépendance affine). La minimisation de la somme du terme de dissemblance et un terme de régularisation définit, à travers les équations d'Euler-Lagrange, un système d'équations fonctionnelles d'évolution. Nous étudions les conditions sous lesquelles ces équations d'évolution sont bien posées, c'est à dire ont une solution unique et montrons que les algorithmes proposés satisfont ces conditions pour deux classes d'opérateurs linéaires régularisants, dont une est conçue pour encourager des variations rapides de la solution le long des contours de l'image de référence. La performance de ces algorithmes est illustrée à travers plusieurs exemples synthétiques et réels, aussi bien sur des images 2D que 3D. Comme le montrent ces exemples, les algorithmes décrits sont applicables à des problèmes qui ne font pas nécessairement intervenir des capteurs de modalités différentes. Ils sont aussi spécialement intéressants pour la communauté de l'imagerie médicale, où le problème de fusionner des données provenant de différentes modalités d'imagerie nécessite souvent de corriger des distorsions non-linéaires.

Variational Methods for Multimodal Image Matching

Introduction

The present thesis deals with a specific problem in the field of image analysis, namely *image matching*. Loosely speaking, the problem is that of establishing correspondences between points in two different images. Solving this problem is a fundamental prerequisite in understanding and exploiting the contents of images. For instance, it is essential in middle-level computer vision tasks such as camera calibration and 3D reconstruction from two or more views. The problem may be viewed in a generic fashion as that of data fusion, i.e. that of putting in correspondence information from several sources. When the sources are of a complementary nature, they share by definition little or no common information, and therefore fusing their outputs becomes particularly difficult. This is a common problem in medical imaging, where several modalities are widely used (Xray Tomography, Magnetic Resonance Imaging (MRI), functional MRI (fMRI), Positron Emission Tomography, etc.). In this context, the problem is called *multimodal* image matching. Other situations in which a comparison of the source outputs becomes difficult are matching under varying illumination conditions, or images produced by physical objects with different responses to the same illumination (e.g. different albedos). A possible approach to the solution of this problem is to define meaningful structures in the image, invariant under transformations of the grey-level intensities, for example edges, corners, etc, and then design low-level methods to extract them from the image. Methods to match these structures can then be devised.

This thesis proposes a variational framework for dense multimodal matching. Rather than working with extracted features, the described techniques rely on the computation of statistical dissimilarity measures between the intensities of corresponding regions. The approach which is followed is that of a continuous modeling of the problem, based on the theory of the calculus of variations and partial differential equations (PDEs). This formalism has proved very fruitful in image processing and analysis through its application in image de-noising, segmentation, and matching, mainly because it de-emphasizes the role of discretization and allows to take profit of many results from these mathematical disciplines. The proposed algorithms are divided into two families, corresponding respectively to global and local statistical dissimilarity criteria. The well posedness of the proposed algorithms is proved by showing existence and uniqueness of the solution to the evolution equations that describe the maximiza-

tion of the similarity criteria.

Contributions

This section aims at situating the present work in its context with respect to existing methods, so that a clear appreciation of its contributions and limits may be established. The amount of literature on the subject of image matching is very large. A good survey in the computer vision domain is provided by Mitiche and Boutemy [59]. Barron, et al. give a performance evaluation of some popular optical flow algorithms in [13]. In the domain of medical image registration, a good and recent survey is provided by Maintz and Viergever [9].

At a conceptual level, most of the existing methods for the recovery of motion rely on the minimization of an energy which encompasses two sources of a priori knowledge: (a) what should a good matching satisfy and (b) a model of the transformation or some other constraint allowing to limit the search for possible matches. As for the first point we can mention the optical flow constraint, the local image differences, or more general block matching strategies [79, 66], which allow to use richer, non-local similarity measures (cross-correlation [35, 36, 21, 64], mutual information [87, 93, 88, 52], correlation-ratio [76], among several others [94, 43, 70, 50]). As for the second point, we can find for instance the search for low-dimensional transformations (e.g. affine, quadratic, or spline-interpolation between a set of control points [57, 78]). Another example of a constrained deformation is the stereo case, in which the knowledge of the fundamental matrix allows to restrict the search for the matching point along the epipolar line [3, 95]. If the searched transformation is a more general function (i.e. not described by parameters), the constraint may consist in requiring some smoothness of the displacement field, possibly preserving discontinuities [81, 3, 72, 6, 55, 56, 11, 10, 32]. Statistical similarity measures have been widely used in the context of image registration through their maximization over a set of low parametric transformations [9]. Mutual information was introduced by Viola et al. [87, 93, 88] and independently by Maes et al. [52]. The correlation ratio was first proposed as a similarity measure for image matching by Roche et al. [76]. Other statistical approaches rely on learning the joint distribution of intensities, as done for instance by Leventon et al. [50]. Extensions to more complex (non-rigid) transformations using statistical similarity measures include approaches relying on more complex parametric transformations [57, 78], block-matching strategies [53, 41, 37], or parametric intensity corrections [74]. Some recent approaches rely on the computation of the gradient of the local cross correlation [21, 64].

Concerning the regularization of dense displacement fields, we distinguish the approaches based on explicit smoothing of the field, as in Thirion's demons algorithm [81] (we refer to [69] for a variational interpretation of this algorithm), from

those considering an additive term in the global energy, yielding (possibly anisotropic) diffusion terms [12, 91]. For a comparison of these two approaches, we refer to the work of Cachier and Ayache [19, 20].

Typically, differential methods are valid only for small displacements and special techniques are required in order to recover large deformations. For instance Alvarez et al. [6] use a scale-space focusing strategy. Christensen et al. [25] adopt a different approach. They look for a continuously invertible mapping which is obtained by the composition of small displacements. Each small displacement is calculated as the solution of an elliptic PDE describing the non-linear kinematics of fluid-elastic materials under deforming forces given by the matching term (in their case the image-differences). Trouvé [84] has generalized this approach using Lie group ideas on diffeomorphisms. Under a similar formalism, a very general framework which also allows for changes in the intensity values is proposed by Miller and Younes [58].

In this thesis, we focus on the study of a family of functional equations resulting from the minimization of global and local statistical dissimilarity measures. The emphasis is put on to the computation of the first variation of these criteria and on the study of the properties of their gradient operators which are important to establish the well posedness of the minimization flows.

Concerning smoothness of the solution, we consider an energy functional composed of the sum of a matching and a regularization term and restrict our study to regularization terms yielding linear operators. We obtain a large family of matching algorithms, each one implying different a priori knowledge about the smoothness of the deformation and the relation between image intensities. We prove that all these problems have a global solution and that the functional equations governing the minimization are well posed in the sense of Hadamard. Interesting generalizations of these results may be obtained for more complex regularization schemes. In this respect we refer to the work of Weickert and Schnörr [91], Trouvé [84] and Miller and Younes [58]. The main contributions of our work are listed below.

- We propose a unifying framework for a family of variational problems for multi-modal image matching. This framework subsumes block matching algorithmic approaches as well as techniques for non-rigid matching based on the global estimation of the intensity relations.
- We formally compute the gradient of local and global statistical dissimilarity measures, which is an essential step in defining and studying the well posedness of their minimization. Contrary to more standard matching terms like intensity differences or the optical flow constraint, these matching terms are non-local, which makes the standard method of the calculus of variations inapplicable.
- We show that the operators defined by the gradients of these criteria satisfy some Lipschitz-continuity conditions which are required for the well posedness of the

associated matching flows.

Document Layout

This manuscript is divided in three parts. Part I (chapters 1 to 3) is devoted to the description of the basic concepts involved in matching two images using statistical dissimilarity measures and provides an overview of the proposed approach. The conditions for the existence and uniqueness of a solution to the minimization problem are established and two regularization operators are studied by showing that they satisfy the required properties. The only part of the algorithms that is not treated is the study of the matching term, coming from the dissimilarity measure. This is the object of part II (chapters 4 to 6), which studies this term of the functional equations in detail, computing the first variation of the six dissimilarity criteria and establishing their good properties in the sense of the well posedness of the minimization process. Finally, part III (chapters 7 to 9) describes in detail the numerical implementation of the resulting algorithms, and presents several experimental results with real and synthetic deformations, involving both 2D and 3D images. In the following, a detailed summary of each chapter is given.

PART I: The Generic Image Matching Problem

CHAPTER 1

This chapter gives an overview of the type of algorithms studied in the thesis. After providing the formal definition of an image adopted in the sequel, the general matching problem is defined. The chapter continues with a discussion of the statistical similarity criteria and their intuitive behavior. It ends by describing the general framework of the calculus of variations and summarizes the approach followed in the thesis by giving the general form of the functional equations which describe the minimization flows.

CHAPTER 2

This chapter is devoted to the study of the minimization problem introduced in chapter 1, within the abstract framework of functional analysis. The chapter starts with a discussion of the functional spaces considered. Then the existence and uniqueness of several kinds of solutions (weak, strong, classical) to the generic evolution problem is shown assuming Lipschitz-continuity of the matching term and regularization operators generating certain types of contraction semigroups of operators (uniformly continuous, analytical).

CHAPTER 3

This chapter studies the regularization part of the algorithms. Two different families of linear operators are considered, including one which is designed to encourage discontinuities of the displacement field along the edges of the reference image. It is shown that these operators generate uniformly continuous, as well as analytical semigroups of contractions and therefore satisfy the required conditions established in chapter 2.

PART II: Study of Statistical Similarity Measures

CHAPTER 4

This chapter introduces the two classes of matching terms considered, which are called local and global. Their definition is given in terms of non-parametric Parzen-window estimates of the joint intensity distribution from either the whole image or corresponding regions around each pixel (voxel). In each case, three similarity measures are defined: cross-correlation, correlation ratio and mutual information. Existence of minimizers for the energy functional obtained is then shown.

CHAPTER 5

In this chapter, the Euler-Lagrange equations are derived for the six dissimilarity measures. Due to the non-standard form of these functionals, an explicit computation of their Gateaux-derivative is necessary.

CHAPTER 6

This chapter is devoted to showing that the gradients of the statistical criteria computed in chapter 5 satisfy the Lipschitz-continuity conditions established in chapter 2, necessary to assert the well-posedness of the evolution equations.

PART III: Implementation Aspects

CHAPTER 7

This chapter describes the numerical schemes employed in implementing the continuous evolution equations, as well as for interpolating image and gradient values.

CHAPTER 8

This chapter discusses the way in which the different parameters of the algorithms are determined, particularly the smoothing parameter for the Parzen window estimates.

CHAPTER 9

This chapter presents experimental results for all the described algorithms using both real and synthetic data. Examples include 2D images for applications in computer vision and 3D images concerning different medical image modalities.

Part I

**A Generic Image Matching
Problem**

Chapter 1

Overview

This chapter gives an overview of the type of algorithms studied in the thesis. After providing the formal definition of an image adopted in the sequel, the general matching problem is defined. The chapter continues with a discussion of the statistical similarity criteria and their intuitive behavior. It ends by describing the general framework of the calculus of variations and summarizes the approach followed in the thesis by giving the general form of the functional equations which describe the minimization flows.

1.1 Definition of Images

Physically, an image is a set of measurements obtained by integration of some density field, for example irradiance or water concentration, over a finite area (pixel) or volume (voxel). Sometimes images are vector valued, as color images for example. We shall restrict ourselves to scalar images. In a computer, an image appears as a set of scalar values ordered in a two or three-dimensional array. The grey-value obtained involves a neighborhood of a point, and the idea of resolution, or scale, is captured by modeling the physical field as a tempered distribution. In practice, this amounts to defining image derivatives by convolution with the derivative of an appropriate kernel. We will view images as functions defined over a two or three dimensional manifold, usually a bounded domain Ω of \mathbb{R}^n ($n = 2, 3$) with smooth boundary $\partial\Omega$. The range of an image will be considered to be the interval $[0, \mathcal{A}]$.

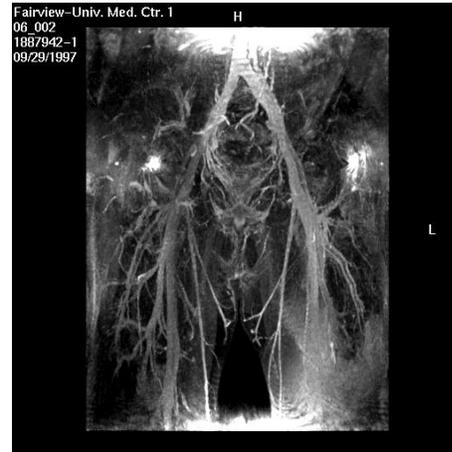
1.2 Image Matching

In many applications, one needs to integrate information coming from different types of sensors, compare data acquired at different times, or put similar structures of two different images into correspondence. These tasks are known respectively as data fu-

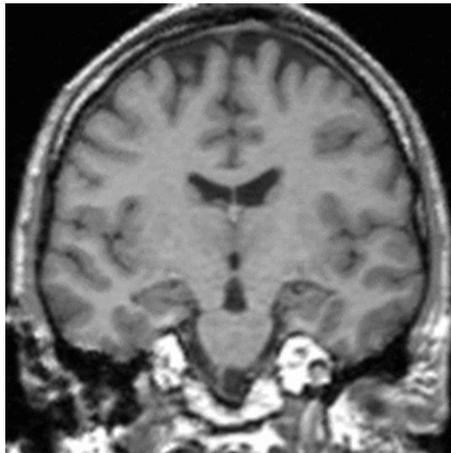
sion, image registration and template matching, and they are all based upon the ability to automatically map points between the respective domains of the images. Additionally, computing the optical flow, reconstructing a 3D scene from (at least) two views, tracking a “feature” or a region in a video sequence and calibrating a camera, also require the ability to establish point correspondences between two images.



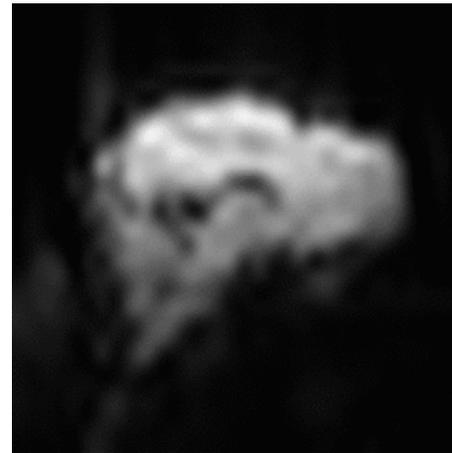
(a) Black and white photography



(b) Magnetic resonance angiography



(c) T1-weighted magnetic resonance



(d) Functional magnetic resonance

Figure 1.1: Examples of different image modalities

The problem can be formulated as follows. Given two sets of points on a manifold (for instance \mathbb{R}^n), we want to be able to automatically put them into correspondence, say by finding a function $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^n$. This function can be constrained in many ways, depending on how much we know about the relation between the two sets. For instance, when matching points from a stereo pair we know that corresponding points should belong to epipolar lines, and that from two views taken with the same center

of perspective (but different viewing orientations), the transformation is a homography [34]. In other cases, other, more complicated functions are needed, but some a priori knowledge may still be available, like the fact that the transformation should be “smooth” and invertible. Consider for instance the images shown in Figure 1.2 on the following page. The first image pair represents a three-dimensional scene with no moving objects, viewed by a projective camera from two different points in space. Consequently, the transformation which links the points in both images is a homography within each plane of the scene. The regions where occlusions occur, are regions where the transformation ϕ is not invertible. The two images of the second example were constructed by calculating and assigning to each pixel its signed distance from two given curves (the curves outlined in red). One possible way of matching points in the first curve with points in the second curve is by matching all the points in these two images. This would require to find a highly nonlinear mapping which however should be smooth and invertible, at least for the points near the curves.

1.3 Multimodality and Statistical Similarity Criteria

The second component in the matching problem (the first one was the nature of the transformation ϕ) is the knowledge about what should be satisfied when two points are to be associated with one another. Coming back to the examples of Figure 1.2, it is clear that for the first image pair a reasonable way of matching the images is by simply comparing the intensities of corresponding pixels. For the second case, since the value of the images is zero for points lying on the curves, it seems also reasonable to match the images by a comparison of the local image intensities.

However, images may be produced by a variety of sensors (Figure 1.1 on the facing page gives some examples), and this simple way of measuring their similarity is no longer adapted. More general ways of comparing the images are therefore needed. This is the role of statistical similarity measures, which have been widely used to cope with the problem of registering different medical image modalities (see the first example of Figure 1.3 on page 35). Nevertheless, these criteria can be used in other situations in which no intensity comparison can be made, even though the acquiring sensors are of similar kind. This is the case for instance when matching images of similar objects which however have different responses to similar lighting conditions (e.g. the two skins with different albedos in the second example of Figure 1.3).

Let us try to give an intuition behind these similarity measures by picturing artificial imaging sensors. The described situation is admittedly far simpler than reality, but the idea behind the similarity criteria can be better grasped in this ideal situation. Formal definitions will be postponed until chapter 4.

Suppose our detectors are sensitive to a physical quantity Q . To fix ideas, we may picture Q as the intensities of a given image (see Figure 1.4 on page 37). We note

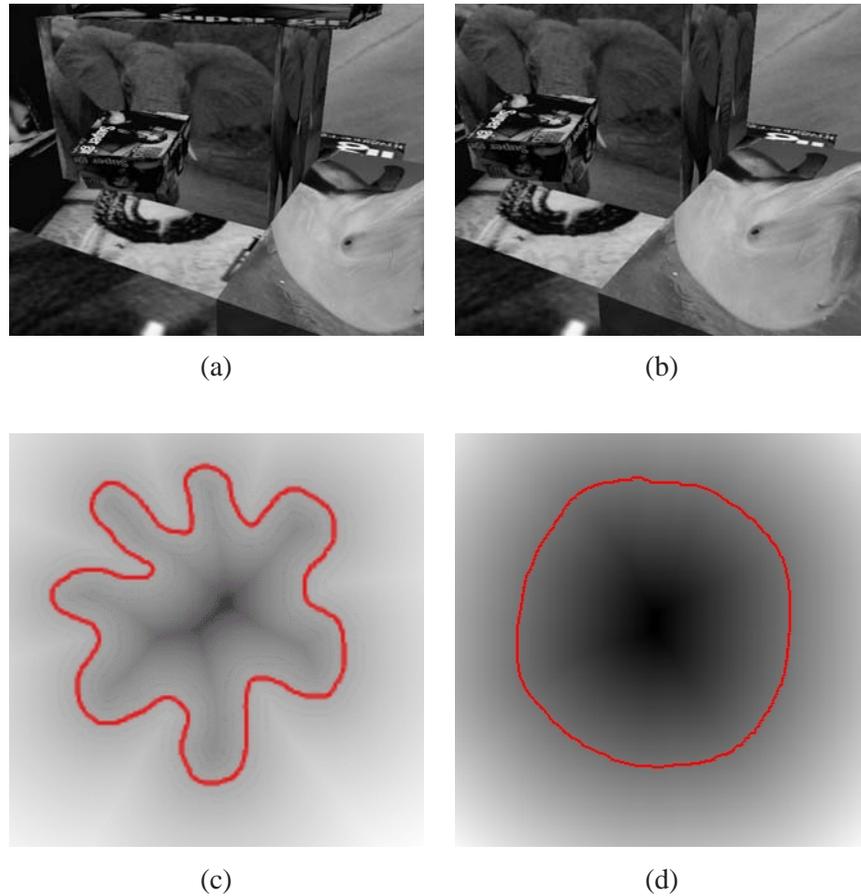


Figure 1.2: Between (a) and (b) the camera has undergone a rigid 3D movement so that, within each plane of the scene, the matching function is a homography. On the other hand, (c) and (d) are constructed as the signed distance functions to the red curves. The matching of these curves requires a highly nonlinear mapping between the two images. The occlusions in the top-row example are regions where the mapping function is not invertible. The mapping between the two curves should on the contrary be invertible everywhere.

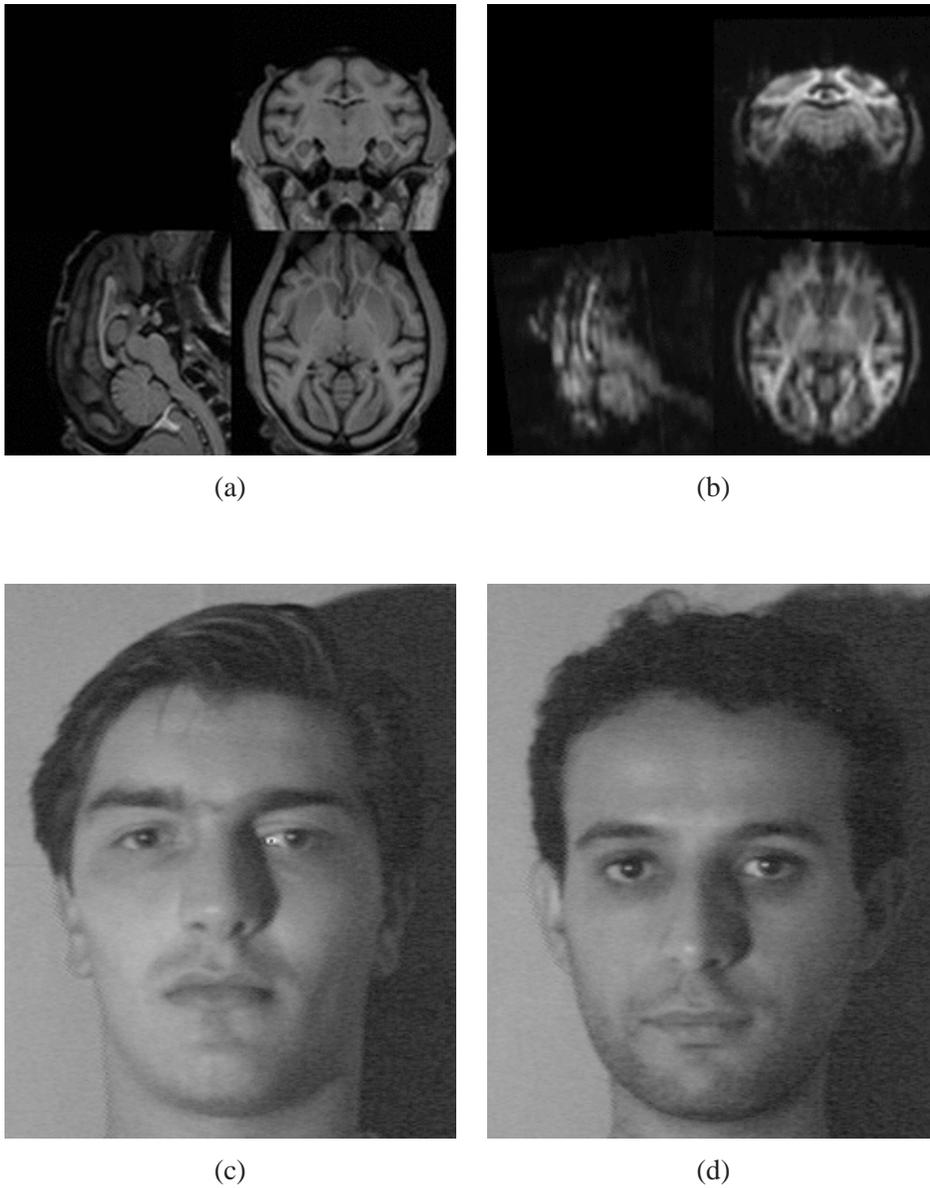


Figure 1.3: Nonrigid “multimodal” matching examples: (a) and (b): T1-weighted anatomical magnetic resonance image (MRI) against functional MRI. (c) and (d) : two human faces (with different skin albedos) under similar illuminating conditions.

the output of two given sensors i_1 and i_2 . If their response is a smooth function of Q , the support of the joint intensity distribution of intensities is generally a curve in the plane $[i_1, i_2]$. A particular case is obtained when one of the responses is an invertible function of Q (say i_1). In this case, the support of the joint distribution has a functional form $f(i_1)$. When both i_1 and i_2 are invertible functions of Q , the support of the joint distribution is also an invertible function and the output of the two sensors may be equalized to yield the same image. This suggests that looking at the joint distribution of intensities and somehow constraining it to be clustered is an appropriate way of matching related outputs.

As will be clear from their expressions, the gradients of the three similarity measures that we consider define three types of clustering processes of the joint distribution according to a hierarchy of constraints on the intensity relations. Roche et al. [75, 73] have clarified the assumptions on which these similarity measures rely by looking for optimal measures from various sensor models. At the most general stage, mutual information is a measure of the statistical dependency between i_1 and i_2 . A more constrained criterion is the correlation ratio, which measures the functional dependency between the intensities. Finally, the cross correlation is still more constrained, as it measures their affine dependency (see Figure 1.5).

1.4 Dense Matching and the Variational Framework

We now summarize the modeling assumptions used in the sequel and define the matching problem in the context of the calculus of variations. We consider two images $I_1^\sigma = I_1 \star G_\sigma$ and $I_2^\sigma = I_2 \star G_\sigma$ at a given scale σ , i.e. resulting from the convolution of two square-integrable functions $I_1 : \mathbb{R}^n \rightarrow \mathbb{R}$ and $I_2 : \mathbb{R}^n \rightarrow \mathbb{R}$ ($n = 2, 3$) with a Gaussian kernel of standard deviation σ . Given a region of interest Ω , a bounded region of \mathbb{R}^n (we may require its boundary $\partial\Omega$ to fulfill some regularity constraints, e.g. that of being of class C^2), we look for a function $\mathbf{h} : \Omega \rightarrow \mathbb{R}^n$ assigning to each point \mathbf{x} in Ω a displacement vector $\mathbf{h}(\mathbf{x}) \in \mathbb{R}^n$. This function is searched for in a set \mathcal{F} of admissible functions such that it minimizes an energy functional $\mathcal{I} : \mathcal{F} \rightarrow \mathbb{R}$ of the form

$$\mathcal{I}(\mathbf{h}) = \mathcal{J}(\mathbf{h}) + \mathcal{R}(\mathbf{h}),$$

where $\mathcal{J}(\mathbf{h})$ measures the “dissimilarity” between I_1^σ and

$$I_2^\sigma \circ (\mathbf{Id} + \mathbf{h})$$

and $\mathcal{R}(\mathbf{h})$ is a measure of the “irregularity” of \mathbf{h} (\mathbf{Id} is the identity mapping of \mathbb{R}^n).

The dissimilarity term will be defined in terms of global or local statistical measures on the intensities of I_1^σ and $I_2^\sigma \circ (\mathbf{Id} + \mathbf{h})$, and the irregularity term will generally be a measure of the variations of \mathbf{h} in Ω . For example if \mathbf{h} is differentiable, $\mathcal{R}(\mathbf{h})$ could

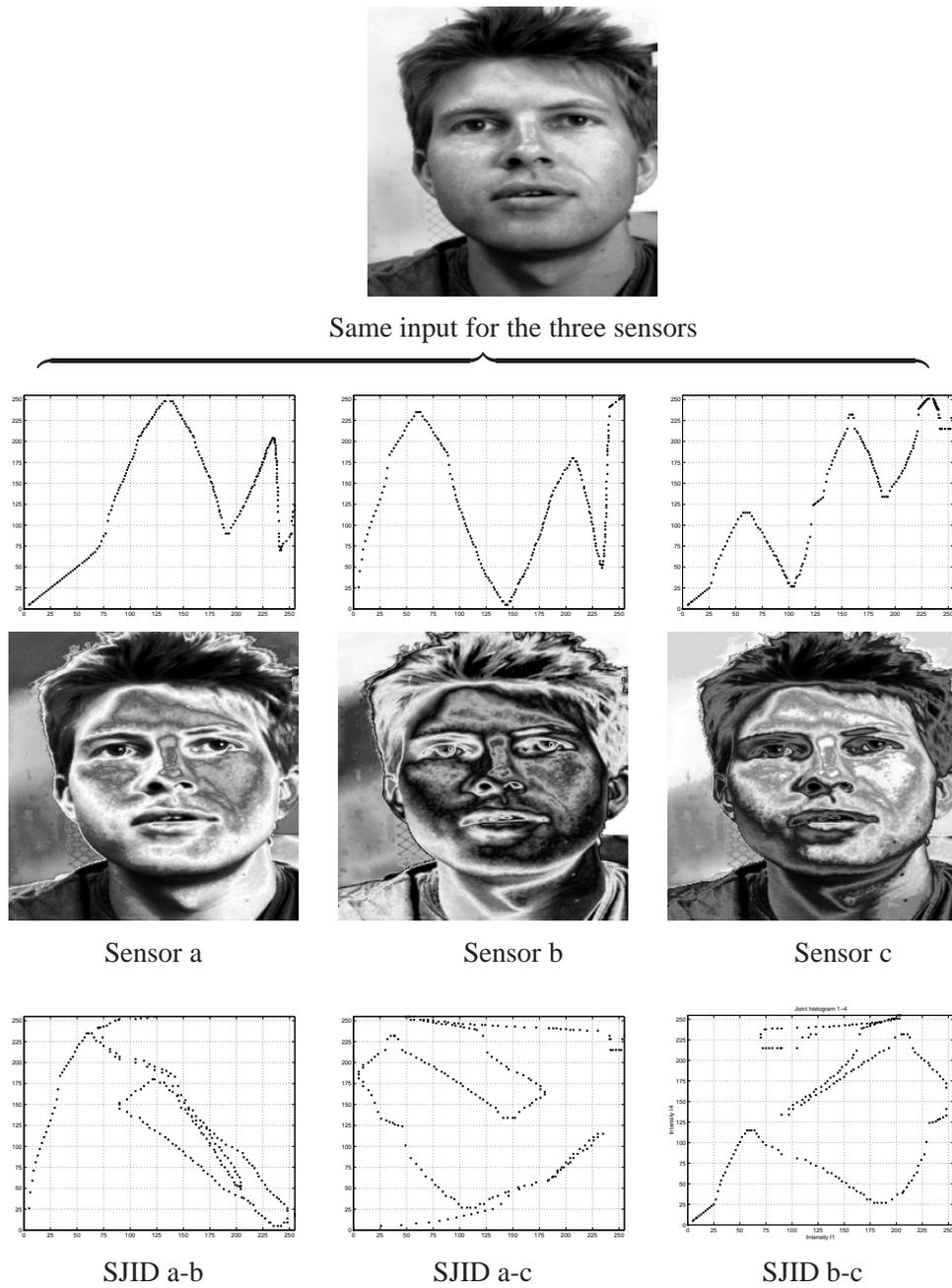


Figure 1.4: Synthetic sensors and the support of the joint intensity distribution (SJID) of their outputs. The second and third rows represent the response and output of three synthetic sensors.

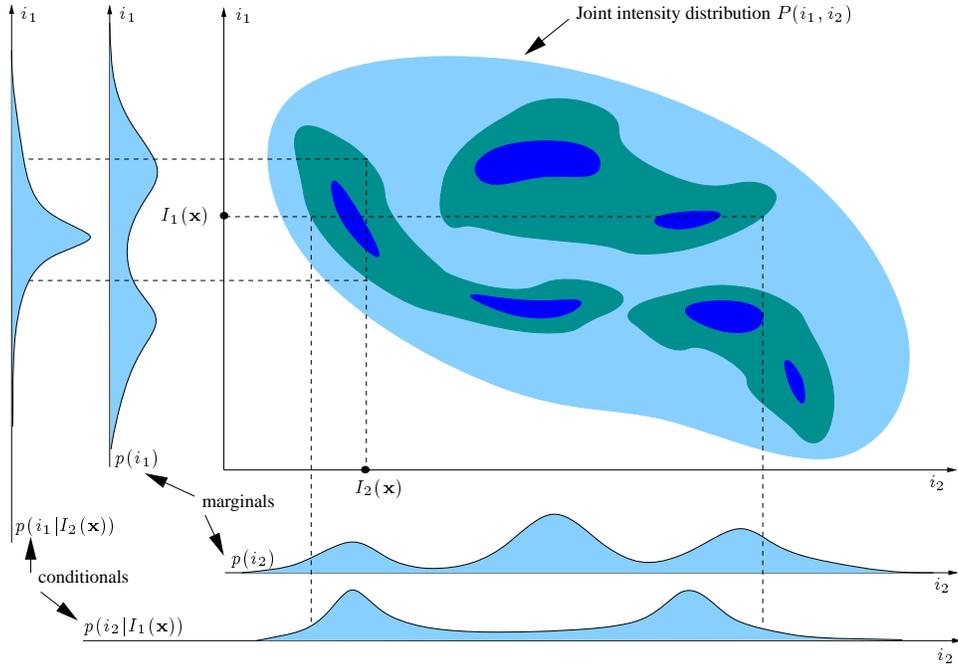


Figure 1.5: Schematic joint intensity distribution. The three criteria give a hierarchy of measures to compare image intensities. The cross correlation measures their affine dependency, so that maximizing this criterion amounts to trying to fit an affine function to the joint density. The correlation ratio measures their functional dependency, so that the optimal density can have the shape of a nonlinear function. Finally, their mutual information gives an estimate of their statistical dependency; maximizing this criterion tends to cluster P .

be defined as a certain norm of its Jacobian $D\mathbf{h}$. In summary, the matching problem is defined as the solution of the following minimization problem:

$$\mathbf{h}^* = \arg \min_{\mathbf{h} \in \mathcal{F}} \mathcal{I}(\mathbf{h}) = \arg \min_{\mathbf{h} \in \mathcal{F}} (\mathcal{J}(\mathbf{h}) + \mathcal{R}(\mathbf{h})). \quad (1.1)$$

Assuming that \mathcal{I} is sufficiently regular, its first variation at $\mathbf{h} \in \mathcal{F}$ in the direction of $\mathbf{k} \in \mathcal{F}$ is defined by

$$\delta \mathcal{I}(\mathbf{h}, \mathbf{k}) = \lim_{\epsilon \rightarrow 0} \frac{\mathcal{I}(\mathbf{h} + \epsilon \mathbf{k}) - \mathcal{I}(\mathbf{h})}{\epsilon} = \left. \frac{d\mathcal{I}(\mathbf{h} + \epsilon \mathbf{k})}{d\epsilon} \right|_{\epsilon=0}.$$

If a minimizer \mathbf{h}^* of \mathcal{I} exists, then $\delta \mathcal{I}(\mathbf{h}^*, \mathbf{k}) = 0$ must hold for every $\mathbf{k} \in \mathcal{F}$. The equations $\delta \mathcal{I}(\mathbf{h}^*, \mathbf{k}) = 0$ are called the Euler-Lagrange equations associated with the energy functional \mathcal{I} . Assuming that \mathcal{F} is a linear subspace of a Hilbert space H , endowed with a scalar product $(\cdot, \cdot)_H$, we define the gradient $\nabla_H \mathcal{I}(\mathbf{h})$ of \mathcal{I} by requiring that

$$\forall \mathbf{k} \in \mathcal{F}, \quad \left. \frac{d\mathcal{I}(\mathbf{h} + \epsilon \mathbf{k})}{d\epsilon} \right|_{\epsilon=0} = (\nabla_H \mathcal{I}(\mathbf{h}), \mathbf{k})_H.$$

The Euler equations are then equivalent to $\nabla_H \mathcal{I}(\mathbf{h}^*) = 0$. Rather than solving them directly (which is usually impossible), the search for a minimizer of \mathcal{I} is done using a “gradient descent” strategy. Given an initial estimate $\mathbf{h}_0 \in H$, we introduce time and a differentiable function, also noted \mathbf{h} from the interval $[0, T]$ into H (we say that $\mathbf{h} \in C^1([0, T]; H)$) and we solve the following initial value problem:

$$\begin{cases} \frac{d\mathbf{h}}{dt} = -\nabla_H \mathcal{I}(\mathbf{h}) = -\left(\nabla_H \mathcal{J}(\mathbf{h}) + \nabla_H \mathcal{R}(\mathbf{h})\right), \\ \mathbf{h}(0)(\cdot) = \mathbf{h}_0(\cdot). \end{cases} \quad (1.2)$$

That is, we start from the initial field \mathbf{h}_0 and follow the gradient of the functional \mathcal{I} (the minus sign is because we are minimizing). The solution of the matching problem is then taken as the asymptotic state (i.e. when $t \rightarrow \infty$) of $\mathbf{h}(t)$, provided that $\mathbf{h}(t) \in \mathcal{F}$ for a sufficiently large t .

The boundary conditions, i.e. the values of $\mathbf{h}(t)(\cdot)$ in $\partial\Omega$, must also be specified. This will be done along with the choice of the space \mathcal{F} of admissible functions in Chapter 3. We assume for the moment (since this is the case we shall treat) that $\nabla_H \mathcal{R}(\mathbf{h})$ is a linear application from a linear subspace of H into H . In Chapter 3, concrete functional spaces \mathcal{F} and H will be chosen and two families of regularization operators will be studied.

The computation and study of the properties of $\nabla_H \mathcal{J}(\mathbf{h})$ for a set of statistical dissimilarity measures will be the object of Part II of this manuscript. In the following chapter, we study the existence and uniqueness of a solution of (1.2) from an abstract viewpoint, by borrowing tools from the theory of semigroups generated by unbounded linear operators on a Hilbert space.

Chapter 2

Study of the Abstract Matching Flow

In the previous chapter, $\nabla_H \mathcal{I}$ was defined by assuming that \mathbf{h} belongs to a Hilbert space denoted H . Consequently, equation (1.2) may be viewed as a first-order ordinary differential equation with values in H . It turns out that studying it from such an abstract viewpoint allows to prove the existence and uniqueness of several types of solutions (mild, strong, classical) of (1.2), by borrowing tools from functional analysis and the theory of semigroups of linear operators. We refer to the books of Brezis [18] and Pazy [68] for formal studies of these subjects. In the present chapter, we study the generic minimization flow (1.2) within this abstract framework. The linear operator $-\nabla_H \mathcal{R}(\mathbf{h})$ defined by the regularization term will be simply noted A and the non-linear matching term $-\nabla_H \mathcal{J}$ will be generically noted F . The unknown of the problem is an H valued function $\mathbf{h} : [0, +\infty[\rightarrow H$ defined on \mathbb{R}^+ . The goal of this chapter is to establish the properties required by A and F in order for equation (1.2), which is now written as a semilinear abstract initial value problem of the form

$$\begin{cases} \frac{d\mathbf{h}}{dt} - A\mathbf{h}(t) = F(\mathbf{h}(t)), & t > 0 \\ \mathbf{h}(0) = u_0 \in H, \end{cases} \quad (2.1)$$

to have a unique solution (in a sense to be defined). That these conditions are met will be the object of Chapter 3 concerning two different families of linear regularization operators A , and of Chapter 6 concerning six different matching functions F .

2.1 Definitions and Notations

We begin by introducing some definitions and notations. H will denote a complex Hilbert space with scalar product $(\cdot, \cdot)_H \in \mathbb{C}$, i.e. satisfying for u and v in H ,

$(u, v)_H = (v, u)_H^*$, where λ^* denotes the complex conjugate of λ . The real and imaginary parts of $\lambda \in \mathbb{C}$ will be noted $\operatorname{Re}(\lambda)$ and $\operatorname{Im}(\lambda)$. The norm of H , induced by the Hilbert product, will be noted $\|\cdot\|_H = (\cdot, \cdot)_H^{1/2}$.

If E and F denote two Banach spaces, a linear operator is any linear application $A : \mathcal{D}(A) \subset E \rightarrow F$ from its domain $\mathcal{D}(A)$, a linear subspace of E , into F . We shall restrict ourselves to densely defined linear operators, i.e. for which $\mathcal{D}(A)$ is dense in E . In the following, we consider a linear operator $A : \mathcal{D}(A) \subset H \rightarrow H$.

The range of A is the linear subspace of H

$$\operatorname{Ran}(A) = \{f \in H : f = Au, u \in \mathcal{D}(A)\}$$

and its graph is the set of pairs

$$\Gamma(A) = \{[u, Au], u \in \mathcal{D}(A)\} \subset H \times H.$$

A is said to be *closed* if $\Gamma(A)$ is a closed subset of $H \times H$. It is said to be *bounded* if there exists $c \geq 0$ such that

$$\|Au\|_H \leq c \|u\|_H, \quad \forall u \in \mathcal{D}(A).$$

The smallest such c will be denoted $\|A\|$. The graph norm of A is the norm $\| \cdot \|_A$ defined, for $u \in \mathcal{D}(A)$, as

$$\|u\|_A = \|u\|_H + \|Au\|_H$$

and its numerical range is the set

$$Q(A) = \{(Au, u)_H, \|u\|_H = 1\} \subset \mathbb{C}.$$

A is said to be *invertible* if, for all $f \in H$, there exists a unique $u \in \mathcal{D}(A)$ such that $Au = f$. It implies that $\operatorname{Ran}(A) = H$. We note $u = A^{-1}f$ and readily verify that A^{-1} is a linear application from H into $\mathcal{D}(A)$. If an invertible operator A is closed, it follows (Proposition 2.3) that A^{-1} is a bounded linear operator.

Finally, if I denotes the identity operator on H , the resolvent set $\rho(A)$ of a closed linear operator A is the set of all $\lambda \in \mathbb{C}$ for which $\lambda I - A$ is invertible, i.e. $(\lambda I - A)^{-1}$ is a bounded linear operator. The family

$$R(\lambda : A) = (\lambda I - A)^{-1}, \quad \lambda \in \rho(A)$$

of bounded linear operators is called the resolvent of A .

2.2 Basic Properties

We now state some basic properties of densely defined closed linear operators that will be useful in the sequel. In all this section, A denotes a densely defined *closed* linear operator from $\mathcal{D}(A) \subset H$ into H .

Proposition 2.1 $\mathcal{D}(A)$, endowed with the graph norm of A , is a Banach space.

Proof : Consider a Cauchy sequence $\{u_n\}$ in $\mathcal{D}(A)$, i.e. such that

$$\|u_n - u_p\|_H + \|Au_n - Au_p\|_H \xrightarrow{n,p \rightarrow \infty} 0. \quad (2.2)$$

We must prove that $\{u_n\}$ converges to $u \in \mathcal{D}(A)$. Because of (2.2), we have that $\|u_n - u_p\|_H \rightarrow 0$ and $\|Au_n - Au_p\|_H \rightarrow 0$, i.e. we have two Cauchy sequences in H , which are convergent since H is complete. We therefore have $[u_n, Au_n] \rightarrow [u, f]$, where $u \in H$ and $f \in H$. Since $\Gamma(A)$ is closed, we have that $[u, f] \in \Gamma(A)$. This means that (a) $f = Au$, which implies that $\|u_n - u\|_A \rightarrow 0$, and (b) $u \in \mathcal{D}(A)$ which completes the proof. \square

We next recall the closed graph theorem.

Theorem 2.2 (Closed graph theorem) *Let E and F be two Banach spaces and let $T : E \rightarrow F$ be a linear operator. If the graph of T is closed then there exists $c > 0$ such that $\|Tu\|_F \leq c \|u\|_E$, i.e. T is continuous.*

Proof : The proof can be found for example in Theorem II.7 of the book of Brezis [18]. \square

The closed graph theorem allows to prove the following.

Proposition 2.3 *If A is invertible then A^{-1} is a bounded linear operator.*

Proof : We have $A^{-1} : H \rightarrow \mathcal{D}(A)$ is a linear application. Since A is closed, $\mathcal{D}(A)$ endowed with the graph norm of A is a Banach space (Proposition 2.1). Now since $\text{Ran}(A) = H$ and $\forall f \in H, A^{-1}Af = f$, we have that

$$\Gamma(A) = \{[u, Au], u \in \mathcal{D}(A)\} = \{[A^{-1}f, f], f \in H\} = \Gamma(A^{-1})$$

and thus A^{-1} is closed. We therefore can apply the closed graph theorem to A^{-1} , which says that there exists $c > 0$ such that

$$\|A^{-1}u\|_H + \|u\|_H \leq c \|u\|_H.$$

This implies that

$$\|A^{-1}u\|_H \leq c \|u\|_H$$

and thus A^{-1} is a bounded linear operator. \square

From Proposition 2.3, the following result readily follows.

Proposition 2.4 *If A is invertible then there exists $c > 0$ such that*

$$\|Au\|_H \geq c \|u\|_H, \quad \forall u \in \mathcal{D}(A).$$

Proof : Since A is invertible, A^{-1} is a bounded linear operator. Therefore there exists $c > 0$ such that

$$\|A^{-1} Au\|_H \leq c \|Au\|_H, \quad \forall u \in \mathcal{D}(A).$$

This completes the proof since $A^{-1} Au = u$. \square

As a direct consequence of Proposition 2.4, we have the following useful result.

Proposition 2.5 *If A is invertible then the graph norm of A , $\| \cdot \|_A$ and the norm $\|A \cdot \|_H$ are equivalent, i.e. there exist $c_1 > 0$ and $c_2 > 0$ such that*

$$\| \|u\|_A c_1 \leq \|Au\|_H \leq c_2 \| \|u\|_A, \quad \forall u \in \mathcal{D}(A).$$

Proof : We have $\| \|u\|_A = \|Au\|_H + \|u\|_H$ and therefore the right part of the inequality is obvious ($c_2 = 1$). For the left part, since A is invertible we apply Proposition 2.4 to A which says that there exists $c > 0$ such that $\|Au\|_H \geq c \| \|u\|_H$. Adding $c \|Au\|_H$ to both sides of this inequality yields the desired estimate. \square

2.3 Semigroups of Linear Operators

Consider a one-parameter family $S(t)$, $0 \leq t \leq +\infty$ of bounded linear operators from H to H . This family is said to be a C_0 semigroup of bounded linear operators if

Definition 2.1

1. $S(0) = I$,
2. $S(t_1 + t_2) = S(t_1) S(t_2)$, $\forall t_1, t_2 \geq 0$. (the semigroup property)
3. $\lim_{t \rightarrow 0^+} S(t)u = u$, $\forall u \in H$.

The Hille-Yosida theorem says that there is a one-to-one correspondence between C_0 semigroups of contractions ($\|S(t)\| \leq 1, \forall t \geq 0$) and maximal monotone operators in a Hilbert space. A linear operator A is maximal monotone if and only if

1. A is monotone: $\operatorname{Re}(Au, u)_H \geq 0$, $\forall u \in \mathcal{D}(A)$,
2. and maximal: $\operatorname{Ran}(I + A) = H$. That is, a linear operator A is maximal if

$$\forall f \in H, \exists u \in \mathcal{D}(A) \text{ such that } u + Au = f.$$

if $-A$ is a maximal monotone operator, A is said to be the infinitesimal generator of the corresponding C_0 semigroup noted $S_A(t)$, $t \geq 0$. The relation between A and $S_A(t)$ is the following. Given $u_0 \in \mathcal{D}(A)$, consider the initial value problem,

$$\begin{cases} \frac{d\mathbf{h}}{dt} - A\mathbf{h}(t) = 0, & t > 0 \\ \mathbf{h}(0) = u_0. \end{cases} \quad (2.3)$$

if $-A$ is a maximal monotone operator, the Hille-Yosida theorem asserts that there exists a unique solution of (2.3), i.e. a unique function $\mathbf{h} : [0, +\infty[\rightarrow H$ such that:

1. $\mathbf{h}(t)$ is continuous and $\mathbf{h}(t) \in \mathcal{D}(A)$ for $t \geq 0$.
2. $\mathbf{h}(t)$ is continuously differentiable for $t \geq 0$.
3. $\mathbf{h}(t)$ satisfies (2.3) for $t \geq 0$.

Moreover, the solution satisfies $\|\mathbf{h}(t)\|_H \leq \|u_0\|_H$ for $t \geq 0$. The first two points above are summarized by saying that

$$\mathbf{h} \in C([0, +\infty[, \mathcal{D}(A)) \cap C^1([0, +\infty[, H).$$

The linear application $S_A(t), \mathcal{D}(A) \rightarrow \mathcal{D}(A)$ is defined by $S_A(t)u_0 = \mathbf{h}(t)$, where $\mathbf{h}(t)$ is the solution of (2.3) at time t . Since $\|S_A(t)u_0\|_H \leq \|u_0\|_H$, it is possible, using the Hahn-Banach theorem and the fact that H is a Hilbert space [18], to extend $S_A(t)$ by continuity and density to a linear continuous operator $H \rightarrow H$. This family of operators, also noted $S_A(t)$, is the C_0 semigroup of contractions corresponding to A .

A property of the C_0 semigroups of bounded operators that we will need later is given next.

Proposition 2.6 *For all $u \in \mathcal{D}(A)$, $S_A(t)u \in \mathcal{D}(A)$ and*

$$\frac{d}{dt} S_A(t) u = A S_A(t) u = S_A(t) A u.$$

Proof : The proof can be found for example in Theorem 1.2.4 of the book of Pazy [68]. \square

We will also make use of analytic semigroups of operators, which are defined as follows. For more details, the interested reader is referred to [68].

Definition 2.2 *Let $\Delta = \{z \in \mathbb{C} : \varphi_1 < \arg z < \varphi_2, \varphi_1 < 0 < \varphi_2\}$ and for $z \in \Delta$, let $S(z)$ be a bounded linear operator. The family $S(z), z \in \Delta$ is an analytic semigroup in Δ if*

1. $z \rightarrow S(z)$ is analytic in Δ .
2. $S(0) = I$ and $\lim_{\substack{z \rightarrow 0 \\ z \in \Delta}} S(z)u = u, \forall u \in H$.
3. $S(z_1 + z_2) = S(z_1) S(z_2) \quad \forall z_1, z_2 \in \Delta$. (the semigroup property)

A semigroup $S(t)$ will be called *analytic* if it is analytic in some sector Δ containing the nonnegative real axis (Figure 2.1). Clearly, the restriction of an analytic semigroup to the real axis is a C_0 semigroup.

We will make use of the following characterization of the infinitesimal generator of an analytic semigroup.

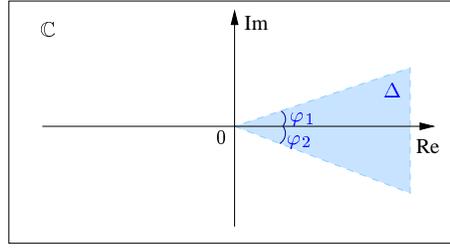


Figure 2.1: The complex plane and the sector Δ of Definition 2.2, containing $[0, +\infty[$. A semigroup $S(t)$ will be called analytic if it is analytic in Δ .

Theorem 2.7 *Let A be the infinitesimal generator of a uniformly bounded C_0 semigroup $S(t)$ and assume $0 \in \rho(A)$. The following statements are equivalent.*

1. $S(t)$ can be extended to an analytic semigroup in a sector $\Delta_\delta = \{z : |\arg z| < \delta\}$ and $\|S(z)\|$ is uniformly bounded in every closed sub-sector $\bar{\Delta}_{\delta'}$, $\delta' < \delta$, of Δ_δ .
2. There exist $0 < \delta < \pi/2$ and $M > 0$ such that

$$\rho(A) \supset \Sigma_\delta = \{\lambda : |\arg \lambda| < \frac{\pi}{2} + \delta\} \cup \{0\}$$

and

$$\|R(\lambda : A)\| \leq \frac{M}{|\lambda|} \quad \text{for } \lambda \in \Sigma_\delta, \lambda \neq 0.$$

Proof : The proof is found in Theorem 2.5.2 of the book of Pazy [68]. \square

Figure 2.2 illustrates the relation between the sectors Σ_δ and Δ_δ of Theorem 2.7.

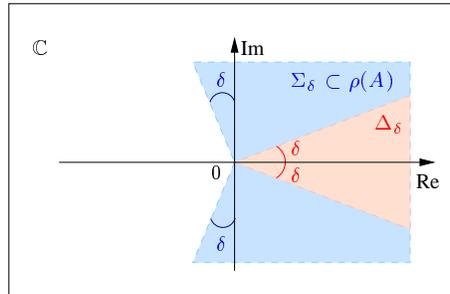


Figure 2.2: The complex plane and the sectors Δ_δ and Σ_δ defined in Theorem 2.7. A C_0 semigroup $S(t)$ can be extended to an analytic semigroup in Δ_δ if the resolvent set $\rho(A)$ of A includes the sector Σ_δ for some $0 < \delta < \pi/2$.

2.4 Solutions of the Abstract Matching Flow

We now consider the initial value problem (2.1):

$$\begin{cases} \frac{d\mathbf{h}}{dt} - A\mathbf{h}(t) = F(\mathbf{h}(t)), & t > 0 \\ \mathbf{h}(0) = u_0 \in H, \end{cases} \quad (2.4)$$

and start by defining four different kinds of solutions.

Definition 2.3 (Global classical solution) A function $\mathbf{h} : [0, +\infty[\rightarrow H$ is a global classical solution of (2.4) if

$$\mathbf{h} \in C([0, +\infty[; H) \cap C^1(]0, +\infty[; H) \cap C(]0, +\infty[; \mathcal{D}(A)),$$

and (2.4) is satisfied for $t > 0$.

Definition 2.4 (Local classical solution) A function $\mathbf{h} : [0, T[\rightarrow H$ is a local classical solution of (2.4) if

$$\mathbf{h} \in C([0, T[; H) \cap C^1(]0, T[; H) \cap C(]0, T[; \mathcal{D}(A)),$$

and (2.4) is satisfied for $0 < t < T$.

Definition 2.5 (Strong solution) A function \mathbf{h} which is differentiable almost everywhere on $[0, T]$ such that $d\mathbf{h}/dt \in L^1(]0, T[; H)$ is called a strong solution of the initial value problem (2.4) if $\mathbf{h}(0) = u_0$ and $d\mathbf{h}/dt - A\mathbf{h}(t) = F(\mathbf{h}(t))$ almost everywhere on $[0, T]$.

Definition 2.6 (Mild solution) A continuous solution \mathbf{h} of the integral equation

$$\mathbf{h}(t) = S_A(t)u_0 + \int_0^t S_A(t-s)F(\mathbf{h}(s)) ds \quad (2.5)$$

is called a mild solution of the initial value problem (2.4).

The last definition is motivated by the following argument. If (2.4) has a classical solution then the H valued function $\mathbf{k}(s) = S_A(t-s)\mathbf{h}(s)$ is differentiable for $0 < s < t$ and (Proposition 2.6):

$$\begin{aligned} \frac{d\mathbf{k}}{ds} &= -AS_A(t-s)\mathbf{h}(s) + S_A(t-s)\mathbf{h}'(s) = \\ &= -AS_A(t-s)\mathbf{h}(s) + S_A(t-s)A\mathbf{h}(s) + S_A(t-s)F(\mathbf{h}(s)) = \\ &= S_A(t-s)F(\mathbf{h}(s)). \end{aligned} \quad (2.6)$$

If $F \circ \mathbf{h} \in L^1([0, T]; H)$ then $S_A(t-s)F(\mathbf{h}(s))$ is integrable and integrating (2.6) from 0 to t yields

$$\mathbf{k}(t) - \mathbf{k}(0) = \mathbf{h}(t) - S_A(t)u_0 = \int_0^t S_A(t-s)F(\mathbf{h}(s)) ds,$$

hence

$$\mathbf{h}(t) = S_A(t)u_0 + \int_0^t S_A(t-s)F(\mathbf{h}(s)) ds.$$

Definition 2.6 is thus natural.

The main goal of this chapter is to establish sufficient conditions on A (in view of the regularization operators that will be studied in the next chapter) and on F in order for the initial value problem (2.4) to have a unique *global classical solution*.

2.4.1 Mild and Strong Solutions

Sufficient conditions on A and F for (2.4) to have a unique mild solution are given by the following theorem.

Theorem 2.8 *Let $F : H \rightarrow H$ be uniformly Lipschitz continuous on H and let $-A$ be a maximal monotone operator. Then the initial value problem (2.4) has a unique mild solution $\mathbf{h} \in C([0, T]; H)$ (given by equation (2.5)). Moreover, the mapping $u_0 \rightarrow \mathbf{h}$ is Lipschitz continuous from H into $C([0, T]; H)$.*

Proof : The proof can be found for example in Theorem 6.1.2 of [68].

□

Since H is a Hilbert space, taking an initial value $u_0 \in \mathcal{D}(A)$ suffices to obtain existence and uniqueness of a strong solution.

Theorem 2.9 *Let F , A and \mathbf{h} be those of Theorem 2.8. Then, if $u_0 \in \mathcal{D}(A)$, \mathbf{h} is the unique strong solution of (2.4).*

Proof : This is a direct consequence of Theorem 6.1.6 in [68] since H , being a Hilbert space, is a reflexive Banach space. □

2.4.2 Classical Solution

To show the existence of a classical solution of (2.4), we will make use of analytic semigroups. If $-A$ generates an analytic semigroup of operators and $0 \in \rho(A)$ (i.e. A is invertible), it is shown in Section 2.2.6 of the book of Pazy [68] that A^α can be defined for $0 < \alpha \leq 1$ and that A^α is a closed linear invertible operator with domain dense in H .

The closedness of A^α implies that its domain, endowed with the graph norm of A^α , is a Banach space (Proposition 2.1). Moreover, since A^α is invertible, its graph norm is equivalent to the norm $\|\cdot\|_\alpha = \|A^\alpha \cdot\|_H$ (Proposition 2.5). Thus $\mathcal{D}(A^\alpha)$, equipped with the norm $\|\cdot\|_\alpha$, is a Banach space which we denote by H_α .

Proposition 2.10 *Let H_α be the Banach space defined above. Then $H_\alpha \subset H$ with continuous embedding.*

Proof : Since A^α is a densely defined closed linear invertible operator from $\mathcal{D}(A^\alpha) \subset H$ into H , we may apply Proposition 2.4 to A^α , which says that there exists $c > 0$ such that $\|A^\alpha u\|_H \geq c \|u\|_H, \forall u \in \mathcal{D}(A^\alpha)$. Therefore there exists $c > 0$ such that

$$\|u\|_H \leq c \|u\|_\alpha, \quad \forall u \in H_\alpha. \quad (2.7)$$

□

The importance of the continuous embedding of H_α into H lies in the fact that if the function F in (2.4) is Lipschitz continuous in H , i.e. if it satisfies for some $L_F > 0$:

$$\|F(u_1) - F(u_2)\|_H \leq L_F \|u_1 - u_2\|_H, \quad \forall u_1, u_2 \in H,$$

then it follows from equation (2.7) that it is also Lipschitz continuous in H_α . Moreover, if F is bounded in H , i.e. if it satisfies for some $K_F > 0$:

$$\|F(u)\|_H \leq K_F, \quad \forall u \in H,$$

then F is well defined in H_α . The main result that we will use is the following, which is a special case of Theorems 6.3.1 and 6.3.3 in [68].

Theorem 2.11 *Assume that A generates an analytic semigroup $S(t)$ satisfying $\|S(t)\| \leq M$ and that $0 \in \rho(A)$, so that the Banach space H_α above is well defined. Assume further that for some $L_F > 0$ and $K_F > 0$ and for $0 \leq \alpha_0 < \alpha < 1$, the function F satisfies*

1. $\|F(u_1) - F(u_2)\|_H \leq L_F \|u_1 - u_2\|_\alpha \quad \forall u_1, u_2 \in H_\alpha.$
2. $\|F(u)\|_H \leq K_F \quad \forall u \in H_\alpha.$

Then for every $u_0 \in H_\alpha$, the initial value problem (2.4) has a unique global classical solution

$$\mathbf{h} \in C([0, +\infty[; H) \cap C([0, +\infty[; \mathcal{D}(A)) \cap C^1([0, +\infty[; H).$$

Moreover, the function $t \rightarrow d\mathbf{h}/dt$ from $]0, +\infty[$ into H_α is Hölder continuous.

Proof : This follows directly from Theorem 6.3.1 (existence of a local classical solution) and Theorem 6.3.3 (extension to a global solution using the boundedness of F) in [68], with $k(t) = K_F$ in Theorem 6.3.3. The Hölder continuity follows from corollary 6.3.2 in [68] which also shows that the Hölder exponent β verifies $0 < \beta < 1 - \alpha$. \square

We are thus interested in the possibility of defining A^α , i.e. that of extending a given C_0 semigroup to an analytic semigroup in some sector around the nonnegative real axis. In order to do that, we will use Theorem 2.7 on page 46, together with the following one.

Theorem 2.12 *Let A be a densely defined closed linear operator in H . Let $\overline{Q(A)}$ be the closure in \mathbb{C} of the numerical range of A and Σ its complement, i.e. $\Sigma = \mathbb{C} \setminus \overline{Q(A)}$. If Σ_0 is a connected component of Σ satisfying $\rho(A) \cap \Sigma_0 \neq \emptyset$ then $\rho(A) \supseteq \Sigma_0$ and*

$$\|R(\lambda : A)\| \leq \frac{1}{d(\lambda : \overline{Q(A)})},$$

where $d(\lambda : \overline{Q(A)})$ is the distance of λ from $\overline{Q(A)}$.

Proof : The proof is found in Theorem 1.3.9 of [68]. \square

In view of the regularization operators studied in the next chapter, the following theorem establishes sufficient assumptions for A to be the infinitesimal generator of an analytic semigroup.

Theorem 2.13 *Let A be the infinitesimal generator of a C_0 semigroup of contractions on H (i.e. let $-A$ be a maximal monotone operator). We assume that A is invertible, i.e. that $0 \in \rho(A)$ and that:*

1. $(Au, v)_H = (u, Av)_H, \forall u, v \in \mathcal{D}(A)$ (A is called symmetric).
2. $\operatorname{Re}(Au, u)_H \leq -c \|u\|_H$ for some $c > 0$ (coerciveness).

Then A is the infinitesimal generator of an analytic semigroup of operators on H .

Proof : From the two assumptions about $(Au, u)_H$, it follows that the numerical range $Q(A) = \{(Au, u)_H, \|u\|_H = 1\}$ of A is a subset of the interval $(-\infty, -c]$ for some $c > 0$ (since the first assumption implies, by the definition of the scalar product, that $(Au, u)_H \in \mathbb{R}, \forall u \in H$). Choosing $0 < \delta < \pi/2$ and denoting $\Sigma_\delta = \{\lambda \in \mathbb{C} : |\arg \lambda| < \pi/2 + \delta\}$ (see Figure 2.3 on the facing page), there exists a constant C_δ such that

$$d(\lambda : Q(A)) \geq C_\delta |\lambda| \quad \text{for all } \lambda \in \Sigma_\delta. \quad (2.8)$$

This is clear from Figure 2.3, where we see that $d(\lambda : Q(A)) \geq d_1 \geq d_0 = |\lambda| \cos \delta$, so we can set $C_\delta = \cos \delta$. Moreover, $d(0 : Q(A)) = c$ and therefore $\Sigma_\delta \subset \mathbb{C} \setminus \overline{Q(A)}$.

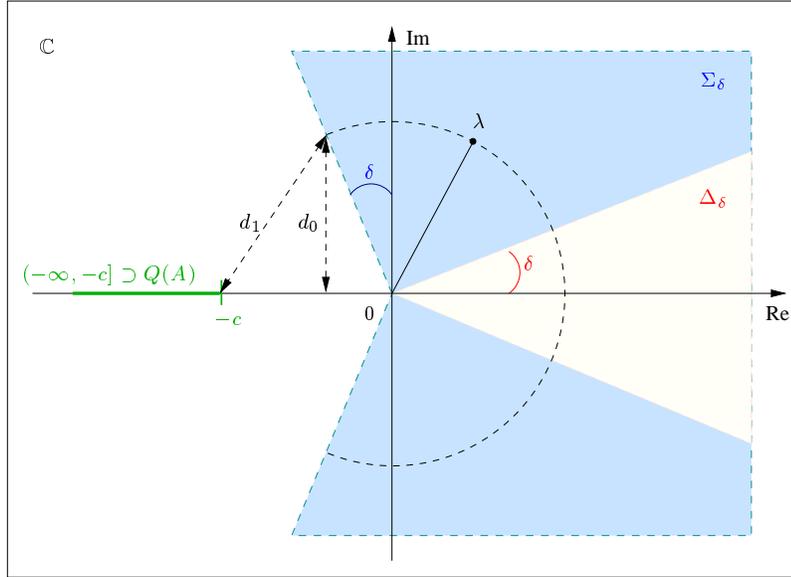


Figure 2.3: The complex plane and the sectors Σ_δ and Δ_δ defined in Theorem 2.13.

The fact that $0 \in \rho(A)$ shows that Σ_δ , which contains 0, is a connected component of $\mathbb{C} \setminus \overline{Q(A)}$ that has a nonempty intersection with $\rho(A)$. This implies by Theorem 2.12 on the facing page that $\rho(A) \supseteq \Sigma_\delta$ and that for every λ in Σ_δ , $\lambda \neq 0$,

$$\|R(\lambda : A)\| \leq \frac{1}{d(\lambda : \overline{Q(A)})} \leq \frac{1}{C_\theta |\lambda|}.$$

We can therefore apply Theorem 2.7 on page 46, which allows us to conclude that the C_0 semigroup generated by A can be extended to an analytic semigroup $S(z)$ in the sector $\Delta_\delta = \{z \in \mathbb{C} : |\arg z| < \delta\}$ (see Figure 2.3), and that $\|S(z)\|$ is uniformly bounded in every closed sub-sector $\bar{\Delta}_{\delta'}$, $\delta' < \delta$, of Δ_δ . \square

As a summary of the results of this chapter, we end it by stating the main result arrived at in the form of a single theorem.

Theorem 2.14 (Main result) *If the following assumptions are satisfied,*

1. *The linear operator $A : \mathcal{D}(A) \subset H \rightarrow H$ is the infinitesimal generator of a C_0 semigroup of contractions on H ($-A$ is maximal monotone) and A is invertible.*
2. *$\forall u, v \in \mathcal{D}(A)$, $(Au, v)_H = (v, Au)_H$ and there exists $c > 0$ such that $(Au, u)_H \leq -c \|u\|_H$.*
3. *F is bounded and Lipschitz continuous in H .*

then, for each $u_0 \in H_\alpha$, the initial value problem (2.4) has a unique global classical solution as defined in Definition 2.3 on page 47.

Proof : Assumptions 1 and 2 are the assumptions of Theorem 2.13 and therefore A generates an analytic semigroup $S(t)$ satisfying $\|S(t)\| \leq M$. This is the first assumption of Theorem 2.11 and, since we assume that A is invertible, we also have that $0 \in \rho(A)$. Now assumption 3 together with equation (2.7) implies the two remaining assumptions of Theorem 2.11 and the proof is complete. \square

Chapter 3

Regularization Operators

This chapter studies the regularization part of the initial value problem (1.2), i.e. the term $\nabla_H \mathcal{R}(\mathbf{h})$. Two families of regularization operators are considered, including one which encourages the preservation of edges of the displacement field along the edges of the reference image. In view of the results of the previous chapter, we choose concrete functional spaces \mathcal{F} and H and specify the domain of the regularization operators. We then show that these operators satisfy the properties of A which are sufficient to assert the existence of a classical solution of (2.1) according to the main result of the previous chapter.

3.1 Functional Spaces

We begin by a brief description of the functional spaces that will be appropriate for our purposes. In doing this, we will make reference to Sobolev spaces, denoted $W^{k,p}(\Omega)$. We refer to the books of Evans [33] and Brezis [18] for formal definitions and in-depth studies of the properties of these functional spaces.

For the definition of $\nabla_H \mathcal{I}$, we use the Hilbert space

$$H = \mathbf{L}^2(\Omega) = (W^{0,2}(\Omega))^n.$$

The regularization functionals that we consider are of the form

$$\mathcal{R}(\mathbf{h}) = \alpha \int_{\Omega} \varphi(D\mathbf{h}(\mathbf{x})) \, d\mathbf{x}, \quad (3.1)$$

where $D\mathbf{h}(\mathbf{x})$ is the Jacobian of \mathbf{h} at \mathbf{x} , φ is a quadratic form of the elements of the matrix $D\mathbf{h}(\mathbf{x})$ and $\alpha > 0$. Therefore the set of admissible functions \mathcal{F} will be contained in the space

$$\mathbf{H}^1(\Omega) = (W^{1,2}(\Omega))^n.$$

Additionally, the boundary conditions for \mathbf{h} will be specified in \mathcal{F} . Assuming for definiteness that $\mathbf{h} = \partial_i \mathbf{h} = 0$ almost everywhere on $\partial\Omega$, we set

$$\mathcal{F} = \mathbf{H}_0^1(\Omega) = (W_0^{1,2}(\Omega))^n.$$

As will be seen, due to the special form of $\mathcal{R}(\mathbf{h})$, the regularization operators are second order differential operators, and we therefore will need the space

$$\mathbf{H}^2(\Omega) = (W^{2,2}(\Omega))^n$$

for the definition of their domain.

3.2 Notations

We introduce in this section some notations that will be used in the sequel. Recall the general form of $\mathcal{R}(\mathbf{h})$ given by (3.1). The quadratic form $\varphi : \mathcal{M}_{n \times n} \rightarrow \mathbb{R}^+$ is defined on the set $\mathcal{M}_{n \times n}$ of $n \times n$ matrices with real coefficients. The components of a vector $\mathbf{x} \in \mathbb{R}^n$ will be noted x_i , and $\partial_i f$ will denote the i^{th} partial derivative of a scalar function f , so that its gradient ∇f is given by $\nabla f = [\partial_1 f, \dots, \partial_n f]^T$. The mapping $\varphi(D\mathbf{h}(\mathbf{x}))$ is given by

$$\varphi(D\mathbf{h}(\mathbf{x})) = \sum_{i,j,k,l} a_{ijkl}(\mathbf{x}) \partial_i \mathbf{h}_j(\mathbf{x}) \partial_k \mathbf{h}_l(\mathbf{x}),$$

where a_{ijkl} are n^4 scalar functions defined in Ω . The divergence of a vector field $\mathbf{h} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is denoted $\text{div}(\mathbf{h}) = \nabla \cdot \mathbf{h} = \sum_i \partial_i \mathbf{h}_i$. For a matrix $\mathbf{T} \in \mathcal{M}_{n \times n}$, composed of row vectors $\mathbf{t}_{\{1\}} \dots \mathbf{t}_{\{n\}}$, i.e. $\mathbf{T} = [\mathbf{t}_{\{1\}} \dots \mathbf{t}_{\{n\}}]^T$, we note

$$\text{div}(\mathbf{T}) = [\nabla \cdot \mathbf{t}_{\{1\}}, \dots, \nabla \cdot \mathbf{t}_{\{n\}}]^T,$$

so that the following relations hold:

$$\begin{cases} \text{div}(D\mathbf{h}^T) = \text{div}((\nabla \cdot \mathbf{h}) \text{Id}) = \nabla(\nabla \cdot \mathbf{h}), \\ \text{div}(D\mathbf{h}) = [\Delta \mathbf{h}_1, \dots, \Delta \mathbf{h}_n]^T \equiv \Delta \mathbf{h}. \end{cases} \quad (3.2)$$

Given $\mathcal{R}(\mathbf{h})$ as in (3.1), the computation of $\nabla_H \mathcal{R}(\mathbf{h})$ is standard:

$$\nabla_H \mathcal{R}(\mathbf{h}) = -\alpha \text{div}(D\varphi(D\mathbf{h})).$$

3.3 Image Driven Anisotropic Diffusion

The first regularization functional that we consider is defined by

$$\varphi_1(D\mathbf{h}) = \frac{1}{2} \text{Tr}(D\mathbf{h} \mathbf{T}_{I_1^\sigma} D\mathbf{h}^T), \quad (3.3)$$

where $\mathbf{T}_{I_1^\sigma}$ is a $n \times n$ symmetric matrix defined at every point of Ω by the following expression:

$$\mathbf{T}_f = \frac{(\lambda + |\nabla f|^2)\mathbf{Id} - \nabla f \nabla f^T}{(n-1)|\nabla f|^2 + n\lambda}, \quad \text{for } f : \mathbb{R}^n \rightarrow \mathbb{R}.$$

This matrix is a regularized projector in the plane perpendicular to ∇f . It was first proposed by Nagel and Enkelmann [62] for computing optical flow while preserving the discontinuities of the deforming template. As pointed out by Alvarez et al. [6], applying the smoothness constraint to the reference image (here I_1^σ) instead of the deforming one (here I_2^σ) allows to avoid artifacts which appear when recovering large displacements. The matrix \mathbf{T}_f has one eigenvector equal to ∇f , while the remaining eigenvectors span the plane perpendicular to ∇f . Its eigenvalues λ_i verify $\sum_i \lambda_i = 1$ independently of ∇f .

It is straightforward to verify that

$$\mathbf{div}(D\varphi_1(D\mathbf{h})) = \begin{pmatrix} \mathbf{div}(\mathbf{T}_{I_1^\sigma} \nabla \mathbf{h}_1) \\ \vdots \\ \mathbf{div}(\mathbf{T}_{I_1^\sigma} \nabla \mathbf{h}_n) \end{pmatrix}.$$

Thus, the regularization operator $\nabla_H \mathcal{R}(\mathbf{h})$ yields a linear diffusion term with \mathbf{T}_f as diffusion tensor. In regions where $\nabla \mathbf{h}_i$ is small compared to the parameter λ in \mathbf{T}_f , the diffusion tensor is almost isotropic and so is the regularization. At the edges of f , where $|\nabla f| \gg \lambda$, the diffusion takes place mainly along these edges. This operator is thus well suited for encouraging large variations of \mathbf{h} along the edges of the reference image I_1^σ .

We define our first regularization operator as follows.

Definition 3.1 *The linear operator $A_1 : \mathcal{D}(A_1) \rightarrow H$ is defined as*

$$\left\{ \begin{array}{l} \mathcal{D}(A_1) = \mathbf{H}_0^1(\Omega) \cap \mathbf{H}^2(\Omega), \\ A_1 \mathbf{h} = \begin{pmatrix} \mathbf{div}(\mathbf{T}_{I_1^\sigma} \nabla \mathbf{h}_1) \\ \vdots \\ \mathbf{div}(\mathbf{T}_{I_1^\sigma} \nabla \mathbf{h}_n) \end{pmatrix}. \end{array} \right.$$

We now check that $-A_1$ is a symmetric maximal monotone invertible operator, applying the standard variational approach [33].

Proposition 3.1 *The operator $(I - A_1)$ defines a bilinear form B_1 on the space $\mathbf{H}_0^1(\Omega)$ which is continuous and coercive (elliptic).*

Proof : Because of the form of the operator A_1 , it is sufficient to work on one of the coordinates and consider the operator $a_1 : \mathcal{D}(a_1) \rightarrow L^2(\Omega)$ defined by

$$a_1 u = \mathbf{div}(\mathbf{T}_{I_1^\sigma} \nabla u),$$

and to show that the operator $u \rightarrow (u - a_1 u)$ defines a bilinear form b_1 on the space $H_0^1(\Omega)$ which is continuous and coercive. Indeed, we define

$$b_1(u, v) = \int_{\Omega} (uv - v \operatorname{div}(\mathbf{T}_{I_1^\sigma} \nabla u)) \, d\mathbf{x}.$$

We integrate by parts the divergence term, use the fact that $v \in H_0^1(\Omega)$, and obtain

$$b_1(u, v) = \int_{\Omega} (uv + \nabla v^T \mathbf{T}_{I_1^\sigma} \nabla u) \, d\mathbf{x}.$$

Because the coefficients of $\mathbf{T}_{I_1^\sigma}$ are all bounded, we obtain, by applying Cauchy-Schwarz:

$$|b_1(u, v)| \leq c_1 \|u\|_{H^1(\Omega)} \|v\|_{H^1(\Omega)},$$

where c_1 is a positive constant, hence continuity.

Because the eigenvalues of the symmetric matrix $\mathbf{T}_{I_1^\sigma}$ are strictly positive, we have $\mathbf{T}_{I_1^\sigma} \geq c_{\mathbf{T}} \mathbf{Id}$, where $c_{\mathbf{T}}$ is a positive constant. This implies that

$$\int_{\Omega} \nabla u^T \mathbf{T}_{I_1^\sigma} \nabla u \, d\mathbf{x} = b_1(u, u) - \|u\|_{L^2(\Omega)}^2 \geq c_{\mathbf{T}} \|\nabla u\|_{L^2(\Omega)}^2,$$

from which it follows that

$$b_1(u, u) \geq c_3 \|u\|_{H^1(\Omega)}^2,$$

for some positive constant $c_3 > 0$ and hence we have coerciveness. \square

We can therefore apply the Lax-Milgram theorem and state the existence and uniqueness of a weak solution in $\mathbf{H}_0^1(\Omega)$ to the equation $\mathbf{h} - A_1 \mathbf{h} = \mathbf{f}$ for all $\mathbf{f} \in \mathbf{L}^2(\Omega)$. Since Ω is regular (in particular C^2), the coefficients of $\mathbf{T}_{I_1^\sigma}$ in $C^1(\overline{\Omega})$, the solution is in $\mathbf{H}_0^1(\Omega) \cap \mathbf{H}^2(\Omega)$ and is a strong solution (see e.g. [33]).

Proposition 3.2 $-A_1$ is a maximal monotone self-adjoint operator from $\mathcal{D}(A_1) = \mathbf{H}_0^1(\Omega) \cap \mathbf{H}^2(\Omega)$ into $\mathbf{L}^2(\Omega)$.

Proof : Monotonicity follows from the coerciveness of B_1 proved in the previous proposition. Maximality also follows from the proof of proposition 3.1. According to the same proposition, we have $\mathcal{D}(A_1) = \mathbf{H}_0^1(\Omega) \cap \mathbf{H}^2(\Omega)$ and $\operatorname{Ran}(\mathbf{Id} - A_1) = H$ (application of the Lax-Milgram theorem). In order to prove that the operator is self-adjoint, it is sufficient, since it is maximal monotone, to prove that it is symmetric ([18], proposition VII.6), i.e. that $(-A_1 \mathbf{h}, \mathbf{k})_{\mathbf{L}^2(\Omega)} = (\mathbf{h}, -A_1 \mathbf{k})_{\mathbf{L}^2(\Omega)}$ and this is obvious from the proof of proposition 3.1. \square

Lemma 3.3 The linear operator αA_1 is invertible for all $\alpha > 0$.

Proof : It is sufficient to show that the equation $\alpha A_1 \mathbf{h} = \mathbf{f}$ has a unique solution for all $\mathbf{f} \in \mathbf{L}^2(\Omega)$. The proof of proposition 3.1 shows that the bilinear form associated to the operator αA_1 is continuous and coercive in $\mathbf{H}^1(\Omega)$, hence the Lax-Milgram theorem tells us that the equation $\alpha A_1 \mathbf{h} = \mathbf{f}$ has a unique weak solution in $\mathbf{H}_0^1(\Omega)$ for all $\mathbf{f} \in \mathbf{L}^2(\Omega)$. Since Ω is regular the weak solution is in $\mathbf{H}_0^1(\Omega) \cap \mathbf{H}^2(\Omega)$ and is a strong solution. \square

3.4 The Linearized Elasticity Operator

The second regularization operator that we propose is inspired from the equilibrium equations of linearized elasticity (we refer to Ciarlet [27] for a formal study of three-dimensional elasticity theory), which are of the form:

$$\mu \Delta \mathbf{h} + (\lambda + \mu) \nabla(\nabla \cdot \mathbf{h}) = 0. \quad (3.4)$$

The constants λ and μ are known as the Lamé coefficients. Rather than modeling the domain Ω as an elastic material¹, the idea in this section is simply to view the left-hand side of (3.4) as a kind of “diffusion” operator and use it as an instance of $\nabla_{\mathcal{I}} \mathcal{R}(\mathbf{h})$ in (1.2). What interests us is the flexibility gained by the relative weight which we can give to the two operators $\Delta \mathbf{h}$ and $\nabla(\nabla \cdot \mathbf{h})$, so that a single parameter (controlling this weight) is a priori needed. Also, in order to assert the existence of a minimizer of the functional \mathcal{I} obtained, it is desirable to define $\varphi(D\mathbf{h})$ in such a way that it is convex in the variable $D\mathbf{h}$. To this end, we consider the one-parameter family ($\frac{1}{2} < \xi \leq 1$) of functions of the form

$$\varphi_2(D\mathbf{h}) = \frac{1}{2} \left(\xi \operatorname{Tr}(D\mathbf{h}^T D\mathbf{h}) + (1 - \xi) \operatorname{Tr}(D\mathbf{h}^2) \right), \quad (3.5)$$

for which we have

$$\operatorname{div}(D\varphi_2(D\mathbf{h})) = \xi \Delta \mathbf{h} + (1 - \xi) \nabla(\nabla \cdot \mathbf{h}).$$

Thus, we define the second regularization operator as follows.

Definition 3.2 *The linear operator $A_2 : \mathcal{D}(A_2) \rightarrow H$ is defined as*

$$\begin{cases} \mathcal{D}(A_2) = \mathbf{H}_0^1(\Omega) \cap \mathbf{H}^2(\Omega), \\ A_2 \mathbf{h} = \xi \Delta \mathbf{h} + (1 - \xi) \nabla(\nabla \cdot \mathbf{h}). \end{cases}$$

for $\frac{1}{2} < \xi \leq 1$.

¹This would require a more complex modeling since true elasticity is always non-linear [27].

We now check that $-A_2$ is a symmetric maximal monotone invertible operator.

Proposition 3.4 *The operator $(I - A_2)$ defines a bilinear form B_2 on the space \mathbf{H}_0^1 which is continuous and coercive (elliptic).*

Proof : We consider the bilinear form C_2 defined as

$$C_2(\mathbf{h}, \mathbf{k}) = - \int_{\Omega} \mathbf{k}^T A_2 \mathbf{h} \, d\mathbf{x},$$

where \mathbf{h} and \mathbf{k} are functions in \mathbf{H}_0^1 . Integrating by parts, we find

$$C_2(\mathbf{h}, \mathbf{k}) = \int_{\Omega} (\xi \operatorname{Tr}(D\mathbf{h}^T D\mathbf{k}) + (1 - \xi) \operatorname{Tr}(D\mathbf{h} D\mathbf{k})) \, d\mathbf{x},$$

and $B_2(\mathbf{h}, \mathbf{k}) = C_2(\mathbf{h}, \mathbf{k}) + \int_{\Omega} \mathbf{h}(\mathbf{x}) \cdot \mathbf{k}(\mathbf{x}) \, d\mathbf{x}$. We have

$$|C_2(\mathbf{h}, \mathbf{k})| \leq \sum_{ijkl} |a_{ijkl}| \int_{\Omega} |\partial_i \mathbf{h}_j \partial_k \mathbf{h}_l| \, d\mathbf{x},$$

where the constants a_{ijkl} are all bounded. Thus, by applying several times Cauchy-Schwarz, we find that

$$|C_2(\mathbf{h}, \mathbf{k})| \leq c_2 \|\mathbf{h}\|_{\mathbf{H}^1(\Omega)} \|\mathbf{k}\|_{\mathbf{H}^1(\Omega)}, \quad c_2 > 0,$$

and hence, using Cauchy-Schwarz again,

$$|B_2(\mathbf{h}, \mathbf{k})| \leq b_2 \|\mathbf{h}\|_{\mathbf{H}^1(\Omega)} \|\mathbf{k}\|_{\mathbf{H}^1(\Omega)}, \quad b_2 > 0.$$

This proves the continuity of B_2 . Next we note that

$$B_2(\mathbf{h}, \mathbf{h}) \geq \xi \|\mathbf{h}\|_{\mathbf{H}^1(\Omega)}^2,$$

which proves the coerciveness of B_2 . \square

Proposition 3.5 *$-A_2$ is a maximal monotone self-adjoint operator from $\mathcal{D}(A_2) = \mathbf{H}_0^1(\Omega) \cap \mathbf{H}^2(\Omega)$ into $\mathbf{L}^2(\Omega)$.*

Proof : Monotonicity follows from the coerciveness of B_2 proved in the previous proposition. More precisely, since $(-A_2 \mathbf{h}, \mathbf{h})_{\mathbf{L}^2(\Omega)} = C_2(\mathbf{h}, \mathbf{h})$, the proof shows that $(-A_2 \mathbf{h}, \mathbf{h})_{\mathbf{L}^2(\Omega)} \geq \xi \int_{\Omega} \operatorname{Tr}(D\mathbf{h}^T D\mathbf{h}) \, d\mathbf{x} \geq 0$.

Regarding maximality, proposition 3.4 shows that the bilinear form B_2 associated to the operator $\operatorname{Id} - A_2$ is continuous and coercive in $\mathbf{H}^1(\Omega)$. We can therefore apply the Lax-Milgram theorem and state the existence and uniqueness of a weak solution in $\mathbf{H}_0^1(\Omega)$ of the equation $\mathbf{h} - A_2 \mathbf{h} = \mathbf{f}$ for all $\mathbf{f} \in \mathbf{L}^2(\Omega)$. Since Ω is regular (in

particular C^2), the solution is in $\mathbf{H}_0^1(\Omega) \cap \mathbf{H}^2(\Omega)$ and is a strong solution (see e.g. [27], Theorem 6.3-6).

Therefore we have $\mathcal{D}(A_2) = \mathbf{H}_0^1(\Omega) \cap \mathbf{H}^2(\Omega)$ and $\text{Ran}(I - A_2) = H$. Finally, $-A_1$ is self-adjoint for the same reasons as those indicated in the proof of proposition 3.5. \square

Lemma 3.6 *The linear operator αA_2 is invertible for all $\alpha > 0$.*

Proof : It is sufficient to show that the equation $\alpha A_2 \mathbf{h} = \mathbf{f}$ has a unique solution for all $\mathbf{f} \in \mathbf{L}^2(\Omega)$. The proof of proposition 3.4 shows that the bilinear form associated to the operator αA_2 is continuous and coercive in $\mathbf{H}^1(\Omega)$, hence the Lax-Milgram theorem tells us that the equation $\alpha A_2 \mathbf{h} = \mathbf{f}$ has a unique weak solution in $\mathbf{H}_0^1(\Omega)$ for all $\mathbf{f} \in \mathbf{L}^2(\Omega)$. Since Ω is regular the weak solution is in $\mathbf{H}_0^1(\Omega) \cap \mathbf{H}^2(\Omega)$ and is a strong solution. \square

3.5 Existence of Minimizers

Having defined the regularization functionals, we discuss in this section the existence of minimizers of the global energy functional

$$\mathcal{I}(\mathbf{h}) = \mathcal{J}(\mathbf{h}) + \alpha \int_{\Omega} \varphi(D\mathbf{h}(\mathbf{x})) \, d\mathbf{x}, \quad (3.6)$$

where φ is either φ_1 defined in Equation (3.3) or φ_2 defined in Equation (3.5). We assume that $\mathcal{J}(\mathbf{h})$ is continuous in \mathbf{h} , and bounded below. These properties will be shown for the statistical dissimilarity functionals $\mathcal{J}(\mathbf{h})$ that we study in Part II. In the following, we use the notion of weak convergence, defined as follows (see e.g. [33]).

Definition 3.3 *Let E be a Banach space and let E^* be its dual. We say that a sequence $\{\mathbf{h}_k\} \subset E$ weakly converges to $\mathbf{h} \in E$, written*

$$\mathbf{h}_k \rightharpoonup \mathbf{h},$$

if $\langle \mathbf{k}^, \mathbf{h}_k \rangle \rightarrow \langle \mathbf{k}^*, \mathbf{h} \rangle$ for each bounded linear functional $\mathbf{k}^* \in E^*$.*

The main result that we will use is given by the following theorem, found in Ciarlet [27].

Theorem 3.7 *Let Ω be a bounded open subset of \mathbb{R}^n and $\beta \in \mathbb{R}$. Assume that the function $\varphi : \Omega \times \mathbb{R}^\mu \rightarrow [\beta, \infty]$ satisfies the following two conditions:*

1. $\varphi(\mathbf{x}, \cdot) : \mathbf{u} \in \mathbb{R}^\mu \rightarrow \varphi(\mathbf{x}, \mathbf{u})$ is convex and continuous for almost all $\mathbf{x} \in \Omega$.

2. $\varphi(\cdot, \mathbf{u}) : \mathbf{x} \in \Omega \rightarrow \varphi(\mathbf{x}, \mathbf{u})$ is measurable for all $\mathbf{u} \in \mathbb{R}^d$.

Then

$$\mathbf{u}_k \rightharpoonup \mathbf{u} \text{ in } \mathbf{L}^1(\Omega) \Rightarrow \int_{\Omega} \varphi(\mathbf{u}) \, d\mathbf{x} \leq \liminf_{k \rightarrow \infty} \int_{\Omega} \varphi(\mathbf{u}_k) \, d\mathbf{x}.$$

Proof : The proof is found in Theorem 7.3-1 of [27]. \square

The following theorem is a slight modification of Theorem 7.3-2 in [27], which assumes that $\mathcal{J}(\mathbf{h})$ is a linear continuous functional.

Theorem 3.8 Given $\mathcal{I}(\mathbf{h})$ as in (3.6), assume that φ is convex and coercive, i.e. that there exist $\alpha > 0$ and β such that

$$\varphi(\mathbf{F}) \geq \alpha \|\mathbf{F}\|^2 + \beta, \quad \text{for all } \mathbf{F} \in \mathcal{M}_{n \times n}.$$

Assume further that $\mathcal{J}(\mathbf{h})$ is continuous in \mathbf{h} and bounded below, and that

$$\inf_{\mathbf{k} \in \mathbf{H}_0^1(\Omega)} \mathcal{I}(\mathbf{k}) < \infty.$$

Then there exists at least one function $\mathbf{h} \in \mathbf{H}_0^1(\Omega)$ satisfying

$$\mathbf{h} = \inf_{\mathbf{k} \in \mathbf{H}_0^1(\Omega)} \mathcal{I}(\mathbf{k}).$$

Proof : First, by the coerciveness of the function φ and the fact that $\mathcal{J}(\mathbf{h})$ is bounded below, we have (using the inequality of Poincaré)

$$\mathcal{I}(\mathbf{h}) \geq c \|\mathbf{h}\|_{\mathbf{H}^1(\Omega)}^2 + d,$$

for all $\mathbf{h} \in \mathbf{H}_0^1(\Omega)$ and some constants $c > 0$ and d .

Let $\{\mathbf{h}_k\}$ be a minimizing sequence for \mathcal{I} , i.e.

$$\mathbf{h}_k \in \mathbf{H}_0^1(\Omega) \quad \forall k, \quad \text{and} \quad \lim_{k \rightarrow \infty} \mathcal{I}(\mathbf{h}_k) = \inf_{\mathbf{k} \in \mathbf{H}_0^1(\Omega)} \mathcal{I}(\mathbf{k}) = m.$$

The assumption that $m < \infty$ and the relation $\mathcal{I}(\mathbf{k}) \rightarrow \infty$ as $\|\mathbf{k}\|_{\mathbf{H}^1(\Omega)} \rightarrow \infty$ imply together that $\{\mathbf{h}_k\}$ is bounded in the reflexive Banach space $\mathbf{H}^1(\Omega)$. Hence $\{\mathbf{h}_k\}$ contains a subsequence $\{\mathbf{h}_p\}$ that weakly converges to an element $\mathbf{h} \in \mathbf{H}^1(\Omega)$. The closed convex set $\mathbf{H}_0^1(\Omega)$ is weakly closed and thus the weak limit \mathbf{h} belongs to $\mathbf{H}_0^1(\Omega)$. The fact that $\mathbf{h}_p \rightharpoonup \mathbf{h}$ in $\mathbf{H}^1(\Omega)$ implies that $D\mathbf{h}_p \rightharpoonup D\mathbf{h}$ in $\mathbf{L}^2(\Omega)$ and, since Ω is bounded (which implies that $\mathbf{L}^\infty(\Omega) \subset \mathbf{L}^2(\Omega)$), we have

$$D\mathbf{h}_p \rightharpoonup D\mathbf{h} \text{ in } \mathbf{L}^2(\Omega) \Rightarrow D\mathbf{h}_p \rightharpoonup D\mathbf{h} \text{ in } \mathbf{L}^1(\Omega).$$

We conclude from Theorem 3.7 on the preceding page that

$$\int_{\Omega} \varphi(D\mathbf{h}) \, d\mathbf{x} \leq \liminf_{p \rightarrow \infty} \int_{\Omega} \varphi(D\mathbf{h}_p) \, d\mathbf{x}.$$

Since \mathcal{J} is continuous, $\mathcal{J}(\mathbf{h}) = \lim_{p \rightarrow \infty} \mathcal{J}(\mathbf{h}_p)$ and thus $\mathcal{I}(\mathbf{h}) \leq \liminf_{p \rightarrow \infty} \mathcal{I}(\mathbf{h}_p) = m$. But since $\mathbf{h} \in \mathbf{H}_0^1(\Omega)$, we also have $\mathcal{I}(\mathbf{h}) \geq m$ and consequently

$$\mathcal{I}(\mathbf{h}) = m = \inf_{\mathbf{k} \in \mathbf{H}_0^1(\Omega)} \mathcal{I}(\mathbf{k}).$$

□

We now check that φ_1 and φ_2 satisfy the hypotheses of Theorem 3.8. For the case of φ_1 , we consider each of its scalar components since this separation is possible. As pointed out in [6], because of the smoothness of $\partial_t I_1^\sigma$, $\mathbf{T}_{I_1^\sigma}$ has strictly positive eigenvalues and therefore, clearly,

Proposition 3.9 *The mapping*

$$\begin{aligned} \varphi_1 : \quad \mathbb{R}^n &\mapsto \mathbb{R}^+ \\ \mathbf{X} &\mapsto \mathbf{X} \mathbf{T}_{I_1^\sigma} \mathbf{X}^T \end{aligned}$$

is convex.

The coerciveness of φ_1 readily follows.

Proposition 3.10 *The functional*

$$\mathcal{R}_1(\mathbf{h}) = \int_{\Omega} \varphi_1(D\mathbf{h}(\mathbf{x})) \, d\mathbf{x},$$

satisfies the coerciveness inequality, i.e. $\exists c_1 > 0, c_2 \geq 0$ such that:

$$\varphi_1(D\mathbf{h}(\mathbf{x})) \geq c_1 |D\mathbf{h}|^2 - c_2.$$

Proof : We have

$$\nabla u^T \mathbf{T}_{I_1^\sigma} \nabla u \geq \theta |\nabla u|^2 \quad \forall \mathbf{x} \in \Omega$$

Where $\theta > 0$ is the smallest eigenvalue of $\mathbf{T}_{I_1^\sigma}$. □

We now turn to φ_2 .

Proposition 3.11 *The mapping*

$$\begin{aligned} \varphi_2 : \quad \mathcal{M}^{n \times n} &\mapsto \mathbb{R}^+ \\ \mathbf{X} &\mapsto \xi \operatorname{Tr}(\mathbf{X}^T \mathbf{X}) + (1 - \xi) \operatorname{Tr}(\mathbf{X}^2), \end{aligned}$$

is convex.

Proof : We write φ_2 as a quadratic form of the components X_k of \mathbf{X} ,

$$\varphi_2(\mathbf{X}) = \sum_i^{n^2} \sum_j^{n^2} a_{ij} X_i X_j$$

and notice that the smallest eigenvalue of the matrix a_{ij} is equal to $2\xi - 1$. The result follows from the fact that $\frac{1}{2} < \xi \leq 1$. \square

Proposition 3.12 *The functional*

$$\mathcal{R}_2(\mathbf{h}) = \int_{\Omega} \varphi_2(D\mathbf{h}(\mathbf{x})) \, d\mathbf{x},$$

satisfies the coerciveness inequality, i.e. $\exists c_1 > 0, c_2 \geq 0$ such that:

$$\varphi_2(D\mathbf{h}(\mathbf{x})) \geq c_1 |D\mathbf{h}|^2 - c_2.$$

Proof : We choose c_1 equal to the smallest eigenvalue of φ_2 and $c_2 = 0$. \square

Part II

Study of Statistical Similarity Measures

Chapter 4

Definition of the Statistical Measures

As mentioned before, a general way of comparing the intensities of two images is by using some statistical or information-theoretic similarity measures. Among numerous criteria, the cross correlation, the correlation ratio and the mutual information provide us with a convenient hierarchy in the relation they assume between intensities [75, 73].

The cross correlation has been widely used as a robust comparison function for image matching. Within recent energy-minimization approaches relying on the computation of its gradient, we can mention for instance the works of Faugeras and Keriven [36], Cachier and Pennec [21] and Netsch et al. [64]. The cross correlation is the most constrained of the three criteria, as it is a measure of the *affine* dependency between the intensities.

The correlation ratio was introduced by Roche et.al [76, 77] as a similarity measure for multi-modal registration. This criterion relies on a slightly different notion of similarity. From its definition given two random variables X and Y ,

$$\mathbf{CR} = \frac{\mathbf{Var}[\mathbf{E}[X|Y]]}{\mathbf{Var}[X]}, \quad (4.1)$$

the correlation ratio can intuitively be described as the proportion of energy in X which is “*explained*” by Y . More formally, this measure is bounded ($0 \leq \mathbf{CR} \leq 1$) and expresses the level of *functional* dependence between X and Y :

$$\begin{cases} \mathbf{CR} = 1 & \Leftrightarrow \exists \phi X = \phi(Y) \\ \mathbf{CR} = 0 & \Leftrightarrow \mathbf{E}[X|Y] = \mathbf{E}[X]. \end{cases}$$

The concept of mutual information is borrowed from information theory, and was introduced in the context of multi-modal registration by Viola and Wells III [88]. Given

two random variables X and Y , their mutual information is defined as

$$\mathbf{MI} = \mathcal{H}[X] + \mathcal{H}[Y] - \mathcal{H}[X, Y],$$

where \mathcal{H} stands for the differential entropy. The mutual information is positive and symmetric, and measures how the intensity distributions of two images fail to be independent.

We analyze these criteria from two different perspectives, namely that of computing them globally for the entire image, or locally within corresponding regions. Both types of similarity functionals are based upon the use of an estimate of the joint probability of the grey levels in the two images. This joint probability, noted $R_{\mathbf{h}}(i_1, i_2)$, is estimated by the Parzen window method [67]. It depends upon the mapping \mathbf{h} since we estimate the joint probability distribution between the images $I_2^\sigma \circ (\mathbf{Id} + \mathbf{h})$ and I_1^σ . To be compatible with the scale-space idea and for computational convenience, we choose a Gaussian window with variance $\beta > 0$ as the Parzen kernel. We will often use the notation $\mathbf{i} \equiv (i_1, i_2)$ and

$$G_\beta(\mathbf{i}) = g_\beta(i_1)g_\beta(i_2) = \frac{1}{2\pi\beta} \exp\left(-\frac{|\mathbf{i}|^2}{2\beta}\right) = \frac{1}{\sqrt{2\pi\beta}} \exp\left(-\frac{i_1^2}{2\beta}\right) \frac{1}{\sqrt{2\pi\beta}} \exp\left(-\frac{i_2^2}{2\beta}\right).$$

Notice that G_β and all its partial derivatives are bounded and Lipschitz continuous. We will need in Chapter 6 the infinite norms $\|g_\beta\|_\infty$ and $\|g'_\beta\|_\infty$. For conciseness, we will sometimes use the following notation when making reference to a pair of grey-level intensities at a point \mathbf{x} :

$$\mathbf{I}_{\mathbf{h}}(\mathbf{x}) \equiv (I_1^\sigma(\mathbf{x}), I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x}))).$$

4.1 Global Criteria

We note $X_{I_1^\sigma}$ the random variable whose samples are the values $I_1^\sigma(\mathbf{x})$ and $X_{I_2^\sigma, \mathbf{h}}$ the random variable whose samples are the values $I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x}))$.

The joint probability density function of $X_{I_1^\sigma}^g$ and $X_{I_2^\sigma, \mathbf{h}}^g$ (the upper index g stands for global) is defined by the function $P_{\mathbf{h}} : \mathbb{R}^2 \rightarrow [0, 1]$:

$$P_{\mathbf{h}}(\mathbf{i}) = \frac{1}{|\Omega|} \int_{\Omega} G_\beta(\mathbf{I}_{\mathbf{h}}(\mathbf{x}) - \mathbf{i}) d\mathbf{x}. \quad (4.2)$$

Notice that the usual property $\int_{\mathbb{R}} P_{\mathbf{h}}(\mathbf{i}) d\mathbf{i} = 1$ holds true.

With the help of the estimate (4.2), we define the cross correlation between the two images $I_2^\sigma \circ (\mathbf{Id} + \mathbf{h})$ and I_1^σ , noted $\mathbf{CC}^g(\mathbf{h})$, the correlation ratio, noted $\mathbf{CR}^g(\mathbf{h})$ and the mutual information, noted $\mathbf{MI}^g(\mathbf{h})$. In order to do this we need to introduce more random variables besides $X_{I_1^\sigma}^g$ and $X_{I_2^\sigma, \mathbf{h}}^g$. They are summarized in Table 4.1. We have introduced in this table the conditional law of $X_{I_2^\sigma, \mathbf{h}}^g$ with respect to $X_{I_1^\sigma}^g$, noted $P_{\mathbf{h}}(i_2|i_1)$:

$$P_{\mathbf{h}}(i_2|i_1) = \frac{P_{\mathbf{h}}(\mathbf{i})}{p(i_1)}, \quad (4.3)$$

Random variable	Value	PDF
$(X_{I_1^\sigma}^g, X_{I_2^\sigma, \mathbf{h}}^g)$	(\mathbf{i})	$P_{\mathbf{h}}(\mathbf{i})$
$X_{I_1^\sigma}^g$	i_1	$p(i_1) = \int_{\mathbb{R}} P_{\mathbf{h}}(\mathbf{i}) di_2$
$X_{I_2^\sigma, \mathbf{h}}^g$	i_2	$p_{\mathbf{h}}(i_2) = \int_{\mathbb{R}} P_{\mathbf{h}}(\mathbf{i}) di_1$
$\mathbf{E}[X_{I_2^\sigma, \mathbf{h}}^g X_{I_1^\sigma}^g]$	$\mu_{2 1}(i_1, \mathbf{h}) \equiv \int_{\mathbb{R}} i_2 P_{\mathbf{h}}(i_2 i_1) di_2$	$p(i_1)$
$\mathbf{Var}[X_{I_2^\sigma, \mathbf{h}}^g X_{I_1^\sigma}^g]$	$v_{2 1}(i_1, \mathbf{h}) \equiv \int_{\mathbb{R}} i_2^2 P_{\mathbf{h}}(i_2 i_1) di_2 - \mu_{2 1}(i_1, \mathbf{h})^2$	$p(i_1)$

Table 4.1: Random variables: global case.

and the conditional expectation $\mathbf{E}[X_{I_2^\sigma, \mathbf{h}}^g | X_{I_1^\sigma}^g]$ of the intensity in the second image $I_2^\sigma(\mathbf{Id} + \mathbf{h})$ conditionally to the intensity in the first image I_1^σ . We note the value of this random variable $\mu_{2|1}(i_1, \mathbf{h})$, indicating that it depends on the intensity value i_1 and on the field \mathbf{h} . Similarly the conditional variance of the intensity in the second image conditionally to the intensity in the first image is noted $\mathbf{Var}[X_{I_2^\sigma, \mathbf{h}}^g | X_{I_1^\sigma}^g]$ and its value is abbreviated $v_{2|1}(i_1, \mathbf{h})$. The mean and variance of the images will also be used. Note that these are not random variables and that, for the second image, they are functions of \mathbf{h} :

$$\mu_2(\mathbf{h}) \equiv \int_{\mathbb{R}} i_2 p_{\mathbf{h}}(i_2) di_2, \quad (4.4)$$

$$v_2(\mathbf{h}) \equiv \int_{\mathbb{R}} i_2^2 p_{\mathbf{h}}(i_2) di_2 - (\mu_2(\mathbf{h}))^2. \quad (4.5)$$

Their counterparts for the first image do not depend on \mathbf{h} :

$$\mu_1 \equiv \int_{\mathbb{R}} i_1 p(i_1) di_1, \quad (4.6)$$

$$v_1 \equiv \int_{\mathbb{R}} i_1^2 p(i_1) di_1 - (\mu_1)^2. \quad (4.7)$$

The covariance of $X_{I_1^\sigma}^g$ and $X_{I_2^\sigma, \mathbf{h}}^g$ will be noted

$$v_{1,2}(\mathbf{h}) \equiv \int_{\mathbb{R}^2} i_1 i_2 P_{\mathbf{h}}(\mathbf{i}) d\mathbf{i} - \mu_1 \mu_2(\mathbf{h}). \quad (4.8)$$

The three similarity measures may now be defined in terms of these quantities¹:

$$\mathbf{CC}^g(\mathbf{h}) = \frac{v_{1,2}(\mathbf{h})^2}{v_1 v_2(\mathbf{h})}, \quad (4.9)$$

$$\mathbf{CR}^g(\mathbf{h}) = 1 - \frac{1}{v_2(\mathbf{h})} \int_{\mathbb{R}} v_{2|1}(i_1, \mathbf{h}) p(i_1) di_1, \quad (4.10)$$

$$\mathbf{MI}^g(\mathbf{h}) = \int_{\mathbb{R}^2} P_{\mathbf{h}}(\mathbf{i}) \log \frac{P_{\mathbf{h}}(\mathbf{i})}{p(i_1)p_{\mathbf{h}}(i_2)} d\mathbf{i}. \quad (4.11)$$

The three criteria are positive and should be maximized with respect to the field \mathbf{h} . Therefore we propose the following definition.

Definition 4.1 *The three global dissimilarity measures based on the cross correlation, the correlation ratio and the mutual information are as follows:*

$$\mathcal{J}_{\mathbf{CC}^g}(\mathbf{h}) = -\mathbf{CC}^g(\mathbf{h}),$$

$$\mathcal{J}_{\mathbf{CR}^g}(\mathbf{h}) = -\mathbf{CR}^g(\mathbf{h}) + 1,$$

$$\mathcal{J}_{\mathbf{MI}^g}(\mathbf{h}) = -\mathbf{MI}^g(\mathbf{h}).$$

Note that this definition shows that the mappings $\mathbf{h} \rightarrow \mathcal{J}_{\mathbf{CC}^g}(\mathbf{h})$, $\mathbf{h} \rightarrow \mathcal{J}_{\mathbf{CR}^g}(\mathbf{h})$ and $\mathbf{h} \rightarrow \mathcal{J}_{\mathbf{MI}^g}(\mathbf{h})$ are not of the form $\mathbf{h} \rightarrow \int_{\Omega} L(\mathbf{h}(\mathbf{x})) d\mathbf{x}$, for some smooth function $L : \mathbb{R}^n \rightarrow \mathbb{R}$. Therefore the Euler-Lagrange equations will be slightly more complicated to compute than in this classical case.

4.2 Local Criteria

An interesting generalization of the ideas developed in the previous section is to make the estimator (4.2) local. This allows us to take into account non-stationarities in the distributions of the intensities. We weight our estimate (4.2) with a spatial Gaussian of variance $\gamma > 0$ centered at \mathbf{x}_0 . This means that for each point \mathbf{x}_0 in Ω we have two random variables, noted $X_{I_1^\sigma, \mathbf{x}_0}^l$ and $X_{I_2^\sigma, \mathbf{x}_0, \mathbf{h}}^l$ (the upper index l stands for local) whose joint pdf is defined by:

$$P_{\mathbf{h}}(\mathbf{i}, \mathbf{x}_0) = \frac{1}{\mathcal{G}_{\gamma}(\mathbf{x}_0)} \int_{\Omega} G_{\beta}(\mathbf{I}_{\mathbf{h}}(\mathbf{x}) - \mathbf{i}) G_{\gamma}(\mathbf{x} - \mathbf{x}_0) d\mathbf{x}, \quad (4.12)$$

where

$$G_{\gamma}(\mathbf{x} - \mathbf{x}_0) = \frac{1}{(\sqrt{2\pi\gamma})^n} \exp\left(-\frac{|\mathbf{x} - \mathbf{x}_0|^2}{2\gamma}\right),$$

and

$$\mathcal{G}_{\gamma}(\mathbf{x}_0) = \int_{\Omega} G_{\gamma}(\mathbf{x} - \mathbf{x}_0) d\mathbf{x} \leq |\Omega| G_{\gamma}(0). \quad (4.13)$$

¹Note that instead of using the original definition of \mathbf{CR} , we use the total variance theorem to obtain $\mathbf{CR} = 1 - \frac{\mathbf{E}[\mathbf{Var}[X|Y]]}{\mathbf{Var}[Y]}$. This transformation was suggested in [77], and turns out to be more convenient.

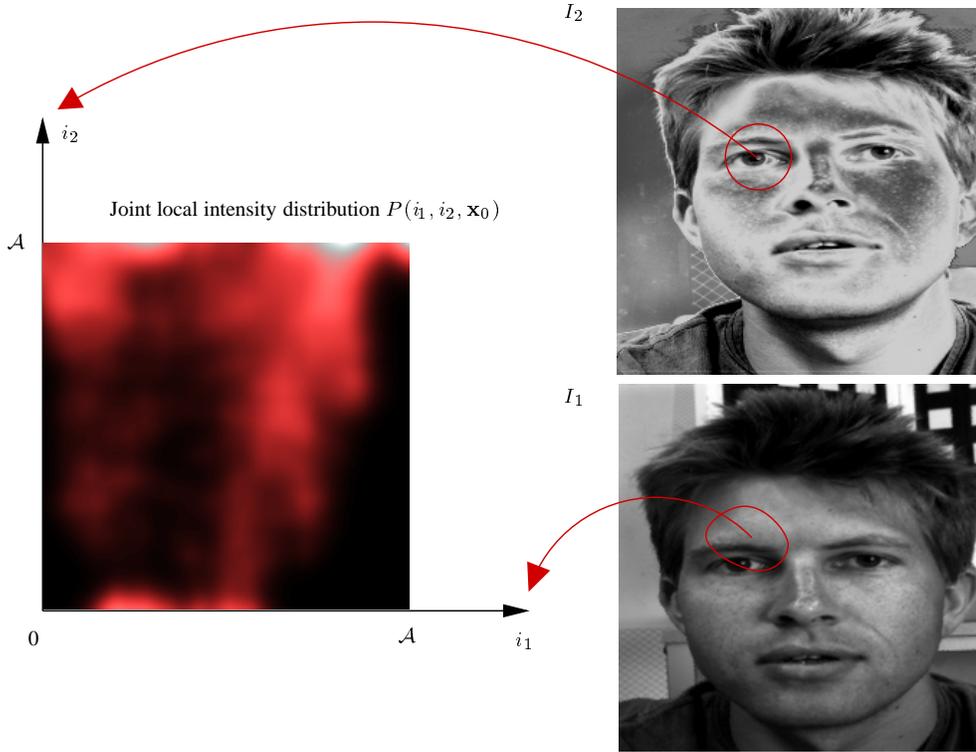


Figure 4.1: Local joint intensity distribution.

The pdf defined by expression (4.12) is in the line of the ideas discussed by Koenderink and Van Doorn in [47], except that we now have a bidimensional histogram calculated around each point (see figure 4.1). With the help of this estimate, we define at every point \mathbf{x}_0 of Ω the local cross correlation between the two images I_1^l and $I_2^l \circ (\text{Id} + \mathbf{h})$, noted $\mathbf{CC}^l(\mathbf{h}, \mathbf{x}_0)$, the local correlation ratio, noted $\mathbf{CR}^l(\mathbf{h}, \mathbf{x}_0)$ and the local mutual information, noted $\mathbf{MI}^l(\mathbf{h}, \mathbf{x}_0)$. In order to do this, just as in the global case, we need to introduce more random variables besides $X_{I_1^l, \mathbf{x}_0}^l$ and $X_{I_2^l, \mathbf{x}_0, \mathbf{h}}^l$. We summarize our notations and definitions in Table 4.2.

As in the global case, we define the mean and variance of $X_{I_1^l, \mathbf{x}_0}^l$ (note that they are not random variables but they are functions of \mathbf{x}_0):

$$\mu_1(\mathbf{x}_0) \equiv \int_{\mathbb{R}} i_1 p_{\mathbf{h}}(i_1, \mathbf{x}_0) di_1, \quad (4.14)$$

$$v_1(\mathbf{x}_0) \equiv \int_{\mathbb{R}} i_1^2 p_{\mathbf{h}}(i_1, \mathbf{x}_0) di_1 - (\mu_1(\mathbf{x}_0))^2, \quad (4.15)$$

the mean and variance of $X_{I_2^l, \mathbf{x}_0, \mathbf{h}}^l$ (note that these quantities depend additionally on the displacement field \mathbf{h}):

$$\mu_2(\mathbf{h}, \mathbf{x}_0) \equiv \int_{\mathbb{R}} i_2 p_{\mathbf{h}}(i_2, \mathbf{x}_0) di_2, \quad (4.16)$$

Random variable	Value	PDF
$(X_{I_1^\sigma}^l, \mathbf{x}_0, X_{I_2^\sigma}^l, \mathbf{x}_0, \mathbf{h})$	(i)	$P_{\mathbf{h}}(i_1, i_2, \mathbf{x}_0)$
$X_{I_1^\sigma}^l, \mathbf{x}_0$	i_1	$p(i_1, \mathbf{x}_0) = \int_{\mathbb{R}} P_{\mathbf{h}}(i_1, i_2, \mathbf{x}_0) di_2$
$X_{I_2^\sigma}^l, \mathbf{x}_0, \mathbf{h}$	i_2	$p_{\mathbf{h}}(i_2, \mathbf{x}_0) = \int_{\mathbb{R}} P_{\mathbf{h}}(i_1, i_2, \mathbf{x}_0) di_1$
$\mathbf{E}[X_{I_2^\sigma}^l, \mathbf{x}_0, \mathbf{h} X_{I_1^\sigma}^l, \mathbf{x}_0]$	$\mu_{2 1}(i_1, \mathbf{h}, \mathbf{x}_0) \equiv \int_{\mathbb{R}} i_2 P_{\mathbf{h}}(i_2, \mathbf{x}_0 i_1) di_2$	$p(i_1, \mathbf{x}_0)$
$\mathbf{Var}[X_{I_2^\sigma}^l, \mathbf{x}_0, \mathbf{h} X_{I_1^\sigma}^l, \mathbf{x}_0]$	$v_{2 1}(i_1, \mathbf{h}, \mathbf{x}_0) \equiv \int_{\mathbb{R}} i_2^2 P_{\mathbf{h}}(i_2, \mathbf{x}_0 i_1) di_2 - \mu_{2 1}(i_1, \mathbf{h}, \mathbf{x}_0)^2$	$p(i_1, \mathbf{x}_0)$

Table 4.2: Random variables: local case.

$$v_2(\mathbf{h}, \mathbf{x}_0) \equiv \int_{\mathbb{R}} i_2^2 p_{\mathbf{h}}(i_2, \mathbf{x}_0) di_2 - (\mu_2(\mathbf{h}, \mathbf{x}_0))^2, \quad (4.17)$$

as well as their covariance:

$$v_{1,2}(\mathbf{h}, \mathbf{x}_0) \equiv \int_{\mathbb{R}} i_1 i_2 P_{\mathbf{h}}(i_2, \mathbf{x}_0) di - \mu_1(\mathbf{x}_0) \mu_2(\mathbf{h}, \mathbf{x}_0). \quad (4.18)$$

The semi-local similarity measures (i.e. depending on \mathbf{x}_0) can be written in terms of these quantities:

$$\mathbf{CC}^l(\mathbf{h}, \mathbf{x}_0) = \frac{v_{1,2}(\mathbf{h}, \mathbf{x}_0)^2}{v_1(\mathbf{x}_0) v_2(\mathbf{h}, \mathbf{x}_0)}, \quad (4.19)$$

$$\mathbf{CR}^l(\mathbf{h}, \mathbf{x}_0) = 1 - \frac{1}{v_2(\mathbf{h}, \mathbf{x}_0)} \int_{\mathbb{R}} v_{2|1}(i_1, \mathbf{h}, \mathbf{x}_0) p(i_1, \mathbf{x}_0) di_1, \quad (4.20)$$

$$\mathbf{MI}^l(\mathbf{h}, \mathbf{x}_0) = \int_{\mathbb{R}^2} P_{\mathbf{h}}(\mathbf{i}, \mathbf{x}_0) \log \frac{P_{\mathbf{h}}(\mathbf{i}, \mathbf{x}_0)}{p(i_1, \mathbf{x}_0) p_{\mathbf{h}}(i_2, \mathbf{x}_0)} di. \quad (4.21)$$

We define global similarity functionals by aggregating these local measures:

$$\mathbf{CC}^l(\mathbf{h}) = \int_{\Omega} \mathbf{CC}^l(\mathbf{h}, \mathbf{x}_0) d\mathbf{x}_0,$$

$$\mathbf{CR}^l(\mathbf{h}) = \int_{\Omega} \mathbf{CR}^l(\mathbf{h}, \mathbf{x}_0) d\mathbf{x}_0,$$

$$\mathbf{MI}^l(\mathbf{h}) = \int_{\Omega} \mathbf{MI}^l(\mathbf{h}, \mathbf{x}_0) d\mathbf{x}_0.$$

The three criteria are positive and should be maximized with respect to the field \mathbf{h} . In order to define a minimization problem, we propose the following definition.

Definition 4.2 *The three local dissimilarity measures based on the cross correlation, the correlation ratio and the mutual information are as follows:*

$$\mathcal{J}_{\mathbf{CC}^l}(\mathbf{h}) = -\mathbf{CC}^l(\mathbf{h}),$$

$$\mathcal{J}_{\mathbf{CR}^l}(\mathbf{h}) = -\mathbf{CR}^l(\mathbf{h}) + |\Omega|,$$

$$\mathcal{J}_{\mathbf{MI}^l}(\mathbf{h}) = -\mathbf{MI}^l(\mathbf{h}).$$

Note that, just as in the global case, this definition shows that the mappings $\mathbf{h} \rightarrow \mathcal{J}_{\mathbf{CC}^l}(\mathbf{h})$, $\mathbf{h} \rightarrow \mathcal{J}_{\mathbf{CR}^l}(\mathbf{h})$ and $\mathbf{h} \rightarrow \mathcal{J}_{\mathbf{MI}^l}(\mathbf{h})$ are not of the form $\mathbf{h} \rightarrow \int_{\Omega} L(\mathbf{h}(\mathbf{x})) d\mathbf{x}$, for some differentiable function $L : \mathbb{R}^n \rightarrow \mathbb{R}$. Therefore the Euler-Lagrange equations will be more complicated to compute than in this classical case. This will be the object of the next chapter.

4.3 Continuity of \mathbf{MI}^g and \mathbf{MI}^l

Recall that the existence of minimizers for $\mathcal{I}(\mathbf{h})$ was discussed in the end of Chapter 3 by assuming continuity and boundedness of $\mathcal{J}(\mathbf{h})$. This is proved in Theorems 6.29 on page 104 and 6.61 on page 117 for the cross correlation in the global and local cases, respectively, and in Theorems 6.21 on page 100 and 6.53 on page 115 for the correlation ratio in the global and local cases, respectively. In the case of the mutual information, we have the following.

Proposition 4.1 *Let $\mathbf{h}_n, n = 1, \dots, \infty$ be a sequence of functions of H such that $\mathbf{h}_n \rightarrow \mathbf{h}$ almost everywhere in Ω . Then $\mathbf{MI}^g(\mathbf{h}_n) \rightarrow \mathbf{MI}^g(\mathbf{h})$.*

Proof : Because I_2^σ and g_β are continuous, $G_\beta(\mathbf{i}_{\mathbf{h}_n}(\mathbf{x}) - \mathbf{i}) \rightarrow G_\beta(\mathbf{i}_{\mathbf{h}}(\mathbf{x}) - \mathbf{i})$ a.e. in $\Omega \times \mathbb{R}^2$. Since $G_\beta(\mathbf{i}_{\mathbf{h}_n}(\mathbf{x}) - \mathbf{i}) \leq g_\beta(0)^2$, the dominated convergence theorem implies

that $P_{\mathbf{h}_n}(\mathbf{i}) \rightarrow P_{\mathbf{h}}(\mathbf{i})$ for all $\mathbf{i} \in \mathbb{R}^2$. A similar reasoning shows that $p_{\mathbf{h}_n}(i_2) \rightarrow p_{\mathbf{h}}(i_2)$ for all $i_2 \in \mathbb{R}$. Hence, the logarithm being continuous,

$$P_{\mathbf{h}_n}(\mathbf{i}) \log \frac{P_{\mathbf{h}_n}(\mathbf{i})}{p(i_1)p_{\mathbf{h}_n}(i_2)} \rightarrow P_{\mathbf{h}}(\mathbf{i}) \log \frac{P_{\mathbf{h}}(\mathbf{i})}{p(i_1)p_{\mathbf{h}}(i_2)} \quad \forall \mathbf{i} \in \mathbb{R}^2.$$

We next consider three cases to find an upper bound for $P_{\mathbf{h}_n}(\mathbf{i}) \left| \log \frac{P_{\mathbf{h}_n}(\mathbf{i})}{p(i_1)p_{\mathbf{h}_n}(i_2)} \right|$:

$i_2 \leq 0$

This is the case where

$$0 \leq |i_2| \leq |i_2 - I_2^\sigma(\mathbf{x} + \mathbf{h}_n(\mathbf{x}))| \leq |i_2 - \mathcal{A}| \quad n \geq 1.$$

Hence

$$g_\beta(i_2 - \mathcal{A}) \leq g_\beta(i_2 - I_2^\sigma(\mathbf{x} + \mathbf{h}_n(\mathbf{x}))) \leq g_\beta(i_2) \quad n \geq 1.$$

This yields

$$\frac{g_\beta(i_2 - \mathcal{A})}{g_\beta(i_2)} \leq \frac{P_{\mathbf{h}_n}(\mathbf{i})}{p(i_1)p_{\mathbf{h}_n}(i_2)} \leq \frac{g_\beta(i_2)}{g_\beta(i_2 - \mathcal{A})}$$

and

$$\left| \log \frac{P_{\mathbf{h}_n}(\mathbf{i})}{p(i_1)p_{\mathbf{h}_n}(i_2)} \right| \leq \log \frac{g_\beta(i_2)}{g_\beta(i_2 - \mathcal{A})},$$

and therefore

$$P_{\mathbf{h}_n}(\mathbf{i}) \left| \log \frac{P_{\mathbf{h}_n}(\mathbf{i})}{p(i_1)p_{\mathbf{h}_n}(i_2)} \right| \leq g_\beta(i_2)p(i_1) \log \frac{g_\beta(i_2)}{g_\beta(i_2 - \mathcal{A})}.$$

The function on the right-hand side is continuous and integrable in $\mathbb{R} \times]-\infty, \mathcal{A}]$.

$0 \leq i_2 \leq \mathcal{A}$

We have

$$0 \leq |i_2 - I_2^\sigma(\mathbf{x} + \mathbf{h}_n(\mathbf{x}))| \leq \mathcal{A} \quad n \geq 1.$$

Hence

$$g_\beta(\mathcal{A}) \leq g_\beta(i_2 - I_2^\sigma(\mathbf{x} + \mathbf{h}_n(\mathbf{x}))) \leq g_\beta(0) \quad n \geq 1.$$

This yields

$$\frac{g_\beta(\mathcal{A})}{g_\beta(0)} \leq \frac{P_{\mathbf{h}_n}(\mathbf{i})}{p(i_1)p_{\mathbf{h}_n}(i_2)} \leq \frac{g_\beta(0)}{g_\beta(\mathcal{A})},$$

and

$$\left| \log \frac{P_{\mathbf{h}_n}(\mathbf{i})}{p(i_1)p_{\mathbf{h}_n}(i_2)} \right| \leq \log \frac{g_\beta(0)}{g_\beta(\mathcal{A})},$$

and therefore

$$P_{\mathbf{h}_n}(\mathbf{i}) \left| \log \frac{P_{\mathbf{h}_n}(\mathbf{i})}{p(i_1)p_{\mathbf{h}_n}(i_2)} \right| \leq g_\beta(0)p(i_1) \log \frac{g_\beta(0)}{g_\beta(\mathcal{A})}.$$

The function on the right-hand side is continuous and integrable in $\mathbb{R} \times [0, \mathcal{A}]$.

$i_2 \geq \mathcal{A}$

This is the case where

$$0 \leq i_2 - \mathcal{A} \leq i_2 - I_2^\sigma(\mathbf{x} + \mathbf{h}_n(\mathbf{x})) \leq i_2 \quad n \geq 1.$$

Hence

$$g_\beta(i_2) \leq g_\beta(i_2 - I_2^\sigma(\mathbf{x} + \mathbf{h}_n(\mathbf{x}))) \leq g_\beta(i_2 - \mathcal{A}) \quad n \geq 1.$$

This yields

$$\frac{g_\beta(i_2)}{g_\beta(i_2 - \mathcal{A})} \leq \frac{P_{\mathbf{h}_n}(\mathbf{i})}{p(i_1)p_{\mathbf{h}_n}(i_2)} \leq \frac{g_\beta(i_2 - \mathcal{A})}{g_\beta(i_2)},$$

and

$$\left| \log \frac{P_{\mathbf{h}_n}(\mathbf{i})}{p(i_1)p_{\mathbf{h}_n}(i_2)} \right| \leq \log \frac{g_\beta(i_2 - \mathcal{A})}{g_\beta(i_2)},$$

and therefore

$$P_{\mathbf{h}_n}(\mathbf{i}) \left| \log \frac{P_{\mathbf{h}_n}(\mathbf{i})}{p(i_1)p_{\mathbf{h}_n}(i_2)} \right| \leq g_\beta(i_2 - \mathcal{A})p(i_1) \log \frac{g_\beta(i_2 - \mathcal{A})}{g_\beta(i_2)}.$$

The function on the right-hand side is continuous and integrable in $\mathbb{R} \times]\mathcal{A}, +\infty]$.

The dominated convergence theorem implies that

$$\begin{aligned} \mathbf{MI}^g(\mathbf{h}_n) &= \int_{\mathbb{R}^2} P_{\mathbf{h}_n}(\mathbf{i}) \log \frac{P_{\mathbf{h}_n}(\mathbf{i})}{p(i_1)p_{\mathbf{h}_n}(i_2)} d\mathbf{i} \rightarrow \\ &\mathbf{MI}^g(\mathbf{h}) = \int_{\mathbb{R}^2} P_{\mathbf{h}}(\mathbf{i}) \log \frac{P_{\mathbf{h}}(\mathbf{i})}{p(i_1)p_{\mathbf{h}}(i_2)} d\mathbf{i}. \end{aligned}$$

□

Concerning the local case, a similar result holds true.

Proposition 4.2 *Let $\mathbf{h}_n, n = 1, \dots, \infty$ be a sequence of functions of H such that $\mathbf{h}_n \rightarrow \mathbf{h}$ almost everywhere in Ω then $\mathbf{MI}^l(\mathbf{h}_n) \rightarrow \mathbf{MI}^l(\mathbf{h})$.*

Proof : The proof is similar to that of proposition 4.1 □

Chapter 5

The Euler-Lagrange Equations

In this chapter, the computation of the first variation of the statistical dissimilarity measures defined in the previous chapter is carried out by considering the variations of the joint density estimates. This provides us with a simple way to quantify the contribution of local infinitesimal variations of \mathbf{h} to these intrinsically non-local dissimilarity criteria.

5.1 Global Criteria

We start with the global criteria, studying them in decreasing order of generality, i.e. starting with mutual information.

5.1.1 Mutual Information

We do an explicit computation of the first variation of $\mathcal{J}_{\text{MI}^\vartheta}(\mathbf{h})$ (definition 4.1 on page 68). For this purpose, we recall that the first variation is defined as $\left. \frac{\partial}{\partial \epsilon} \mathcal{J}_{\text{MI}^\vartheta}(\mathbf{h} + \epsilon \mathbf{k}) \right|_{\epsilon=0}$. From the definition of $\mathcal{J}_{\text{MI}^\vartheta}(\mathbf{h})$, we readily have

$$\begin{aligned} \frac{\partial \mathcal{J}_{\text{MI}^\vartheta}(\mathbf{h} + \epsilon \mathbf{k})}{\partial \epsilon} &= - \int_{\mathbb{R}^2} \frac{\partial}{\partial \epsilon} \left[P_{\mathbf{h} + \epsilon \mathbf{k}}(\mathbf{i}) \log \frac{P_{\mathbf{h} + \epsilon \mathbf{k}}(\mathbf{i})}{p(i_1) p_{\mathbf{h} + \epsilon \mathbf{k}}(i_2)} \right] d\mathbf{i} \\ &= - \int_{\mathbb{R}^2} \left(1 + \log \frac{P_{\mathbf{h} + \epsilon \mathbf{k}}(\mathbf{i})}{p(i_1) p_{\mathbf{h} + \epsilon \mathbf{k}}(i_2)} \right) \frac{\partial P_{\mathbf{h} + \epsilon \mathbf{k}}(\mathbf{i})}{\partial \epsilon} di_1 di_2 \\ &\quad - \int_{\mathbb{R}^2} \frac{P_{\mathbf{h} + \epsilon \mathbf{k}}(\mathbf{i})}{p_{\mathbf{h} + \epsilon \mathbf{k}}(i_2)} \frac{\partial p_{\mathbf{h} + \epsilon \mathbf{k}}(i_2)}{\partial \epsilon} d\mathbf{i}. \end{aligned}$$

We notice that the second term on the right-hand side is zero, since we have

$$\int_{\mathbb{R}} \frac{\partial p_{\mathbf{h}+\epsilon\mathbf{k}}(i_2)}{\partial \epsilon} \frac{1}{p_{\mathbf{h}+\epsilon\mathbf{k}}(i_2)} \underbrace{\int_{\mathbb{R}} P_{\mathbf{h}+\epsilon\mathbf{k}}(\mathbf{i}) \, di_1}_{p_{\mathbf{h}+\epsilon\mathbf{k}}(i_2)} \, di_2 = \frac{\partial}{\partial \epsilon} \underbrace{\int_{\mathbb{R}} p_{\mathbf{h}+\epsilon\mathbf{k}}(i_2) \, di_2}_1 = 0.$$

Thus, we may write the first variation of $\mathcal{J}_{\mathbf{M}\mathbf{I}^g}(\mathbf{h})$ as

$$\left. \frac{\partial \mathcal{J}_{\mathbf{M}\mathbf{I}^g}(\mathbf{h} + \epsilon\mathbf{k})}{\partial \epsilon} \right|_{\epsilon=0} = \int_{\mathbb{R}^2} E_{\mathbf{h}}^{\mathbf{M}\mathbf{I}}(\mathbf{i}) \left. \frac{\partial P_{\mathbf{h}+\epsilon\mathbf{k}}(\mathbf{i})}{\partial \epsilon} \right|_{\epsilon=0} \, d\mathbf{i}, \quad (5.1)$$

where

$$E_{\mathbf{h}}^{\mathbf{M}\mathbf{I}}(\mathbf{i}) = - \left(1 + \log \frac{P_{\mathbf{h}}(\mathbf{i})}{p(i_1) p_{\mathbf{h}}(i_2)} \right).$$

The function $P_{\mathbf{h}+\epsilon\mathbf{k}}(\mathbf{i})$ is given by equation (4.2):

$$P_{\mathbf{h}+\epsilon\mathbf{k}}(\mathbf{i}) = \frac{1}{|\Omega|} \int_{\Omega} G_{\beta}(\mathbf{I}_{\mathbf{h}+\epsilon\mathbf{k}}(\mathbf{x}) - \mathbf{i}) \, d\mathbf{x}, \quad (5.2)$$

and therefore

$$\frac{\partial P_{\mathbf{h}+\epsilon\mathbf{k}}(\mathbf{i})}{\partial \epsilon} = \frac{1}{|\Omega|} \int_{\Omega} \partial_2 G_{\beta}(\mathbf{I}_{\mathbf{h}+\epsilon\mathbf{k}}(\mathbf{x}) - \mathbf{i}) \nabla I_2^{\sigma}(\mathbf{x} + \mathbf{h}(\mathbf{x}) + \epsilon\mathbf{k}(\mathbf{x})) \cdot \mathbf{k}(\mathbf{x}) \, d\mathbf{x}.$$

Thus, we finally have

$$\begin{aligned} \left. \frac{\partial \mathcal{J}_{\mathbf{M}\mathbf{I}^g}(\mathbf{h} + \epsilon\mathbf{k})}{\partial \epsilon} \right|_{\epsilon=0} &= \\ & \frac{1}{|\Omega|} \int_{\mathbb{R}^2} \int_{\Omega} E_{\mathbf{h}}^{\mathbf{M}\mathbf{I}}(\mathbf{i}) \, \partial_2 G_{\beta}(\mathbf{I}(\mathbf{x})_{\mathbf{h}} - \mathbf{i}) \nabla I_2^{\sigma}(\mathbf{x} + \mathbf{h}(\mathbf{x})) \cdot \mathbf{k}(\mathbf{x}) \, d\mathbf{x} \, d\mathbf{i}. \end{aligned}$$

A convolution appears with respect to the intensity variable \mathbf{i} . This convolution commutes with the derivative ∂_2 with respect to the second intensity variable i_2 , and therefore

$$\left. \frac{\partial \mathcal{J}_{\mathbf{M}\mathbf{I}^g}(\mathbf{h} + \epsilon\mathbf{k})}{\partial \epsilon} \right|_{\epsilon=0} = \frac{1}{|\Omega|} \int_{\Omega} \left(G_{\beta} \star \partial_2 E_{\mathbf{h}}^{\mathbf{M}\mathbf{I}} \right) (\mathbf{I}_{\mathbf{h}}(\mathbf{x})) \nabla I_2^{\sigma}(\mathbf{x} + \mathbf{h}(\mathbf{x})) \cdot \mathbf{k}(\mathbf{x}) \, d\mathbf{x}.$$

By identifying this expression with a scalar product in $H = \mathbf{L}^2(\Omega)$, we define the gradient of $\mathcal{J}_{\mathbf{M}\mathbf{I}^g}(\mathbf{h})$, denoted $\nabla_H \mathcal{J}_{\mathbf{M}\mathbf{I}^g}(\mathbf{h})$, with the property that:

$$\left. \frac{\partial \mathcal{J}_{\mathbf{M}\mathbf{I}^g}(\mathbf{h} + \epsilon\mathbf{k})}{\partial \epsilon} \right|_{\epsilon=0} = (\nabla_H \mathcal{J}_{\mathbf{M}\mathbf{I}^g}(\mathbf{h}), \mathbf{k})_{\mathbf{L}^2(\Omega)}.$$

Thus,

$$\nabla_H \mathcal{J}_{\mathbf{M}\mathbf{I}^g}(\mathbf{h})(\mathbf{x}) = \frac{1}{|\Omega|} \left(G_{\beta} \star \partial_2 E_{\mathbf{h}}^{\mathbf{M}\mathbf{I}} \right) (\mathbf{I}_{\mathbf{h}}(\mathbf{x})) \nabla I_2^{\sigma}(\mathbf{x} + \mathbf{h}(\mathbf{x})),$$

where

$$\partial_2 E_{\mathbf{h}}^{\mathbf{M}\mathbf{I}}(\mathbf{i}) = - \left(\frac{\partial_2 P_{\mathbf{h}}(\mathbf{i})}{P_{\mathbf{h}}(\mathbf{i})} - \frac{p'_{\mathbf{h}}(i_2)}{p_{\mathbf{h}}(i_2)} \right).$$

We define the function $\mathbb{R}^2 \rightarrow \mathbb{R}$:

$$L_{\text{MI},\mathbf{h}}^g(\mathbf{i}) \equiv \frac{1}{|\Omega|} \partial_2 E_{\mathbf{h}}^{\text{MI}}(\mathbf{i}).$$

The gradient of $\mathcal{J}_{\text{MI}^g}(\mathbf{h})$ is a smoothed version of this function, evaluated at the intensity pair $\mathbf{I}_{\mathbf{h}}(\mathbf{x})$, times the vector pointing to the direction of local maximum increase of i_2 , namely $\nabla I_2^g(\mathbf{x} + \mathbf{h}(\mathbf{x}))$. It is therefore of interest to interpret the behavior of this function. Given a point \mathbf{x} , the pair $\mathbf{I}_{\mathbf{h}}(\mathbf{x})$ lies somewhere in the square $[0, \mathcal{A}]$ within the domain of intensities, i.e. \mathbb{R}^2 (see Figure 5.1). The first term in $L_{\text{MI},\mathbf{h}}^g$, namely $\frac{\partial_2 P_{\mathbf{h}}(\mathbf{i})}{P_{\mathbf{h}}(\mathbf{i})}$, tries to make the intensity i_2 move closer to a local maximum of $P_{\mathbf{h}}$. It thus tends to cluster $P_{\mathbf{h}}$. On the contrary, the second term, namely $-\frac{p'_{\mathbf{h}}(i_2)}{p_{\mathbf{h}}(i_2)}$, tries to prevent the marginal law $p_{\mathbf{h}}(i_2)$ from becoming too clustered, i.e. keep $X_{I_2^g, \mathbf{h}}$ as unpredictable as possible. The fact that only the value of i_2 is changed implies that these movements take place only along one of the axes. This lack of symmetry is a general problem coming from the way in which the problem is posed. We refer to the works of Trouvé and Younes [83], Cachier and Rey [22], Christensen and He [24] and Alvarez et al. [1] for some recent approaches to overcome this lack of symmetry. The red sketch in Figure 5.1 depicts a possible state of the function $P_{\mathbf{h}}$ after minimization of $\mathcal{J}_{\text{MI}^g}(\mathbf{h})$.

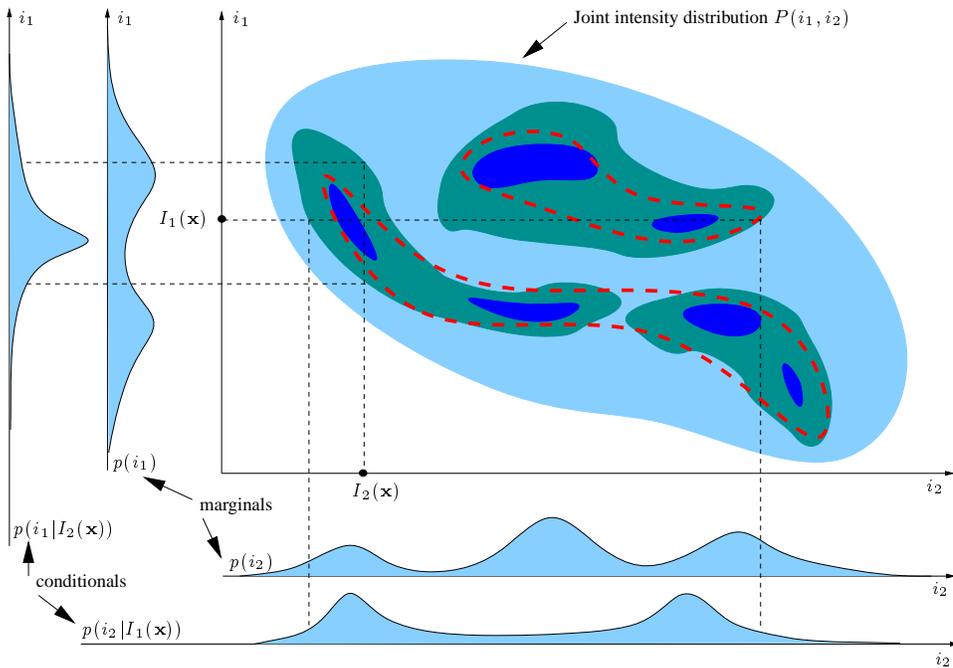


Figure 5.1: Sketch of a possible state of the joint pdf of intensities after minimization of $\mathcal{J}_{\text{MI}^g}(\mathbf{h})$ (see text).

5.1.2 Correlation Ratio

In this section, we compute the first variation of $\mathcal{J}_{\mathbf{CR}^g}(\mathbf{h})$ (definition 4.1 on page 68). To do this, we note

$$w(\mathbf{h}) \equiv \mathbf{E}[\mathbf{Var}[X_{I_2^g, \mathbf{h}} | X_{I_1^g}]] = \int_{\mathbb{R}} v_{2|1}(i_1, \mathbf{h}) p(i_1) di_1,$$

so that

$$\mathcal{J}_{\mathbf{CR}^g}(\mathbf{h}) = \frac{w(\mathbf{h})}{v_2(\mathbf{h})} = \frac{1}{v_2(\mathbf{h})} \left(\int_{\mathbb{R}^2} i_2^2 P_{\mathbf{h}}(\mathbf{i}) d\mathbf{i} - \int_{\mathbb{R}} \mu_{2|1}(i_1, \mathbf{h})^2 p(i_1) di_1 \right),$$

where

$$\mu_{2|1}(i_1, \mathbf{h}) = \int_{\mathbb{R}} i_2 \frac{P_{\mathbf{h}}(\mathbf{i})}{p(i_1)} di_2.$$

Thus, we readily have

$$\frac{\partial \mathcal{J}_{\mathbf{CR}^g}(\mathbf{h} + \epsilon \mathbf{k})}{\partial \epsilon} = \frac{1}{v_2(\mathbf{h} + \epsilon \mathbf{k})} \left(\frac{\partial w(\mathbf{h} + \epsilon \mathbf{k})}{\partial \epsilon} - \mathcal{J}_{\mathbf{CR}^g}(\mathbf{h} + \epsilon \mathbf{k}) \frac{\partial v_2(\mathbf{h} + \epsilon \mathbf{k})}{\partial \epsilon} \right),$$

where

$$\frac{\partial w(\mathbf{h} + \epsilon \mathbf{k})}{\partial \epsilon} = \int_{\mathbb{R}^2} i_2^2 \frac{\partial P_{\mathbf{h} + \epsilon \mathbf{k}}(\mathbf{i})}{\partial \epsilon} d\mathbf{i} - \int_{\mathbb{R}} 2\mu_{2|1}(i_1, \mathbf{h} + \epsilon \mathbf{k}) \int_{\mathbb{R}} i_2 \frac{\partial P_{\mathbf{h} + \epsilon \mathbf{k}}(\mathbf{i})}{\partial \epsilon} d\mathbf{i},$$

and

$$\frac{\partial v_2(\mathbf{h} + \epsilon \mathbf{k})}{\partial \epsilon} = \int_{\mathbb{R}^2} i_2^2 \frac{\partial P_{\mathbf{h} + \epsilon \mathbf{k}}(\mathbf{i})}{\partial \epsilon} d\mathbf{i} - 2\mu_2(\mathbf{h} + \epsilon \mathbf{k}) \int_{\mathbb{R}^2} i_2 \frac{\partial P_{\mathbf{h} + \epsilon \mathbf{k}}(\mathbf{i})}{\partial \epsilon} d\mathbf{i}. \quad (5.3)$$

Similarly to the case of the mutual information (see equation (5.1)), the first variation of $\mathcal{J}_{\mathbf{CR}^g}(\mathbf{h})$ can be put in the form

$$\left. \frac{\partial \mathcal{J}_{\mathbf{CR}^g}(\mathbf{h} + \epsilon \mathbf{k})}{\partial \epsilon} \right|_{\epsilon=0} = \int_{\mathbb{R}^2} E_{\mathbf{h}}^{\mathbf{CR}}(\mathbf{i}) \left. \frac{\partial P_{\mathbf{h} + \epsilon \mathbf{k}}(\mathbf{i})}{\partial \epsilon} \right|_{\epsilon=0} d\mathbf{i},$$

where

$$E_{\mathbf{h}}^{\mathbf{CR}}(\mathbf{i}) = \frac{i_2}{v_2(\mathbf{h})} \left(i_2 - 2\mu_{2|1}(i_1, \mathbf{h}) - \mathcal{J}_{\mathbf{CR}^g}(\mathbf{h})(i_2 - 2\mu_2(\mathbf{h})) \right).$$

The discussion starting before equation (5.2) remains identical in this case. Thus, the gradient of $\mathcal{J}_{\mathbf{CR}^g}(\mathbf{h})$ is given by:

$$\nabla_H \mathcal{J}_{\mathbf{CR}^g}(\mathbf{h})(\mathbf{x}) = \frac{1}{|\Omega|} \left(G_{\beta} \star \partial_2 E_{\mathbf{h}}^{\mathbf{CR}} \right) (\mathbf{I}_{\mathbf{h}}(\mathbf{x})) \nabla I_2^g(\mathbf{x} + \mathbf{h}(\mathbf{x})),$$

where

$$\partial_2 E_{\mathbf{h}}^{\mathbf{CR}}(\mathbf{i}) = \frac{2}{v_2(\mathbf{h})} \left(\mu_2(\mathbf{h}) - \mu_{2|1}(i_1, \mathbf{h}) + \mathbf{CR}^g(\mathbf{h})(i_2 - \mu_2(\mathbf{h})) \right).$$

As for the mutual information, we define the function $\mathbb{R}^2 \rightarrow \mathbb{R}$:

$$L_{\text{CR},\mathbf{h}}^g(\mathbf{i}) \equiv \frac{1}{|\Omega|} \partial_2 E_{\mathbf{h}}^{\text{CR}}(\mathbf{i}),$$

and interpret its behavior as an intensity comparison function. The function $\mu_{2|1}(i_1, \mathbf{h})$ gives the “backbone” of $P_{\mathbf{h}}$. We see that trying to decrease the value of $\mathcal{J}_{\text{CR}^g}(\mathbf{h})$ amounts to making i_2 lie as close as possible to $\mu_{2|1}(i_1, \mathbf{h})$, while keeping this backbone as “complex” as possible (away from $\mu_2(\mathbf{h})$). The red sketch in Figure 5.2 depicts a possible state of the function $P_{\mathbf{h}}$ after minimization of $\mathcal{J}_{\text{CR}^g}(\mathbf{h})$.

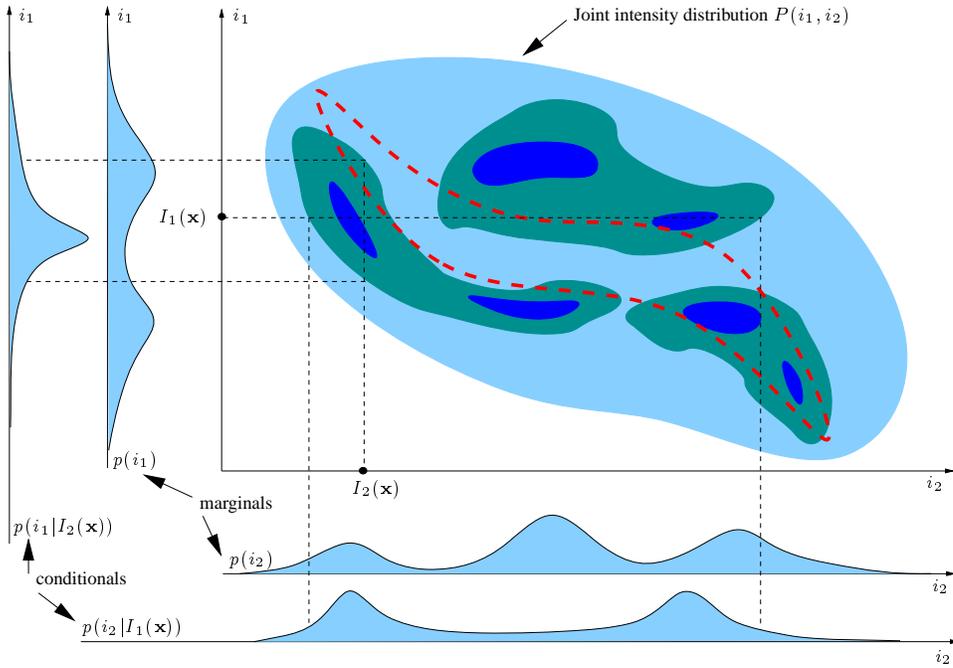


Figure 5.2: Sketch of a possible state of the joint pdf of intensities after minimization of $\mathcal{J}_{\text{CR}^g}(\mathbf{h})$ (see text).

5.1.3 Cross Correlation

In this section, we compute the first variation of $\mathcal{J}_{\text{CC}^g}(\mathbf{h})$ (definition 4.1 on page 68). This case is extremely similar to the previous two. From the definition of $\mathcal{J}_{\text{CC}^g}(\mathbf{h})$, we readily have

$$\begin{aligned} \frac{\partial \mathcal{J}_{\text{CC}^g}(\mathbf{h} + \epsilon \mathbf{k})}{\partial \epsilon} &= \frac{-1}{v_1 v_2(\mathbf{h} + \epsilon \mathbf{k})} \\ &\quad \left(2 v_{1,2}(\mathbf{h} + \epsilon \mathbf{k}) \frac{\partial v_{1,2}(\mathbf{h} + \epsilon \mathbf{k})}{\partial \epsilon} - \text{CC}^g(\mathbf{h} + \epsilon \mathbf{k}) v_1 \frac{\partial v_2(\mathbf{h} + \epsilon \mathbf{k})}{\partial \epsilon} \right), \end{aligned}$$

where

$$\frac{\partial v_{1,2}(\mathbf{h} + \epsilon \mathbf{k})}{\partial \epsilon} = \int_{\mathbb{R}^2} i_1 i_2 \frac{\partial P_{\mathbf{h}+\epsilon \mathbf{k}}(\mathbf{i})}{\partial \epsilon} d\mathbf{i} - \mu_1 \int_{\mathbb{R}^2} i_2 \frac{\partial P_{\mathbf{h}+\epsilon \mathbf{k}}(\mathbf{i})}{\partial \epsilon} d\mathbf{i},$$

and $\frac{\partial}{\partial \epsilon} v_2(\mathbf{h} + \epsilon \mathbf{k})$ is given by equation (5.3). Thus, one more time, we may put the first variation of $\mathcal{J}_{CC^g}(\mathbf{h})$ in the form

$$\left. \frac{\partial \mathcal{J}_{CC^g}(\mathbf{h} + \epsilon \mathbf{k})}{\partial \epsilon} \right|_{\epsilon=0} = \int_{\mathbb{R}^2} E_{\mathbf{h}}^{\text{CC}}(\mathbf{i}) \left. \frac{\partial P_{\mathbf{h}+\epsilon \mathbf{k}}(\mathbf{i})}{\partial \epsilon} \right|_{\epsilon=0} d\mathbf{i},$$

where

$$E_{\mathbf{h}}^{\text{CC}}(\mathbf{i}) = \frac{-1}{v_1 v_2(\mathbf{h})} \left(2 v_{1,2}(\mathbf{h}) i_2 (i_1 - \mu_1) - \mathbf{CC}^g(\mathbf{h}) v_1 i_2 (i_2 - 2 \mu_2(\mathbf{h})) \right).$$

Again, the discussion starting before equation (5.2) remains valid, and therefore the gradient of $\mathcal{J}_{CC^g}(\mathbf{h})$ is given by:

$$\nabla_H \mathcal{J}_{CC^g}(\mathbf{h})(\mathbf{x}) = \frac{1}{|\Omega|} \left(G_\beta \star \partial_2 E_{\mathbf{h}}^{\text{CC}} \right) (\mathbf{I}_{\mathbf{h}}(\mathbf{x})) \nabla I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})),$$

where

$$\partial_2 E_{\mathbf{h}}^{\text{CC}}(\mathbf{i}) = -2 \left[\frac{v_{1,2}(\mathbf{h})}{v_2(\mathbf{h})} \left(\frac{i_1 - \mu_1}{v_1} \right) - \mathbf{CC}^g(\mathbf{h}) \left(\frac{i_2 - \mu_2(\mathbf{h})}{v_2(\mathbf{h})} \right) \right].$$

Notice that in this simple case the convolution may be applied formally, yielding the same expression (since only linear terms are involved), so that in fact

$$\nabla_H \mathcal{J}_{CC^g}(\mathbf{h})(\mathbf{x}) = \frac{1}{|\Omega|} \partial_2 E_{\mathbf{h}}^{\text{CC}}(\mathbf{I}_{\mathbf{h}}(\mathbf{x})) \nabla I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})).$$

As for the previous two criteria, we define the function $\mathbb{R}^2 \rightarrow \mathbb{R}$:

$$L_{\text{CC},\mathbf{h}}^g(\mathbf{i}) \equiv \frac{1}{|\Omega|} \partial_2 E_{\mathbf{h}}^{\text{CC}}(\mathbf{i}),$$

and interpret its behavior as an intensity comparison function. Decreasing the value of $\mathcal{J}_{CC^g}(\mathbf{h})$ amounts to making the pair of intensities lie on a non-vertical straight line in \mathbb{R}^2 (not necessarily passing through the origin). Again, the lack of symmetry in the problem limits the change of intensities to a single direction. The red sketch in Figure 5.3 depicts a possible state of the function $P_{\mathbf{h}}$ after minimization of $\mathcal{J}_{CC^g}(\mathbf{h})$.

5.2 Local Criteria

We now analyse the case of the local criteria. As will be noticed, the reasoning is completely analog to that of the global case, but the functions obtained are significantly more complex.

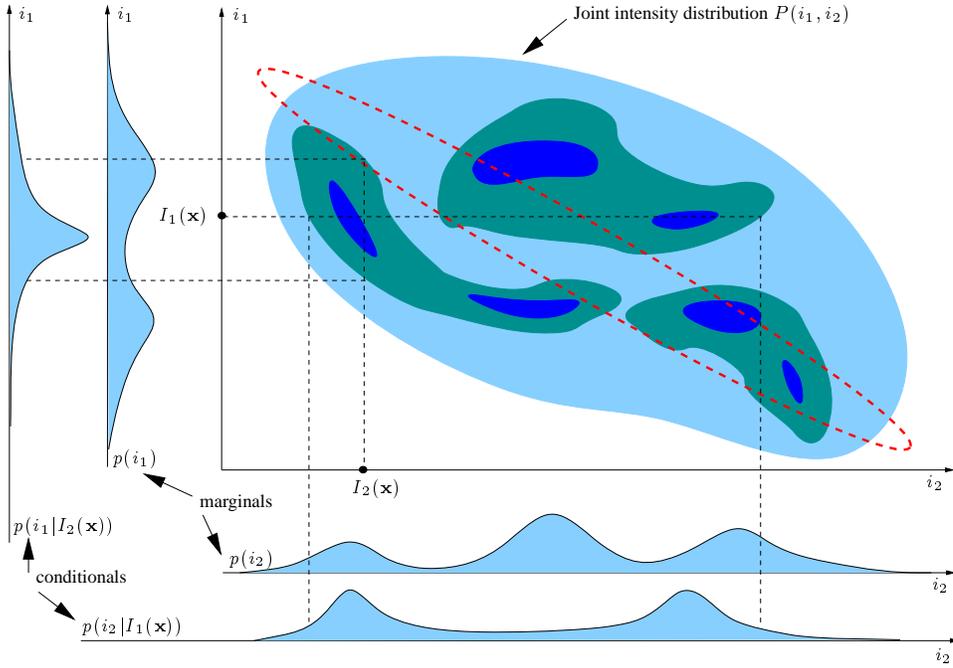


Figure 5.3: Sketch of a possible state of the joint pdf of intensities after minimization of $\mathcal{J}_{CC^g}(\mathbf{h})$ (see text).

5.2.1 Mutual Information

We compute the first variation of $\mathcal{J}_{\text{MI}^t}(\mathbf{h})$ (definition 4.12 on page 68). From this definition, we have

$$\begin{aligned} \frac{\partial \mathcal{J}_{\text{MI}^t}(\mathbf{h} + \epsilon \mathbf{k})}{\partial \epsilon} &= - \int_{\Omega} \int_{\mathbb{R}^2} \frac{\partial}{\partial \epsilon} \left(P_{\mathbf{h} + \epsilon \mathbf{k}}(\mathbf{i}, \mathbf{x}_0) \log \frac{P_{\mathbf{h} + \epsilon \mathbf{k}}(\mathbf{i}, \mathbf{x}_0)}{p(i_1, \mathbf{x}_0) p_{\mathbf{h} + \epsilon \mathbf{k}}(i_2)} \right) d\mathbf{i} d\mathbf{x}_0 \\ &= - \int_{\Omega} \int_{\mathbb{R}^2} \left(1 + \log \frac{P_{\mathbf{h} + \epsilon \mathbf{k}}(\mathbf{i}, \mathbf{x}_0)}{p(i_1, \mathbf{x}_0) p_{\mathbf{h} + \epsilon \mathbf{k}}(i_2, \mathbf{x}_0)} \right) \frac{\partial P_{\mathbf{h} + \epsilon \mathbf{k}}(\mathbf{i}, \mathbf{x}_0)}{\partial \epsilon} d\mathbf{i} d\mathbf{x}_0 \\ &\quad - \underbrace{\int_{\Omega} \int_{\mathbb{R}^2} \frac{P_{\mathbf{h} + \epsilon \mathbf{k}}(\mathbf{i}, \mathbf{x}_0)}{p_{\mathbf{h} + \epsilon \mathbf{k}}(i_2, \mathbf{x}_0)} \frac{\partial p_{\mathbf{h} + \epsilon \mathbf{k}}(i_2, \mathbf{x}_0)}{\partial \epsilon} d\mathbf{i} d\mathbf{x}_0}_{Q}. \end{aligned}$$

We notice that

$$\begin{aligned} Q &= \int_{\mathbb{R}} \frac{\partial p_{\mathbf{h} + \epsilon \mathbf{k}}(i_2, \mathbf{x}_0)}{\partial \epsilon} \frac{1}{p_{\mathbf{h} + \epsilon \mathbf{k}}(i_2, \mathbf{x}_0)} \underbrace{\int_{\mathbb{R}} P_{\mathbf{h} + \epsilon \mathbf{k}}(\mathbf{i}, \mathbf{x}_0) d i_1}_{p_{\mathbf{h} + \epsilon \mathbf{k}}(i_2, \mathbf{x}_0)} d i_2 = \\ &\quad \frac{\partial}{\partial \epsilon} \underbrace{\int_{\mathbb{R}} p_{\mathbf{h} + \epsilon \mathbf{k}}(i_2, \mathbf{x}_0) d i_2}_1 = 0. \end{aligned}$$

Thus, the first variation of $\mathcal{J}_{\mathbf{M}\mathbf{I}^l}(\mathbf{h})$ may be written as

$$\left. \frac{\partial \mathcal{J}_{\mathbf{M}\mathbf{I}^l}(\mathbf{h} + \epsilon \mathbf{k})}{\partial \epsilon} \right|_{\epsilon=0} = \int_{\Omega} \int_{\mathbb{R}^2} E_{\mathbf{h}}^{\mathbf{M}\mathbf{I}}(\mathbf{i}, \mathbf{x}_0) \left. \frac{\partial P_{\mathbf{h}+\epsilon \mathbf{k}}(\mathbf{i}, \mathbf{x}_0)}{\partial \epsilon} \right|_{\epsilon=0} d\mathbf{i} d\mathbf{x}_0,$$

where

$$E_{\mathbf{h}}^{\mathbf{M}\mathbf{I}}(\mathbf{i}, \mathbf{x}_0) = - \left(1 + \log \frac{P_{\mathbf{h}}(\mathbf{i}, \mathbf{x}_0)}{p(i_1, \mathbf{x}_0) p_{\mathbf{h}}(i_2, \mathbf{x}_0)} \right).$$

The law $P_{\mathbf{h}+\epsilon \mathbf{k}}(\mathbf{i}, \mathbf{x}_0)$ is given by equation (4.12):

$$P_{\mathbf{h}+\epsilon \mathbf{k}}(\mathbf{i}, \mathbf{x}_0) = \frac{1}{\mathcal{G}_{\gamma}(\mathbf{x}_0)} \int_{\Omega} G_{\beta}(\mathbf{I}_{\mathbf{h}+\epsilon \mathbf{k}}(\mathbf{x}) - \mathbf{i}) G_{\gamma}(\mathbf{x} - \mathbf{x}_0) d\mathbf{x}. \quad (5.4)$$

Therefore

$$\begin{aligned} \left. \frac{\partial P_{\mathbf{h}+\epsilon \mathbf{k}}(\mathbf{i}, \mathbf{x}_0)}{\partial \epsilon} \right|_{\epsilon=0} &= \\ \frac{1}{\mathcal{G}_{\gamma}(\mathbf{x}_0)} \int_{\Omega} G_{\gamma}(\mathbf{x} - \mathbf{x}_0) \partial_2 G_{\beta}(\mathbf{I}_{\mathbf{h}+\epsilon \mathbf{k}}(\mathbf{x}) - \mathbf{i}) \nabla I_2^{\sigma}(\mathbf{x} + \mathbf{h}(\mathbf{x}) + \epsilon \mathbf{k}(\mathbf{x})) \cdot \mathbf{k}(\mathbf{x}) d\mathbf{x}. \end{aligned}$$

Thus, we finally have

$$\begin{aligned} \left. \frac{\partial \mathcal{J}_{\mathbf{M}\mathbf{I}^l}(\mathbf{h} + \epsilon \mathbf{k})}{\partial \epsilon} \right|_{\epsilon=0} &= \int_{\Omega} \int_{\mathbb{R}^2} \int_{\Omega} \frac{1}{\mathcal{G}_{\gamma}(\mathbf{x}_0)} E_{\mathbf{h}}^{\mathbf{M}\mathbf{I}}(\mathbf{i}, \mathbf{x}_0) G_{\gamma}(\mathbf{x} - \mathbf{x}_0) \partial_2 G_{\beta}(\mathbf{I}_{\mathbf{h}}(\mathbf{x}) - \mathbf{i}) \\ &\quad \nabla I_2^{\sigma}(\mathbf{x} + \mathbf{h}(\mathbf{x})) \cdot \mathbf{k}(\mathbf{x}) d\mathbf{x} d\mathbf{i} d\mathbf{x}_0. \end{aligned}$$

Two convolutions appear, one with respect to the space variable \mathbf{x} and the other one with respect to the intensity variable \mathbf{i} . This last convolution commutes with the partial derivative ∂_2 with respect to the second intensity variable i_2 , and therefore

$$\begin{aligned} \left. \frac{\partial \mathcal{J}_{\mathbf{M}\mathbf{I}^l}(\mathbf{h} + \epsilon \mathbf{k})}{\partial \epsilon} \right|_{\epsilon=0} &= \\ \int_{\Omega} (G_{\gamma} \star (G_{\beta} \star \frac{1}{\mathcal{G}_{\gamma}} \partial_2 E_{\mathbf{h}}^{\mathbf{M}\mathbf{I}}))(\mathbf{I}_{\mathbf{h}}(\mathbf{x}), \mathbf{x}) \nabla I_2^{\sigma}(\mathbf{x} + \mathbf{h}(\mathbf{x})) \cdot \mathbf{k}(\mathbf{x}) d\mathbf{x}. \end{aligned}$$

This expression gives the gradient of $\mathcal{J}_{\mathbf{M}\mathbf{I}^l}(\mathbf{h})$:

$$\nabla_H \mathcal{J}_{\mathbf{M}\mathbf{I}^l}(\mathbf{h})(\mathbf{x}) = (G_{\gamma} \star (G_{\beta} \star \frac{1}{\mathcal{G}_{\gamma}} \partial_2 E_{\mathbf{h}}^{\mathbf{M}\mathbf{I}}))(\mathbf{I}_{\mathbf{h}}(\mathbf{x}), \mathbf{x}) \nabla I_2^{\sigma}(\mathbf{x} + \mathbf{h}(\mathbf{x})),$$

where

$$\partial_2 E_{\mathbf{h}}^{\mathbf{M}\mathbf{I}}(\mathbf{i}, \mathbf{x}) = - \left(\frac{\partial_2 P_{\mathbf{h}}(\mathbf{i}, \mathbf{x})}{P_{\mathbf{h}}(\mathbf{i}, \mathbf{x})} - \frac{p'_{\mathbf{h}}(i_2, \mathbf{x})}{p_{\mathbf{h}}(i_2, \mathbf{x})} \right).$$

We define the function $\mathbb{R}^2 \times \mathbb{R}^n \rightarrow \mathbb{R}$:

$$L_{\mathbf{M}\mathbf{I},\mathbf{h}}^l(\mathbf{i}, \mathbf{x}) \equiv \frac{1}{\mathcal{G}_{\gamma}(\mathbf{x})} \partial_2 E_{\mathbf{h}}^{\mathbf{M}\mathbf{I}}(\mathbf{i}, \mathbf{x}),$$

which plays exactly the same role as its global counterpart, $L_{\mathbf{M}\mathbf{I},\mathbf{h}}^g(\mathbf{i})$.

5.2.2 Correlatio Ratio

In this section, we compute the first variation of $\mathcal{J}_{\mathbf{CR}^l}(\mathbf{h})$ (definition 4.12 on page 68).

If we write

$$w(\mathbf{h}, \mathbf{x}_0) \equiv \int_{\mathbb{R}} v_{2|1}(i_1, \mathbf{h}, \mathbf{x}_0) p(i_1, \mathbf{x}_0) di_1,$$

so that

$$\begin{aligned} \mathcal{J}_{\mathbf{CR}^l}(\mathbf{h}) &= \int_{\Omega} \frac{w(\mathbf{h}, \mathbf{x}_0)}{v_2(\mathbf{h}, \mathbf{x}_0)} d\mathbf{x}_0 = \\ &\int_{\Omega} \left(\frac{1}{v_2(\mathbf{h}, \mathbf{x}_0)} \left(\int_{\mathbb{R}^2} i_2^2 P_{\mathbf{h}}(\mathbf{i}, \mathbf{x}_0) d\mathbf{i} - \int_{\mathbb{R}} \mu_{2|1}(i_1, \mathbf{h}, \mathbf{x}_0)^2 p(i_1, \mathbf{x}_0) di_1 \right) \right) d\mathbf{x}_0, \end{aligned}$$

we immediately see that we are in a situation completely analog to the global case, with the same modification as that of mutual information in the previous section, i.e. that we can write the first variation of $\mathcal{J}_{\mathbf{CR}^l}(\mathbf{h})$ as

$$\left. \frac{\partial \mathcal{J}_{\mathbf{CR}^l}(\mathbf{h} + \epsilon \mathbf{k})}{\partial \epsilon} \right|_{\epsilon=0} = \int_{\Omega} \int_{\mathbb{R}^2} E_{\mathbf{h}}^{\mathbf{CR}}(\mathbf{i}, \mathbf{x}_0) \left. \frac{\partial P_{\mathbf{h}+\epsilon \mathbf{k}}(\mathbf{i}, \mathbf{x}_0)}{\partial \epsilon} \right|_{\epsilon=0} d\mathbf{i} d\mathbf{x}_0,$$

where

$$E_{\mathbf{h}}^{\mathbf{CR}}(\mathbf{i}, \mathbf{x}_0) = \frac{i_2}{v_2(\mathbf{h}, \mathbf{x}_0)} \left(i_2 - 2\mu_{2|1}(i_1, \mathbf{h}, \mathbf{x}_0) - \frac{w(\mathbf{h}, \mathbf{x}_0)}{v_2(\mathbf{h}, \mathbf{x}_0)} (i_2 - 2\mu_2(\mathbf{h}, \mathbf{x}_0)) \right).$$

The discussion starting before equation (5.4) applies directly to this case. Thus, the gradient of $\mathcal{J}_{\mathbf{CR}^l}(\mathbf{h})$ is given by:

$$\nabla_{\mathbf{H}} \mathcal{J}_{\mathbf{CR}^l}(\mathbf{h})(\mathbf{x}) = (G_{\gamma} \star (G_{\beta} \star \frac{1}{\mathcal{G}_{\gamma}} \partial_2 E_{\mathbf{h}}^{\mathbf{CR}})) (\mathbf{I}_{\mathbf{h}}(\mathbf{x}), \mathbf{x}) \nabla I_2^{\sigma}(\mathbf{x} + \mathbf{h}(\mathbf{x})),$$

where

$$\partial_2 E_{\mathbf{h}}^{\mathbf{CR}}(\mathbf{i}, \mathbf{x}) = \frac{2}{v_2(\mathbf{h}, \mathbf{x})} \left(\mu_2(\mathbf{h}, \mathbf{x}) - \mu_{2|1}(i_1, \mathbf{h}, \mathbf{x}) + \mathbf{CR}^l(\mathbf{h}, \mathbf{x}) (i_2 - \mu_2(\mathbf{h}, \mathbf{x})) \right).$$

We define the function $\mathbb{R}^2 \times \mathbb{R}^n \rightarrow \mathbb{R}$:

$$L_{\mathbf{CR}, \mathbf{h}}^l(\mathbf{i}, \mathbf{x}) \equiv \frac{1}{\mathcal{G}_{\gamma}(\mathbf{x})} \partial_2 E_{\mathbf{h}}^{\mathbf{CR}}(\mathbf{i}, \mathbf{x}).$$

which plays exactly the same role as its global counterpart, $L_{\mathbf{CR}, \mathbf{h}}^g(\mathbf{i})$.

5.2.3 Cross Correlation

Again, the situation is quite similar in the case of $\mathcal{J}_{\mathbf{CC}^l}(\mathbf{h})$ (definition 4.1 on page 68).

We can put its first variation in the form

$$\left. \frac{\partial \mathcal{J}_{\mathbf{CC}^l}(\mathbf{h} + \epsilon \mathbf{k})}{\partial \epsilon} \right|_{\epsilon=0} = \int_{\Omega} \int_{\mathbb{R}^2} E_{\mathbf{h}}^{\mathbf{CC}}(\mathbf{i}, \mathbf{x}_0) \left. \frac{\partial P_{\mathbf{h}+\epsilon \mathbf{k}}(\mathbf{i}, \mathbf{x}_0)}{\partial \epsilon} \right|_{\epsilon=0} d\mathbf{i} d\mathbf{x}_0,$$

where

$$E_{\mathbf{h}}^{\text{CC}}(\mathbf{i}, \mathbf{x}_0) = \frac{-1}{v_1(\mathbf{x}_0) v_2(\mathbf{h}, \mathbf{x}_0)}$$

$$\left(2 v_{1,2}(\mathbf{h}, \mathbf{x}_0) i_2 (i_1 - \mu_1(\mathbf{x}_0)) - \mathbf{C}\mathbf{C}^l(\mathbf{h}, \mathbf{x}_0) v_1(\mathbf{x}_0) i_2 (i_2 - 2 \mu_2(\mathbf{h}, \mathbf{x}_0)) \right).$$

The discussion starting at equation (5.4) remains valid, and therefore the gradient of $\mathcal{J}_{\mathbf{C}\mathbf{C}^l}(\mathbf{h})$ is given by:

$$\nabla_{\mathbf{H}} \mathcal{J}_{\mathbf{C}\mathbf{C}^l}(\mathbf{h})(\mathbf{x}) = \left(G_{\gamma} \star \left(G_{\beta} \star \frac{1}{\mathcal{G}_{\gamma}} \partial_2 E_{\mathbf{h}}^{\text{CC}} \right) \right) (\mathbf{I}_{\mathbf{h}}(\mathbf{x}), \mathbf{x}) \nabla I_2^{\sigma}(\mathbf{x} + \mathbf{h}(\mathbf{x})),$$

where

$$\partial_2 E_{\mathbf{h}}^{\text{CC}}(\mathbf{i}, \mathbf{x}) = -2 \left[\frac{v_{1,2}(\mathbf{h}, \mathbf{x})}{v_2(\mathbf{h}, \mathbf{x})} \left(\frac{i_1 - \mu_1(\mathbf{x})}{v_1(\mathbf{x})} \right) - \mathbf{C}\mathbf{C}^l(\mathbf{h}, \mathbf{x}) \left(\frac{i_2 - \mu_2(\mathbf{h}, \mathbf{x})}{v_2(\mathbf{h}, \mathbf{x})} \right) \right].$$

Like in the global case, the convolution in the intensity domain yields the same expression (since it is linear in \mathbf{i}). Thus,

$$\nabla_{\mathbf{H}} \mathcal{J}_{\mathbf{C}\mathbf{C}^l}(\mathbf{h})(\mathbf{x}) = \left(G_{\gamma} \star \frac{1}{\mathcal{G}_{\gamma}} \partial_2 E_{\mathbf{h}}^{\text{CC}} \right) (\mathbf{I}_{\mathbf{h}}(\mathbf{x}), \mathbf{x}) \nabla I_2^{\sigma}(\mathbf{x} + \mathbf{h}(\mathbf{x})).$$

We define the function $\mathbb{R}^2 \times \mathbb{R}^n \rightarrow \mathbb{R}$:

$$L_{\mathbf{C}\mathbf{C},\mathbf{h}}^l(\mathbf{i}, \mathbf{x}) \equiv \frac{1}{\mathcal{G}_{\gamma}(\mathbf{x})} \partial_2 E_{\mathbf{h}}^{\text{CC}}(\mathbf{i}, \mathbf{x}).$$

which plays the same role as its global counterpart, $L_{\mathbf{C}\mathbf{C},\mathbf{h}}^g(\mathbf{i})$.

5.3 Summary

We now summarize the results of this chapter by defining the functions that will be used to specify the function F in the generic matching flow (equation (2.4) on page 47), depending on the various dissimilarity criteria.

Theorem 5.1 *The infinitesimal gradient of the global dissimilarity criteria is given by*

$$F^g(\mathbf{h})(\mathbf{x}) = \nabla_{\mathbf{H}} \mathcal{J}^g(\mathbf{h})(\mathbf{x}) = (G_{\beta} \star L_{\mathbf{h}}^g)(\mathbf{I}_{\mathbf{h}}(\mathbf{x})) \nabla I_2^{\sigma}(\mathbf{x} + \mathbf{h}(\mathbf{x})), \quad (5.5)$$

where the function $L_{\mathbf{h}}^g(\mathbf{i})$ is equal to

$$L_{MI,\mathbf{h}}^g(\mathbf{i}) = -\frac{1}{|\Omega|} \left(\frac{\partial_2 P_{\mathbf{h}}(\mathbf{i})}{P_{\mathbf{h}}(\mathbf{i})} - \frac{p'_{\mathbf{h}}(i_2)}{p_{\mathbf{h}}(i_2)} \right) \quad (5.6)$$

in the case of the mutual information, to

$$L_{CR,\mathbf{h}}^g(\mathbf{i}) = \frac{\mu_2(\mathbf{h}) - \mu_{2|1}(i_1, \mathbf{h}) + \mathbf{C}\mathbf{R}^g(\mathbf{h}) (i_2 - \mu_2(\mathbf{h}))}{\frac{1}{2} |\Omega| v_2(\mathbf{h})} \quad (5.7)$$

in the case of the correlation ratio and to

$$L_{CC,\mathbf{h}}^g(\mathbf{i}) = -\frac{2}{|\Omega|} \left(\frac{v_{1,2}(\mathbf{h})}{v_2(\mathbf{h})} \left(\frac{i_1 - \mu_1}{v_1} \right) - \mathbf{CC}^g(\mathbf{h}) \left(\frac{i_2 - \mu_2(\mathbf{h})}{v_2(\mathbf{h})} \right) \right) \quad (5.8)$$

in the case of the cross correlation. This last case is especially simple since

$$G_\beta \star L_{CC,\mathbf{h}}^g = L_{CC,\mathbf{h}}^g,$$

so that no convolution is required.

This defines three functions $H \rightarrow H$:

$$\begin{aligned} F_{MI}^g(\mathbf{h}) &= (G_\beta \star L_{MI,\mathbf{h}}^g)(I_1^\sigma, I_2^\sigma(\mathbf{Id} + \mathbf{h})) \nabla I_2^\sigma(\mathbf{Id} + \mathbf{h}), \\ F_{CR}^g(\mathbf{h}) &= (G_\beta \star L_{CR,\mathbf{h}}^g)(I_1^\sigma, I_2^\sigma(\mathbf{Id} + \mathbf{h})) \nabla I_2^\sigma(\mathbf{Id} + \mathbf{h}), \\ F_{CC}^g(\mathbf{h}) &= L_{CC,\mathbf{h}}^g(I_1^\sigma, I_2^\sigma(\mathbf{Id} + \mathbf{h})) \nabla I_2^\sigma(\mathbf{Id} + \mathbf{h}). \end{aligned} \quad (5.9)$$

Proof : The only point that has not been proved is the fact that the functions $F_{MI}^g(\mathbf{h})$, $F_{CR}^g(\mathbf{h})$ and $F_{CC}^g(\mathbf{h})$ belong to H . This is a consequence of theorems 6.13, 6.25 and 6.31, respectively. \square

It is worth clarifying how equation (5.5) is interpreted. The function $L_{\mathbf{h}}^g : \mathbb{R}^2 \rightarrow \mathbb{R}$ is convolved with the 2D gaussian G_β and the result is *evaluated* at the intensity pair $(I_1^\sigma(\mathbf{x}), I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})))$. The value of the gradient at the point \mathbf{x} is then obtained by multiplying this value by the gradient of the second image at the point $\mathbf{x} + \mathbf{h}(\mathbf{x})$. For the global cross correlation, no convolution is required.

The case of the local criteria is very similar.

Theorem 5.2 *The infinitesimal gradient of the local criteria is given by*

$$F^l(\mathbf{h})(\mathbf{x}) = \nabla_H \mathcal{J}^l(\mathbf{h})(\mathbf{x}) = (G_\gamma \star (G_\beta \star L_{\mathbf{h}}^l))(\mathbf{I}_{\mathbf{h}}(\mathbf{x}), \mathbf{x}) \nabla I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})), \quad (5.10)$$

where the function $L_{\mathbf{h}}^l(i_1, i_2, \mathbf{x})$ is equal to

$$L_{MI,\mathbf{h}}^l(\mathbf{i}, \mathbf{x}) = -\frac{1}{\mathcal{G}_\gamma(\mathbf{x})} \left(\frac{\partial_2 P_{\mathbf{h}}(\mathbf{i}, \mathbf{x})}{P_{\mathbf{h}}(\mathbf{i}, \mathbf{x})} - \frac{p_{\mathbf{h}}'(i_2, \mathbf{x})}{p_{\mathbf{h}}(i_2, \mathbf{x})} \right), \quad (5.11)$$

in the case of the mutual information, to

$$L_{CR,\mathbf{h}}^l(\mathbf{i}, \mathbf{x}) = \frac{\mu_2(\mathbf{h}, \mathbf{x}) - \mu_{2|1}(i_1, \mathbf{h}, \mathbf{x}) + \mathbf{CR}^l(\mathbf{h}, \mathbf{x}) (i_2 - \mu_2(\mathbf{h}, \mathbf{x}))}{\frac{1}{2} \mathcal{G}_\gamma(\mathbf{x}) v_2(\mathbf{h}, \mathbf{x})} \quad (5.12)$$

in the case of the correlation ratio, and to

$$\begin{aligned} L_{CC,\mathbf{h}}^l(\mathbf{i}, \mathbf{x}) &= \\ &= -\frac{2}{\mathcal{G}_\gamma(\mathbf{x})} \left(\frac{v_{1,2}(\mathbf{h}, \mathbf{x})}{v_2(\mathbf{h}, \mathbf{x})} \left(\frac{i_1 - \mu_1(\mathbf{x})}{v_1(\mathbf{x})} \right) - \mathbf{CC}^l(\mathbf{h}, \mathbf{x}) \left(\frac{i_2 - \mu_2(\mathbf{h}, \mathbf{x})}{v_2(\mathbf{h}, \mathbf{x})} \right) \right) \end{aligned} \quad (5.13)$$

in the case of the cross correlation. The case of the cross correlation is especially simple since

$$G_\beta \star L_{CC,\mathbf{h}}^l = L_{CC,\mathbf{h}}^l,$$

so that only the spatial convolution is necessary.

This defines three functions $H \rightarrow H$:

$$\begin{aligned} F_{MI}^l(\mathbf{h}) &= (G_\gamma \star G_\beta \star L_{MI,\mathbf{h}}^l)(I_1^\sigma, I_2^\sigma(\mathbf{Id} + \mathbf{h}), \mathbf{Id}) \nabla I_2^\sigma(\mathbf{Id} + \mathbf{h}), \\ F_{CR}^l(\mathbf{h}) &= (G_\gamma \star G_\beta \star L_{CR,\mathbf{h}}^l)(I_1^\sigma, I_2^\sigma(\mathbf{Id} + \mathbf{h}), \mathbf{Id}) \nabla I_2^\sigma(\mathbf{Id} + \mathbf{h}), \\ F_{CC}^l(\mathbf{h}) &= (G_\gamma \star L_{CC,\mathbf{h}}^l)(I_1^\sigma, I_2^\sigma(\mathbf{Id} + \mathbf{h}), \mathbf{Id}) \nabla I_2^\sigma(\mathbf{Id} + \mathbf{h}). \end{aligned} \quad (5.14)$$

Proof : The fact that $F_{MI}^l(\mathbf{h})$, $F_{CR}^l(\mathbf{h})$ and $F_{CC}^l(\mathbf{h})$ belong to H is a consequence of theorems 6.41, 6.57 and 6.64, respectively. \square

It is worth clarifying how equation (5.10) is interpreted. The function $L_{\mathbf{h}}^l : \mathbb{R}^2 \times \mathbb{R}^n \rightarrow \mathbb{R}$ is convolved with the 2D gaussian G_β for the first two variables (intensities) and the nD gaussian G_γ for the remaining n variables (spatial), and the result is *evaluated* at the point $((I_1^\sigma(\mathbf{x}), I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x}))), \mathbf{x})$ of $\mathbb{R}^2 \times \mathbb{R}^n$. The value of the gradient at the point \mathbf{x} is then obtained by multiplying this value by the gradient of the second image at the point $\mathbf{x} + \mathbf{h}(\mathbf{x})$. For the local cross correlation, only the spatial convolution is required.

Chapter 6

Properties of the Matching Terms

This chapter is devoted to showing that the gradients of the statistical criteria computed in the previous chapter satisfy the Lipschitz-continuity conditions established in Chapter 2, necessary to assert the well-posedness of the evolution equations.

6.1 Preliminary Results

We begin by some elementary results on Lipschitz-continuous functions that will be used very often in the sequel.

Proposition 6.1 *Let \mathcal{H} be a Banach space and let us denote its norm by $\|\cdot\|_{\mathcal{H}}$. Let $f_i, i = 1, 2 : \mathcal{H} \rightarrow \mathbb{R}$ be two Lipschitz continuous functions. We have the following:*

1. $f_1 + f_2$ is Lipschitz continuous.
2. If f_1 and f_2 are bounded then the product $f_1 f_2$ is Lipschitz continuous.
3. If $f_2 > 0$ and if f_1 and f_2 are bounded, then the ratio $\frac{f_1}{f_2}$ is Lipschitz continuous.

Proof : We prove only 2 and 3. Let \mathbf{h} and \mathbf{h}' be two vectors of \mathcal{H} :

$$\begin{aligned} |f_1(\mathbf{h})f_2(\mathbf{h}) - f_1(\mathbf{h}')f_2(\mathbf{h}')| &= \\ |(f_1(\mathbf{h}) - f_1(\mathbf{h}'))f_2(\mathbf{h}) + f_1(\mathbf{h}')(f_2(\mathbf{h}) - f_2(\mathbf{h}'))| &\leq \\ |f_2(\mathbf{h})| |f_1(\mathbf{h}) - f_1(\mathbf{h}')| + |f_1(\mathbf{h}')| |f_2(\mathbf{h}) - f_2(\mathbf{h}')|, \end{aligned}$$

from which point 2 above follows. Similarly

$$\begin{aligned} \left| \frac{f_1(\mathbf{h})}{f_2(\mathbf{h})} - \frac{f_1(\mathbf{h}')}{f_2(\mathbf{h}')} \right| &= \frac{|f_1(\mathbf{h})f_2(\mathbf{h}') - f_2(\mathbf{h})f_1(\mathbf{h}')|}{f_2(\mathbf{h})f_2(\mathbf{h}')} \leq \\ &= \frac{|f_1(\mathbf{h}) - f_1(\mathbf{h}')|f_2(\mathbf{h}') + |f_1(\mathbf{h}')| |f_2(\mathbf{h}) - f_2(\mathbf{h}')|}{f_2(\mathbf{h})f_2(\mathbf{h}')} \end{aligned}$$

If $f_2 > 0$, there exists $a > 0$ such that $f_2 > a$. Hence

$$\left| \frac{f_1(\mathbf{h})}{f_2(\mathbf{h})} - \frac{f_1(\mathbf{h}')}{f_2(\mathbf{h}')} \right| \leq \frac{1}{a^2} |f_1(\mathbf{h}) - f_1(\mathbf{h}')| f_2(\mathbf{h}') + |f_1(\mathbf{h}')| |f_2(\mathbf{h}) - f_2(\mathbf{h}')|,$$

from which point 3 above follows. \square

In the following, we will need the following definitions and notations.

Definition 6.1 We note $\mathcal{H}_1 = [0, \mathcal{A}] \times H$ and $\mathcal{H}_2 = [0, \mathcal{A}]^2 \times H$ the Banach spaces equipped with the norms $\|(z, \mathbf{h})\|_{\mathcal{H}_1} = |z| + \|\mathbf{h}\|_H$ and $\|(z_1, z_2, \mathbf{h})\|_{\mathcal{H}_2} = |z_1| + |z_2| + \|\mathbf{h}\|_H$, respectively.

We will use several times the following result.

Lemma 6.2 Let $f : \mathcal{H}_2 \rightarrow \mathbb{R}$ be such that $(z_1, z_2) \rightarrow f(z_1, z_2, \mathbf{h})$ is Lipschitz continuous with a Lipschitz constant l_f independent of \mathbf{h} and such that $\mathbf{h} \rightarrow f(z_1, z_2, \mathbf{h})$ is Lipschitz continuous with a Lipschitz constant L_f independent of (z_1, z_2) . Then f is Lipschitz continuous.

Proof : We have

$$\begin{aligned} |f(z_1, z_2, \mathbf{h}) - f(z'_1, z'_2, \mathbf{h}')| &\leq \\ &|f(z_1, z_2, \mathbf{h}) - f(z'_1, z'_2, \mathbf{h})| + |f(z'_1, z'_2, \mathbf{h}) - f(z'_1, z'_2, \mathbf{h}')| \leq \\ &l_f(|z_1 - z'_1| + |z_2 - z'_2|) + L_f \|\mathbf{h} - \mathbf{h}'\|_H \leq \\ &\max(l_f, L_f)(|z_1 - z'_1| + |z_2 - z'_2| + \|\mathbf{h} - \mathbf{h}'\|_H). \end{aligned}$$

\square

In Section 6.3, we will need a slightly more general version of this lemma.

Lemma 6.3 Let $f : [0, \mathcal{A}]^2 \times H \times \Omega \rightarrow \mathbb{R}$ be such that $(z_1, z_2) \rightarrow f(z_1, z_2, \mathbf{h}, \mathbf{x})$ is Lipschitz continuous with a Lipschitz constant l_f independent of \mathbf{x} and \mathbf{h} and such that $\mathbf{h} \rightarrow f(z_1, z_2, \mathbf{h}, \mathbf{x})$ is Lipschitz continuous with a Lipschitz constant L_f independent of (z_1, z_2, \mathbf{x}) . Then f is Lipschitz continuous on $[0, \mathcal{A}]^2 \times H$ uniformly on Ω .

Proof : Indeed,

$$\begin{aligned} |f(z_1, z_2, \mathbf{h}, \mathbf{x}) - f(z'_1, z'_2, \mathbf{h}, \mathbf{x}')| &\leq \\ &|f(z_1, z_2, \mathbf{h}, \mathbf{x}) - f(z'_1, z'_2, \mathbf{h}, \mathbf{x})| + |f(z'_1, z'_2, \mathbf{h}, \mathbf{x}) - f(z'_1, z'_2, \mathbf{h}')| \leq \\ &l_f(|z_1 - z'_1| + |z_2 - z'_2|) + L_f \|\mathbf{h} - \mathbf{h}'\|_H \leq \\ &\max(l_f, L_f)(|z_1 - z'_1| + |z_2 - z'_2| + \|\mathbf{h} - \mathbf{h}'\|_H) \quad \forall \mathbf{x} \in \Omega, \end{aligned}$$

and the Lipschitz constant $\max(l_f, L_f)$ is independent of $\mathbf{x} \in \Omega$. \square

6.2 Global Criteria

We show in this section the Lipschitz continuity of the gradients of the global criteria, as defined in (5.9).

6.2.1 Mutual Information

We first prove that in the mutual information case, there is a neat separation in the definition of the function F between its local and global dependency in the field \mathbf{h} . More precisely we have the following.

Proposition 6.4 *The function $q_{\mathbf{h}} : \mathbb{R} \rightarrow \mathbb{R}$ defined by*

$$q_{\mathbf{h}}(i_2) = \frac{p'_{\mathbf{h}}(i_2)}{p_{\mathbf{h}}(i_2)}$$

satisfies the following equation:

$$q_{\mathbf{h}}(i_2) = a(i_2, \mathbf{h}) - \frac{i_2}{\beta},$$

where the function $0 \leq a(i_2, \mathbf{h}) \leq \frac{\mathcal{A}}{\beta}$.

Proof : $p_{\mathbf{h}}$ is defined by

$$p_{\mathbf{h}}(i_2) = \int_{\Omega} g_{\beta}(I_2^{\sigma}(\mathbf{x} + \mathbf{h}(\mathbf{x})) - i_2) d\mathbf{x},$$

hence

$$p'_{\mathbf{h}}(i_2) = \frac{1}{\beta} \int_{\Omega} (I_2^{\sigma}(\mathbf{x} + \mathbf{h}(\mathbf{x})) - i_2) g_{\beta}(I_2^{\sigma}(\mathbf{x} + \mathbf{h}(\mathbf{x})) - i_2) d\mathbf{x}.$$

The function $a(i_2, \mathbf{h})$ is equal to

$$a(i_2, \mathbf{h}) = \frac{1}{\beta} \frac{\int_{\Omega} I_2^{\sigma}(\mathbf{x} + \mathbf{h}(\mathbf{x})) g_{\beta}(I_2^{\sigma}(\mathbf{x} + \mathbf{h}(\mathbf{x})) - i_2) d\mathbf{x}}{\int_{\Omega} g_{\beta}(I_2^{\sigma}(\mathbf{x} + \mathbf{h}(\mathbf{x})) - i_2) d\mathbf{x}}, \quad (6.1)$$

and the result follows from the fact that $I_2^{\sigma}(\mathbf{x} + \mathbf{h}(\mathbf{x})) \in [0, \mathcal{A}]$. \square

A simple variation of the previous proof shows the truth of the following.

Proposition 6.5 *The function $Q_{\mathbf{h}} : \mathbb{R}^2 \rightarrow \mathbb{R}$ defined by*

$$Q_{\mathbf{h}}(\mathbf{i}) = \frac{\partial_2 P_{\mathbf{h}}(\mathbf{i})}{P_{\mathbf{h}}(\mathbf{i})}$$

satisfies the following equation:

$$Q_{\mathbf{h}}(\mathbf{i}) = A(\mathbf{i}, \mathbf{h}) - \frac{i_2}{\beta},$$

where function

$$A(\mathbf{i}, \mathbf{h}) = \frac{1}{\beta} \frac{\int_{\Omega} I_2^{\sigma}(\mathbf{x} + \mathbf{h}(\mathbf{x})) G_{\beta}(\mathbf{I}_{\mathbf{h}}(\mathbf{x}) - \mathbf{i}) d\mathbf{x}}{\int_{\Omega} G_{\beta}(\mathbf{I}_{\mathbf{h}}(\mathbf{x}) - \mathbf{i}) d\mathbf{x}}, \quad (6.2)$$

satisfies $0 \leq A(\mathbf{i}, \mathbf{h}) \leq \frac{\mathcal{A}}{\beta}$.

In the following, we will use the function $L_{\text{MI}, \mathbf{h}}^g : \mathbb{R}^2 \rightarrow \mathbb{R}$ defined as (see theorem 5.1)

$$L_{\text{MI}, \mathbf{h}}^g(\mathbf{i}) = -\frac{1}{|\Omega|} (Q_{\mathbf{h}}(\mathbf{i}) - q_{\mathbf{h}}(i_2)) = -\frac{1}{|\Omega|} (A(\mathbf{i}, \mathbf{h}) - a(i_2, \mathbf{h})). \quad (6.3)$$

We then consider the result f_{MI}^g of convolving $L_{\text{MI}, \mathbf{h}}^g$ with G_{β} , i.e. the two functions $b : \mathbb{R} \times H \rightarrow \mathbb{R}$ defined as

$$b(z_2, \mathbf{h}) = (g_{\beta} \star a)(z_2, \mathbf{h}) = \int_{\mathbb{R}} g_{\beta}(z_2 - i_2) a(i_2, \mathbf{h}) di_2, \quad (6.4)$$

and $B : \mathbb{R}^2 \times H \rightarrow \mathbb{R}$ defined as

$$B(z_1, z_2, \mathbf{h}) = (G_{\beta} \star A)(\mathbf{z}, \mathbf{h}) = \int_{\mathbb{R}^2} G_{\beta}(\mathbf{z} - \mathbf{i}) A(\mathbf{i}, \mathbf{h}) d\mathbf{i}. \quad (6.5)$$

We prove a series of propositions.

Proposition 6.6 *The function $\mathbb{R} \rightarrow \mathbb{R}^+$ defined by $z_2 \rightarrow b(z_2, \mathbf{h})$ is Lipschitz continuous with a Lipschitz constant $\frac{\mathcal{A}}{\beta}$ which is independent of \mathbf{h} . Moreover, it is bounded by $\frac{\mathcal{A}}{\beta}$.*

Proof : The second part of the proposition follows from the fact that $0 \leq a(i_2, \mathbf{h}) \leq \frac{\mathcal{A}}{\beta} \forall i_2 \in \mathbb{R}$ and $\forall \mathbf{h} \in H$ (proposition 6.4).

In order to prove the first part, we prove that the magnitude of the derivative of the function is bounded independently of \mathbf{h} . Indeed

$$|b'(z_2, \mathbf{h})| = \frac{1}{\beta} \left| \int_{-\infty}^{+\infty} (z_2 - i_2) g_{\beta}(z_2 - i_2) a(i_2, \mathbf{h}) di_2 \right| \leq \frac{\mathcal{A}}{\beta} \int_{-\infty}^{+\infty} |z_2 - i_2| g_{\beta}(z_2 - i_2) di_2.$$

The function on the right-hand side of the inequality is independent of \mathbf{h} and continuous on $[0, \mathcal{A}]$, therefore upperbounded. \square

Proposition 6.7 *The function $\mathbf{h} \longrightarrow b(z_2, \mathbf{h}) : \mathbf{L}^2(\Omega) \longrightarrow \mathbb{R}$ is Lipschitz continuous on $\mathbf{L}^2(\Omega)$ with Lipschitz constant L_b^g , which is independent of $z_2 \in [0, A]$.*

Proof : We consider

$$b(z_2, \mathbf{h}_1) - b(z_2, \mathbf{h}_2) = \int_{\mathbb{R}} g_{\beta}(z_2 - i_2) (a(i_2, \mathbf{h}_1) - a(i_2, \mathbf{h}_2)) di_2 \quad (6.6)$$

According to equation (6.1), $a(i_2, \mathbf{h})$ is the ratio $N(i_2, \mathbf{h})/D(i_2, \mathbf{h})$ of the two functions

$$N(i_2, \mathbf{h}) = \int_{\Omega} I_2^{\sigma}(\mathbf{x} + \mathbf{h}(\mathbf{x})) g_{\beta}(I_2^{\sigma}(\mathbf{x} + \mathbf{h}(\mathbf{x})) - i_2) d\mathbf{x},$$

and

$$D(i_2, \mathbf{h}) = \int_{\Omega} g_{\beta}(I_2^{\sigma}(\mathbf{x} + \mathbf{h}(\mathbf{x})) - i_2) d\mathbf{x}.$$

We ignore the factor $1/\beta$ which is irrelevant in the proof. We write

$$\begin{aligned} |b(z_2, \mathbf{h}_1) - b(z_2, \mathbf{h}_2)| \leq & \int_{\mathbb{R}} g_{\beta}(z_2 - i_2) \frac{|N(i_2, \mathbf{h}_2)| |D(i_2, \mathbf{h}_2) - D(i_2, \mathbf{h}_1)|}{D(i_2, \mathbf{h}_1)D(i_2, \mathbf{h}_2)} di_2 + \\ & \int_{\mathbb{R}} g_{\beta}(z_2 - i_2) \frac{D(i_2, \mathbf{h}_2) |N(i_2, \mathbf{h}_2) - N(i_2, \mathbf{h}_1)|}{D(i_2, \mathbf{h}_1)D(i_2, \mathbf{h}_2)} di_2, \end{aligned} \quad (6.7)$$

and consider the first term of the right-hand side.

$$\begin{aligned} D(i_2, \mathbf{h}_2) - D(i_2, \mathbf{h}_1) = & \int_{\Omega} (g_{\beta}(i_2 - I_2^{\sigma}(\mathbf{x} + \mathbf{h}_2(\mathbf{x}))) - g_{\beta}(i_2 - I_2^{\sigma}(\mathbf{x} + \mathbf{h}_1(\mathbf{x})))) d\mathbf{x} \end{aligned}$$

We use the first order Taylor expansion with integral remainder of the C^1 function g_{β} . This says that

$$g_{\beta}(i + t) = g_{\beta}(i) + t \int_0^1 g'_{\beta}(i + t\alpha) d\alpha,$$

as the reader will easily verify. We can therefore write

$$\begin{aligned} g_{\beta}(i_2 - I_2^{\sigma}(\mathbf{x} + \mathbf{h}_2(\mathbf{x}))) - g_{\beta}(i_2 - I_2^{\sigma}(\mathbf{x} + \mathbf{h}_1(\mathbf{x}))) = & \\ & (I_2^{\sigma}(\mathbf{x} + \mathbf{h}_1(\mathbf{x})) - I_2^{\sigma}(\mathbf{x} + \mathbf{h}_2(\mathbf{x}))) \int_0^1 g'_{\beta}(i_2 - (\alpha I_2^{\sigma}(\mathbf{x} + \mathbf{h}_2(\mathbf{x})) + \\ & (1 - \alpha) I_2^{\sigma}(\mathbf{x} + \mathbf{h}_1(\mathbf{x})))) d\alpha \end{aligned}$$

We use the fact that I_2^{σ} is Lipschitz continuous and write

$$\begin{aligned} |D(i_2, \mathbf{h}_2) - D(i_2, \mathbf{h}_1)| \leq & \\ & Lip(I_2^{\sigma}) \int_{\Omega} (|\mathbf{h}_1(\mathbf{x}) - \mathbf{h}_2(\mathbf{x})| \\ & \left| \int_0^1 g'_{\beta}(i_2 - (\alpha I_2^{\sigma}(\mathbf{x} + \mathbf{h}_2(\mathbf{x})) + (1 - \alpha) I_2^{\sigma}(\mathbf{x} + \mathbf{h}_1(\mathbf{x})))) d\alpha \right|) d\mathbf{x} \end{aligned}$$

Schwarz inequality implies

$$|D(i_2, \mathbf{h}_2) - D(i_2, \mathbf{h}_1)| \leq Lip(I_2^\sigma) \|\mathbf{h}_1 - \mathbf{h}_2\|_H$$

$$\left(\int_{\Omega} \left(\int_0^1 g'_\beta(i_2 - (\alpha I_2^\sigma(\mathbf{x} + \mathbf{h}_2(\mathbf{x})) + (1 - \alpha) I_2^\sigma(\mathbf{x} + \mathbf{h}_1(\mathbf{x})))) d\alpha \right)^2 dx \right)^{\frac{1}{2}}$$

We introduce the function

$$r(i_2, \mathbf{h}_1, \mathbf{h}_2) =$$

$$\left(\int_{\Omega} \left(\int_0^1 |i_2 - (\alpha I_2^\sigma(\mathbf{x} + \mathbf{h}_2(\mathbf{x})) + (1 - \alpha) I_2^\sigma(\mathbf{x} + \mathbf{h}_1(\mathbf{x})))| \right. \right.$$

$$\left. \left. g_\beta(i_2 - (\alpha I_2^\sigma(\mathbf{x} + \mathbf{h}_2(\mathbf{x})) + (1 - \alpha) I_2^\sigma(\mathbf{x} + \mathbf{h}_1(\mathbf{x})))) d\alpha \right)^2 dx \right)^{\frac{1}{2}}$$

We notice that

$$\left(\int_{\Omega} \left(\int_0^1 g'_\beta(i_2 - (\alpha I_2^\sigma(\mathbf{x} + \mathbf{h}_2(\mathbf{x})) + (1 - \alpha) I_2^\sigma(\mathbf{x} + \mathbf{h}_1(\mathbf{x})))) d\alpha \right)^2 dx \right)^{\frac{1}{2}}$$

$$\leq \frac{1}{\beta} r(i_2, \mathbf{h}_1, \mathbf{h}_2).$$

So far we have

$$\int_{\mathbb{R}} g_\beta(z_2 - i_2) \frac{|N(i_2, \mathbf{h}_2)| |D(i_2, \mathbf{h}_2) - D(i_2, \mathbf{h}_1)|}{D(i_2, \mathbf{h}_1) D(i_2, \mathbf{h}_2)} di_2 \leq$$

$$\frac{Lip(I_2^\sigma)}{\beta} \|\mathbf{h}_1 - \mathbf{h}_2\|_H \left(\int_{\mathbb{R}} g_\beta(z_2 - i_2) \frac{|N(i_2, \mathbf{h}_2)| r_H(i_2, \mathbf{h}_1, \mathbf{h}_2)}{D(i_2, \mathbf{h}_1) D(i_2, \mathbf{h}_2)} di_2 \right) \quad (6.8)$$

We study the function of z_2 that is on the right-hand side of this inequality. First we note that the function is well defined since no problems occur when i_2 goes to infinity because "there are three gaussians in the numerator and two in the denominator". We then show that this function is bounded independently of \mathbf{h}_1 and \mathbf{h}_2 for all $z_2 \in [0, \mathcal{A}]$. We consider three cases:

$$i_2 \leq 0$$

This is the case where

$$0 \leq |i_2| \leq |i_2 - I_2^\sigma(\mathbf{x} + \mathbf{h}_j(\mathbf{x}))| \leq |i_2 - \mathcal{A}| \quad j = 1, 2$$

$$0 \leq |i_2| \leq |i_2 - (\alpha I_2^\sigma(\mathbf{x} + \mathbf{h}_2(\mathbf{x})) + (1 - \alpha) I_2^\sigma(\mathbf{x} + \mathbf{h}_1(\mathbf{x})))| \leq |i_2 - \mathcal{A}| \quad 0 \leq \alpha \leq 1$$

Hence

$$g_\beta(i_2 - \mathcal{A}) \leq g_\beta(i_2 - I_2^\sigma(\mathbf{x} + \mathbf{h}_j(\mathbf{x}))) \leq g_\beta(i_2) \quad j = 1, 2$$

$$g_\beta(i_2 - \mathcal{A}) \leq g_\beta(i_2 - (\alpha I_2^\sigma(\mathbf{x} + \mathbf{h}_2(\mathbf{x})) + (1 - \alpha) I_2^\sigma(\mathbf{x} + \mathbf{h}_1(\mathbf{x})))) \leq g_\beta(i_2) \quad 0 \leq \alpha \leq 1$$

This yields

$$\int_{-\infty}^0 g_\beta(z_2 - i_2) \frac{|N(i_2, \mathbf{h}_2)| r(i_2, \mathbf{h}_1, \mathbf{h}_2)}{D(i_2, \mathbf{h}_1)D(i_2, \mathbf{h}_2)} di_2 \leq$$

$$|\Omega|^{1/2} \mathcal{A} \int_{-\infty}^0 g_\beta(z_2 - i_2) |i_2 - \mathcal{A}| \left(\frac{g_\beta(i_2)}{g_\beta(i_2 - \mathcal{A})} \right)^2 di_2,$$

The integral on the right-hand side is well-defined and defines a continuous function of z_2 .

$$0 \leq i_2 \leq \mathcal{A}$$

$$0 \leq |i_2 - I_2^\sigma(\mathbf{x} + \mathbf{h}_j(\mathbf{x}))| \leq \mathcal{A} \quad j = 1, 2$$

$$0 \leq |i_2 - (\alpha I_2^\sigma(\mathbf{x} + \mathbf{h}_2(\mathbf{x})) + (1 - \alpha)I_2^\sigma(\mathbf{x} + \mathbf{h}_1(\mathbf{x})))| \leq \mathcal{A} \quad 0 \leq \alpha \leq 1$$

Hence

$$g_\beta(\mathcal{A}) \leq g_\beta(i_2 - I_2^\sigma(\mathbf{x} + \mathbf{h}_j(\mathbf{x}))) \leq g_\beta(0) \quad j = 1, 2$$

$$g_\beta(\mathcal{A}) \leq g_\beta(i_2 - (\alpha I_2^\sigma(\mathbf{x} + \mathbf{h}_2(\mathbf{x})) + (1 - \alpha)I_2^\sigma(\mathbf{x} + \mathbf{h}_1(\mathbf{x})))) \leq g_\beta(0) \quad 0 \leq \alpha \leq 1$$

This yields

$$\int_0^{\mathcal{A}} g_\beta(z_2 - i_2) \frac{|N(i_2, \mathbf{h}_2)| r(i_2, \mathbf{h}_1, \mathbf{h}_2)}{D(i_2, \mathbf{h}_1)D(i_2, \mathbf{h}_2)} di_2 \leq$$

$$|\Omega|^{1/2} \left(\frac{\mathcal{A} g_\beta(0)}{g_\beta(\mathcal{A})} \right)^2 \int_0^{\mathcal{A}} g_\beta(z_2 - i_2) di_2,$$

The integral on the right-hand side is convergent and defines a continuous function of z_2 .

$$i_2 \geq \mathcal{A}$$

This is the case where

$$0 \leq i_2 - \mathcal{A} \leq i_2 - I_2^\sigma(\mathbf{x} + \mathbf{h}_j(\mathbf{x})) \leq i_2 \quad j = 1, 2$$

$$0 \leq i_2 - \mathcal{A} \leq i_2 - (\alpha I_2^\sigma(\mathbf{x} + \mathbf{h}_2(\mathbf{x})) + (1 - \alpha)I_2^\sigma(\mathbf{x} + \mathbf{h}_1(\mathbf{x}))) \leq i_2 \quad 0 \leq \alpha \leq 1$$

Hence

$$g_\beta(i_2) \leq g_\beta(i_2 - I_2^\sigma(\mathbf{x} + \mathbf{h}_j(\mathbf{x}))) \leq g_\beta(i_2 - \mathcal{A}) \quad j = 1, 2$$

$$g_\beta(i_2) \leq g_\beta(i_2 - (\alpha I_2^\sigma(\mathbf{x} + \mathbf{h}_2(\mathbf{x})) + (1 - \alpha)I_2^\sigma(\mathbf{x} + \mathbf{h}_1(\mathbf{x})))) \leq g_\beta(i_2 - \mathcal{A}) \quad 0 \leq \alpha \leq 1$$

This yields

$$\int_{\mathcal{A}}^{+\infty} g_{\beta}(z_2 - i_2) \frac{|N(i_2, \mathbf{h}_2)| r(i_2, \mathbf{h}_1, \mathbf{h}_2)}{D(i_2, \mathbf{h}_1)D(i_2, \mathbf{h}_2)} di_2 \leq |\Omega|^{1/2} \mathcal{A} \int_{\mathcal{A}}^{+\infty} g_{\beta}(z_2 - i_2) i_2 \left(\frac{g_{\beta}(i_2 - \mathcal{A})}{g_{\beta}(i_2)} \right)^2 di_2,$$

The integral on the right-hand side is convergent and defines a continuous function of z_2 .

In all three cases, the functions of z_2 appearing on the right-hand side are continuous, independent of \mathbf{h}_1 and \mathbf{h}_2 , therefore upperbounded on $[0, \mathcal{A}]$ by a constant independent of \mathbf{h}_1 and \mathbf{h}_2 . Returning to inequality (6.8), we have proved that there existed a positive constant C independent of z_2 such that

$$\int_{\mathbb{R}} g_{\beta}(z_2 - i_2) \frac{|N(i_2, \mathbf{h}_2)| |D(i_2, \mathbf{h}_2) - D(i_2, \mathbf{h}_1)|}{D(i_2, \mathbf{h}_1)D(i_2, \mathbf{h}_2)} di_2 \leq C \|\mathbf{h}_1 - \mathbf{h}_2\|_H \quad \forall z_2 \in [0, \mathcal{A}] \quad \forall \mathbf{h}_1, \mathbf{h}_2 \in H$$

A similar proof can be developed for the second term in the right-hand side of the inequality (6.7). In conclusion we have proved that there existed a constant L_b^g , independent of z_2 such that

$$|b(z_2, \mathbf{h}_1) - b(z_2, \mathbf{h}_2)| \leq L_b^g \|\mathbf{h}_1 - \mathbf{h}_2\|_H \quad \forall z_2 \in [0, \mathcal{A}] \quad \forall \mathbf{h}_1, \mathbf{h}_2 \in H$$

□

Thus, we can state the following.

Proposition 6.8 *The function $b : \mathcal{H}_1 \rightarrow \mathbb{R}$ is Lipschitz continuous.*

Proof : The proof follows from propositions 6.6, 6.7 and lemma 6.2. □

We now proceed with showing the same kind of properties for the function B .

Proposition 6.9 *The function $[0, \mathcal{A}]^2 \rightarrow \mathbb{R}^+$ defined by $(z_1, z_2) \rightarrow B(z_1, z_2, \mathbf{h})$ is Lipschitz continuous with a Lipschitz constant L_B^g which is independent of \mathbf{h} . Moreover, it is bounded by $\frac{\mathcal{A}}{\beta}$.*

Proof : The second part of the proposition follows from the fact that, $\forall i_1, i_2 \in \mathbb{R}^2$ and $\forall \mathbf{h} \in H$, we have (proposition 6.5):

$$0 \leq A(i_1, i_2, \mathbf{h}) \leq \frac{\mathcal{A}}{\beta}.$$

The first part follows the same pattern as the proof of proposition 6.6. □

Proposition 6.10 *The function $\mathbf{h} \longrightarrow B(z_1, z_2, \mathbf{h}), H \longrightarrow \mathbb{R}$ is Lipschitz continuous with a Lipschitz constant L_B^g which is independent of $(z_1, z_2) \in [0, \mathcal{A}]^2$.*

Proof : The proof follows the same pattern as the one of proposition 6.7. \square

Therefore we can state the following result.

Proposition 6.11 *The function $B : \mathcal{H}_2 \rightarrow \mathbb{R}$ is Lipschitz continuous.*

Proof : The proof follows of propositions 6.9, 6.10 and lemma 6.2. \square

From propositions 6.8, 6.11 and 6.1 we obtain the following.

Corollary 6.12 *The function $f_{MI}^g : \mathcal{H}_2 \longrightarrow \mathbb{R}$ defined by $(z_1, z_2, \mathbf{h}) \longrightarrow -\frac{1}{|\Omega|}(B(z_1, z_2, \mathbf{h}) - b(z_2, \mathbf{h}))$ is Lipschitz continuous and bounded by $2\mathcal{A}/\beta|\Omega|$. We note $Lip(f_{MI}^g)$ the corresponding Lipschitz constant.*

We can now state the main result of this section:

Theorem 6.13 *The function $F_{MI}^g : H \longrightarrow H$ defined by*

$$F_{MI}^g(\mathbf{h}) = f_{MI}^g(I_1^\sigma, I_2^\sigma(\mathbf{Id} + \mathbf{h}), \mathbf{h})\nabla I_2^\sigma(\mathbf{Id} + \mathbf{h}) = \\ -\frac{1}{|\Omega|}(B(I_1^\sigma, I_2^\sigma(\mathbf{Id} + \mathbf{h}), \mathbf{h}) - b(I_2^\sigma(\mathbf{Id} + \mathbf{h}), \mathbf{h}))\nabla I_2^\sigma(\mathbf{Id} + \mathbf{h}),$$

is Lipschitz continuous and bounded.

Proof : Boundedness comes from the fact that b and B are bounded (propositions 6.6 and 6.9, respectively) and that $|\nabla I_2^\sigma|$ is bounded. This implies that $F_{MI}^g(\mathbf{h}) \in H = L^2(\Omega) \forall \mathbf{h} \in H$.

We consider the i th component $F_{MI}^{g,i}$ of F_{MI}^g :

$$F_{MI}^{g,i}(\mathbf{h}_1)(\mathbf{x}) - F_{MI}^{g,i}(\mathbf{h}_2)(\mathbf{x}) = -\frac{1}{|\Omega|}(S_1T_1 - S_2T_2),$$

with

$$S_j = B(I_1^\sigma(\mathbf{x}), I_2^\sigma(\mathbf{x} + \mathbf{h}_j(\mathbf{x})), \mathbf{h}_j) - b(I_2^\sigma(\mathbf{x} + \mathbf{h}_j(\mathbf{x})), \mathbf{h}_j) \\ T_j = \partial_i I_2^\sigma(\mathbf{x} + \mathbf{h}_j(\mathbf{x})),$$

and $j = 1, 2$. We continue with

$$|F_{MI}^{g,i}(\mathbf{h}_1)(\mathbf{x}) - F_{MI}^{g,i}(\mathbf{h}_2)(\mathbf{x})| \leq \frac{1}{|\Omega|}(|S_1 - S_2||T_1| + |S_2||T_1 - T_2|)$$

Because $\partial_i I_2^\sigma$ is bounded, $|T_j| \leq \|\partial_i I_2^\sigma\|_\infty$. Because of propositions 6.6 and 6.9, $|S_2| \leq 2\frac{\mathcal{A}}{\beta}$. Because $\partial_i I_2^\sigma$ is Lipschitz continuous $|T_1 - T_2| \leq Lip(\partial_i I_2^\sigma)|\mathbf{h}_1(\mathbf{x}) - \mathbf{h}_2(\mathbf{x})|$.

Finally, because of corollary 6.12 and the fact that I_2^g is Lipschitz continuous,

$$|S_1 - S_2| \leq Lip(f_{MI}^g) (Lip(I_2^\sigma) |\mathbf{h}_1(\mathbf{x}) - \mathbf{h}_2(\mathbf{x})| + \|\mathbf{h}_1 - \mathbf{h}_2\|_H).$$

Collecting all terms we obtain

$$|F_{MI}^{g^i}(\mathbf{h}_1)(\mathbf{x}) - F_{MI}^{g^i}(\mathbf{h}_2)(\mathbf{x})| \leq C^i (|\mathbf{h}_1(\mathbf{x}) - \mathbf{h}_2(\mathbf{x})| + \|\mathbf{h}_1 - \mathbf{h}_2\|_H),$$

for some positive constant C^i , $i = 1, \dots, n$. The last inequality yields, through the application of Cauchy-Schwarz:

$$\|F_{MI}^g(\mathbf{h}_1) - F_{MI}^g(\mathbf{h}_2)\|_H \leq L_F^g \|\mathbf{h}_1 - \mathbf{h}_2\|_H$$

for some positive constant L_F^g and this completes the proof. \square

The following proposition will be needed later.

Proposition 6.14 *The function $\Omega \rightarrow \mathbb{R}^n$ such that $\mathbf{x} \rightarrow F_{MI}^g(\mathbf{h}(\mathbf{x}))$ satisfies*

$$|F_{MI}^g(\mathbf{h}(\mathbf{x})) - F_{MI}^g(\mathbf{h}(\mathbf{y}))| \leq K (|\mathbf{x} - \mathbf{y}| + |\mathbf{h}(\mathbf{x}) - \mathbf{h}(\mathbf{y})|),$$

for some constant $K > 0$.

Proof : We write

$$\begin{aligned} F_{MI}^g(\mathbf{h}(\mathbf{x})) - F_{MI}^g(\mathbf{h}(\mathbf{y})) = & \\ & f_{MI}^g(\mathbf{h}(\mathbf{x})) \nabla I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) - f_{MI}^g(\mathbf{h}(\mathbf{x})) \nabla I_2^\sigma(\mathbf{y} + \mathbf{h}(\mathbf{y})) + \\ & f_{MI}^g(\mathbf{h}(\mathbf{x})) \nabla I_2^\sigma(\mathbf{y} + \mathbf{h}(\mathbf{y})) - f_{MI}^g(\mathbf{h}(\mathbf{y})) \nabla I_2^\sigma(\mathbf{y} + \mathbf{h}(\mathbf{y})). \end{aligned}$$

Hence

$$\begin{aligned} |F_{MI}^g(\mathbf{h}(\mathbf{x})) - F_{MI}^g(\mathbf{h}(\mathbf{y}))| \leq & \\ & |f_{MI}^g(\mathbf{h}(\mathbf{x}))| |\nabla I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) - \nabla I_2^\sigma(\mathbf{y} + \mathbf{h}(\mathbf{y}))| + \\ & |\nabla I_2^\sigma(\mathbf{y} + \mathbf{h}(\mathbf{y}))| |f_{MI}^g(\mathbf{h}(\mathbf{x})) - f_{MI}^g(\mathbf{h}(\mathbf{y}))|. \end{aligned}$$

Corollary 6.12 and the fact that the functions I_1^σ , I_2^σ and its first order derivative, are Lipschitz continuous imply

$$\begin{aligned} |F_{MI}^g(\mathbf{h}(\mathbf{x})) - F_{MI}^g(\mathbf{h}(\mathbf{y}))| \leq & \\ \frac{2\mathcal{A}}{\beta|\Omega|} Lip(\nabla I_2^\sigma) (|\mathbf{x} - \mathbf{y}| + |\mathbf{h}(\mathbf{x}) - \mathbf{h}(\mathbf{y})|) + & \|\nabla I_2^\sigma\|_\infty Lip(f_{MI}^g) |\mathbf{h}(\mathbf{x}) - \mathbf{h}(\mathbf{y})|, \end{aligned}$$

hence the result. \square

6.2.2 Correlation Ratio

We produce a simple expression of $\mathbf{CR}^g(\mathbf{h})$ in terms of the two images I_1^σ and I_2^σ and use it to prove that the correlation ratio is Lipschitz continuous as a function of \mathbf{h} . In the sequel, we drop the upper index g . We begin with some estimates.

Lemma 6.15

$$0 \leq \mu_2(\mathbf{h}) = \mathbf{E}[X_{I_2^\sigma, \mathbf{h}}] = \frac{1}{|\Omega|} \int_{\Omega} I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) \, d\mathbf{x} \leq \mathcal{A} \quad (6.9)$$

$$\beta \leq v_2(\mathbf{h}) = \mathbf{Var}[X_{I_2^\sigma, \mathbf{h}}] = \beta + \theta(\mathbf{h}) \leq \beta + \mathcal{A}^2, \quad (6.10)$$

where

$$\theta(\mathbf{h}) = \frac{1}{|\Omega|} \int_{\Omega} I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x}))^2 \, d\mathbf{x} - \left(\frac{1}{|\Omega|} \int_{\Omega} I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) \, d\mathbf{x} \right)^2. \quad (6.11)$$

Proof : Because of equation (4.4) we have

$$\mu_2(\mathbf{h}) = \frac{1}{|\Omega|} \int_{\mathbb{R}} i_2 \left(\int_{\Omega} g_\beta(i_2 - I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x}))) \, d\mathbf{x} \right) di_2 = \frac{1}{|\Omega|} \int_{\Omega} I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) \, d\mathbf{x}.$$

This yields the first part of the lemma. For the second part, we use equation (4.5):

$$v_2(\mathbf{h}) = \mathbf{Var}[X_{I_2^\sigma, \mathbf{h}}] = \int_{\mathbb{R}} i_2^2 \left(\frac{1}{|\Omega|} \int_{\Omega} g_\beta(i_2 - I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x}))) \, d\mathbf{x} \right) di_2 - \mu_2(\mathbf{h})^2,$$

and hence

$$\mathbf{Var}[X_{I_2^\sigma, \mathbf{h}}] = \beta + \frac{1}{|\Omega|} \int_{\Omega} I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x}))^2 \, d\mathbf{x} - \left(\frac{1}{|\Omega|} \int_{\Omega} I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) \, d\mathbf{x} \right)^2,$$

from which the upper and lower bounds of the lemma follow. \square

We next take care of $\mathbf{E}[\mathbf{Var}[X_{I_2^\sigma, \mathbf{h}} | X_{I_1^\sigma}]]$ with the following lemma.

Lemma 6.16

$$\mathbf{E}[\mathbf{Var}[X_{I_2^\sigma, \mathbf{h}} | X_{I_1^\sigma}]] = \beta + \frac{1}{|\Omega|} \int_{\Omega} I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x}))^2 \, d\mathbf{x} - M(\mathbf{h}),$$

where

$$M(\mathbf{h}) = \int_{\Omega \times \Omega} f(\mathbf{x}, \mathbf{x}') I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) I_2^\sigma(\mathbf{x}' + \mathbf{h}(\mathbf{x}')) \, d\mathbf{x} \, d\mathbf{x}',$$

and

$$f(\mathbf{x}, \mathbf{x}') = \frac{1}{|\Omega|^2} \int_{\mathbb{R}} \frac{g_\beta(i_1 - I_1^\sigma(\mathbf{x})) g_\beta(i_1 - I_1^\sigma(\mathbf{x}'))}{p(i_1)} di_1,$$

is such that $\int_{\Omega} f(\mathbf{x}, \mathbf{x}') \, d\mathbf{x} = \int_{\Omega} f(\mathbf{x}, \mathbf{x}') \, d\mathbf{x}' = \frac{1}{|\Omega|}$.

Proof : According to Table 4.1 we have $v_{2|1}(i_1, \mathbf{h}) = S(i_1) - T^2(i_1)$ and

$$\mathbf{E}[\mathbf{Var}[X_{I_2^\sigma, \mathbf{h}} | X_{I_1^\sigma}]] = \int_{\mathbb{R}} v_{2|1}(i_1, \mathbf{h}) p(i_1) di_1,$$

with

$$S(i_1) = \int_{\mathbb{R}} i_2^2 \frac{P_{\mathbf{h}}(\mathbf{i})}{p(i_1)} di_2,$$

and

$$T(i_1) = \int_{\mathbb{R}} i_2 \frac{P_{\mathbf{h}}(\mathbf{i})}{p(i_1)} di_2 = \mu_{2|1}(i_1, \mathbf{h}).$$

It is straightforward to show that

$$\int_{\mathbb{R}} S(i_1) p(i_1) di_1 = \beta + \frac{1}{|\Omega|} \int_{\Omega} I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x}))^2 d\mathbf{x}.$$

It is also straightforward to show that

$$T(i_1) = \frac{1}{|\Omega| p(i_1)} \int_{\Omega} g_\beta(i_1 - I_1^\sigma(\mathbf{x})) I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) d\mathbf{x},$$

and hence that

$$\int_{\mathbb{R}} T^2(i_1) p(i_1) di_1 = \frac{1}{|\Omega|^2} \int_{\mathbb{R}} \frac{1}{p(i_1)} \left(\int_{\Omega} g_\beta(i_1 - I_1^\sigma(\mathbf{x})) I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) d\mathbf{x} \right)^2 di_1.$$

We next write

$$\begin{aligned} & \left(\int_{\Omega} g_\beta(i_1 - I_1^\sigma(\mathbf{x})) I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) d\mathbf{x} \right)^2 = \\ & \int_{\Omega \times \Omega} g_\beta(i_1 - I_1^\sigma(\mathbf{x})) I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) g_\beta(i_1 - I_1^\sigma(\mathbf{x}')) I_2^\sigma(\mathbf{x}' + \mathbf{h}(\mathbf{x}')) d\mathbf{x} d\mathbf{x}', \end{aligned}$$

commute the integration with respect to i_1 with that with respect to \mathbf{x} and \mathbf{x}' to obtain the result. \square

We pursue with another lemma.

Lemma 6.17 *The function $M : H \longrightarrow \mathbb{R}$ defined in lemma 6.16 is bounded and Lipschitz continuous.*

Proof : For the first part, $|M(\mathbf{h})| \leq \mathcal{A}^2 \int_{\Omega \times \Omega} f(\mathbf{x}, \mathbf{x}') d\mathbf{x} d\mathbf{x}' = \mathcal{A}^2$, according to lemma 6.16. For the second part we compute

$$\begin{aligned} M[\mathbf{h}_1] - M[\mathbf{h}_2] = & \int_{\Omega \times \Omega} f(\mathbf{x}, \mathbf{x}') \left(I_2^\sigma(\mathbf{x} + \mathbf{h}_1(\mathbf{x})) I_2^\sigma(\mathbf{x}' + \mathbf{h}_1(\mathbf{x}')) - \right. \\ & \left. I_2^\sigma(\mathbf{x} + \mathbf{h}_2(\mathbf{x})) I_2^\sigma(\mathbf{x}' + \mathbf{h}_2(\mathbf{x}')) \right) d\mathbf{x} d\mathbf{x}'. \end{aligned}$$

$$|M[\mathbf{h}_1] - M[\mathbf{h}_2]| \leq \int_{\Omega \times \Omega} f(\mathbf{x}, \mathbf{x}') (|I_2^\sigma(\mathbf{x} + \mathbf{h}_1(\mathbf{x})) - I_2^\sigma(\mathbf{x} + \mathbf{h}_2(\mathbf{x}))| I_2^\sigma(\mathbf{x}' + \mathbf{h}_1(\mathbf{x}')) + |I_2^\sigma(\mathbf{x}' + \mathbf{h}_1(\mathbf{x}')) - I_2^\sigma(\mathbf{x}' + \mathbf{h}_2(\mathbf{x}'))| I_2^\sigma(\mathbf{x} + \mathbf{h}_2(\mathbf{x}))) d\mathbf{x} d\mathbf{x}'.$$

Because I_2^σ is Lipschitz continuous and bounded

$$\begin{aligned} |M[\mathbf{h}_1] - M[\mathbf{h}_2]| &\leq \|I_2^\sigma\|_\infty \text{Lip}(I_2^\sigma) \int_{\Omega \times \Omega} f(\mathbf{x}, \mathbf{x}') (|\mathbf{h}_1(\mathbf{x}) - \mathbf{h}_2(\mathbf{x})| + |\mathbf{h}_1(\mathbf{x}') - \mathbf{h}_2(\mathbf{x}')|) d\mathbf{x} d\mathbf{x}' = \\ &= \frac{\|I_2^\sigma\|_\infty \text{Lip}(I_2^\sigma)}{|\Omega|} \left(\int_{\Omega} |\mathbf{h}_1(\mathbf{x}) - \mathbf{h}_2(\mathbf{x})| d\mathbf{x} + \int_{\Omega} |\mathbf{h}_1(\mathbf{x}') - \mathbf{h}_2(\mathbf{x}')| d\mathbf{x}' \right) = \\ &= \frac{2\|I_2^\sigma\|_\infty \text{Lip}(I_2^\sigma)}{|\Omega|} \int_{\Omega} |\mathbf{h}_1(\mathbf{x}) - \mathbf{h}_2(\mathbf{x})| d\mathbf{x}. \end{aligned}$$

The previous to the last equality is obtained from lemma 6.16. Therefore we have from Cauchy-Schwarz:

$$|M[\mathbf{h}_1] - M[\mathbf{h}_2]| \leq \frac{2\|I_2^\sigma\|_\infty \text{Lip}(I_2^\sigma)}{|\Omega|^{1/2}} \|\mathbf{h}_1 - \mathbf{h}_2\|_H.$$

□

Lemma 6.18 *The functions $H \rightarrow \mathbb{R}^+$ defined by*

$$\mathbf{h} \rightarrow \frac{1}{|\Omega|} \int_{\Omega} I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) d\mathbf{x} \quad \text{and} \quad \mathbf{h} \rightarrow \frac{1}{|\Omega|} \int_{\Omega} I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x}))^2 d\mathbf{x}$$

are bounded and Lipschitz continuous.

Proof : Boundedness has been proved in lemma 6.15 for the first function. For the second we have

$$\frac{1}{|\Omega|} \int_{\Omega} I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x}))^2 d\mathbf{x} \leq \mathcal{A}^2.$$

Next, for Lipschitz continuity:

$$\begin{aligned} \left| \frac{1}{|\Omega|} \int_{\Omega} I_2^\sigma(\mathbf{x} + \mathbf{h}_1(\mathbf{x})) d\mathbf{x} - \frac{1}{|\Omega|} \int_{\Omega} I_2^\sigma(\mathbf{x} + \mathbf{h}_2(\mathbf{x})) d\mathbf{x} \right| &\leq \\ &= \text{Lip}(I_2^\sigma) \frac{1}{|\Omega|} \int_{\Omega} |\mathbf{h}_1(\mathbf{x}) - \mathbf{h}_2(\mathbf{x})| d\mathbf{x}, \end{aligned}$$

because I_2^σ is Lipschitz, and hence (Cauchy-Schwarz):

$$\left| \frac{1}{|\Omega|} \int_{\Omega} I_2^\sigma(\mathbf{x} + \mathbf{h}_1(\mathbf{x})) d\mathbf{x} - \frac{1}{|\Omega|} \int_{\Omega} I_2^\sigma(\mathbf{x} + \mathbf{h}_2(\mathbf{x})) d\mathbf{x} \right| \leq \frac{\text{Lip}(I_2^\sigma)}{|\Omega|^{1/2}} \|\mathbf{h}_1 - \mathbf{h}_2\|_H.$$

Similarly (Cauchy-Schwarz),

$$\left| \frac{1}{|\Omega|} \int_{\Omega} I_2^{\sigma}(\mathbf{x} + \mathbf{h}_1(\mathbf{x}))^2 d\mathbf{x} - \frac{1}{|\Omega|} \int_{\Omega} I_2^{\sigma}(\mathbf{x} + \mathbf{h}_2(\mathbf{x}))^2 d\mathbf{x} \right| \leq 2 \frac{\mathcal{ALip}(I_2^{\sigma})}{|\Omega|^{1/2}} \|\mathbf{h}_1 - \mathbf{h}_2\|_H.$$

□

From this lemma, proposition 6.1 and lemma 6.15 we deduce the following.

Corollary 6.19 *The function $H \rightarrow \mathbb{R}$ defined by $\mathbf{h} \rightarrow \mathbf{Var}[X_{I_2^{\sigma}, \mathbf{h}}]$ is Lipschitz continuous.*

We also prove the following.

Lemma 6.20 *The function $H \rightarrow \mathbb{R}$ defined by $\mathbf{h} \rightarrow \mathbf{E}[\mathbf{Var}[X_{I_2^{\sigma}, \mathbf{h}} | X_{I_1^{\sigma}}^g]]$ is bounded and Lipschitz continuous.*

Proof : This follows from lemmas 6.16, 6.17, and 6.18. □

We can now prove the important intermediary result that the correlation ratio, as a function of the field \mathbf{h} , is Lipschitz continuous.

Theorem 6.21 *The function $H \rightarrow \mathbb{R}$ defined by $\mathbf{h} \rightarrow \frac{\mathbf{E}[\mathbf{Var}[X_{I_2^{\sigma}, \mathbf{h}} | X_{I_1^{\sigma}}^g]]}{\mathbf{Var}[X_{I_2^{\sigma}, \mathbf{h}}]}$ is Lipschitz continuous.*

Proof : This follows from proposition 6.1 and from lemmas 6.20, 6.15 and corollary 6.19. □

We pursue with another lemma.

Lemma 6.22 *The function $f_{CR}^g = G_{\beta} \star L_{CR}^g : \mathcal{H}_2 \rightarrow \mathbb{R}$, where L_{CR}^g is given by equation (5.7), is equal to the following expression:*

$$f_{CR}^g(z_1, z_2, \mathbf{h}) = \frac{2}{|\Omega| v_2(\mathbf{h})} (\mu_2(\mathbf{h}) - d(z_1, \mathbf{h}) + \mathbf{CR}(\mathbf{h})(z_2 - \mu_2(\mathbf{h}))),$$

where

$$d(z_1, \mathbf{h}) = \int_{\mathbb{R}} g_{\beta}(z_1 - i_1) \left(\frac{\int_{\Omega} g_{\beta}(i_1 - I_1^{\sigma}(\mathbf{x})) I_2^{\sigma}(\mathbf{x} + \mathbf{h}(\mathbf{x})) d\mathbf{x}}{\int_{\Omega} g_{\beta}(i_1 - I_1^{\sigma}(\mathbf{x})) d\mathbf{x}} \right) di_1.$$

Proof : We use equation (5.7) and apply the convolution to it. The value of d is obtained from:

$$\begin{aligned} d(z_1, \mathbf{h}) &= \int_{\mathbb{R}} g_{\beta}(z_1 - i_1) \mu_{2|1}(i_1, \mathbf{h}) di_1 = \\ &= \int_{\mathbb{R}} \frac{g_{\beta}(z_1 - i_1)}{p(i_1)} \left(\int_{\mathbb{R}} i_2 P_{\mathbf{h}}(\mathbf{i}) di_2 \right) di_1 = \\ &= \int_{\mathbb{R}} \frac{g_{\beta}(z_1 - i_1)}{p(i_1)} \left(\int_{\Omega} I_2^{\sigma}(\mathbf{x} + \mathbf{h}(\mathbf{x})) g_{\beta}(i_1 - I_1^{\sigma}(\mathbf{x})) d\mathbf{x} \right) di_1. \end{aligned}$$

□

We next prove the following.

Lemma 6.23 *The function $d : \mathcal{H}_1 \longrightarrow \mathbb{R}$ is bounded and Lipschitz continuous.*

Proof : The proof of the first part uses exactly the same ideas as those of the second part of proposition 6.7. For the second part, we first prove that the function $H \longrightarrow \mathbb{R}$, $\mathbf{h} \longrightarrow d(z_1, \mathbf{h})$ is Lipschitz continuous with a Lipschitz constant L_d that is independent of $z_1 \in [0, \mathcal{A}]$ and second prove that the function $[0, \mathcal{A}] \longrightarrow \mathbb{R}$, $z_1 \longrightarrow \frac{\partial d}{\partial z_1}$ is upperbounded independently of $\mathbf{h} \in H$. Indeed,

$$\begin{aligned} |d(z_1, \mathbf{h}_1) - d(z_1, \mathbf{h}_2)| &\leq \\ &\int_{\mathbb{R}} g_{\beta}(z_1 - i_1) \left(\frac{\int_{\Omega} g_{\beta}(i_1 - I_1^{\sigma}(\mathbf{x})) |I_2^{\sigma}(\mathbf{x} + \mathbf{h}_1(\mathbf{x})) - I_2^{\sigma}(\mathbf{x} + \mathbf{h}_2(\mathbf{x}))| d\mathbf{x}}{\int_{\Omega} g_{\beta}(i_1 - I_1^{\sigma}(\mathbf{x})) d\mathbf{x}} \right) di_1. \end{aligned}$$

Because I_2^{σ} is Lipschitz continuous and of Schwarz inequality, we have

$$\begin{aligned} |d(z_1, \mathbf{h}_1) - d(z_1, \mathbf{h}_2)| &\leq \\ &|\Omega| Lip(I_2^{\sigma}) \|\mathbf{h}_1 - \mathbf{h}_2\|_H \int_{\mathbb{R}} g_{\beta}(z_1 - i_1) \left(\frac{(\int_{\Omega} g_{\beta}(i_1 - I_1^{\sigma}(\mathbf{x}))^2 d\mathbf{x})^{\frac{1}{2}}}{\int_{\Omega} g_{\beta}(i_1 - I_1^{\sigma}(\mathbf{x})) d\mathbf{x}} \right) di_1. \end{aligned}$$

The function of z_1 that appears on the right-hand side of this inequality does not depend on \mathbf{h} , is continuous and therefore bounded on $[0, \mathcal{A}]$.

We now notice that

$$\frac{\partial d}{\partial z_1} = \frac{1}{\beta} \int_{\mathbb{R}} (i_1 - z_1) g_{\beta}(z_1 - i_1) \left(\frac{\int_{\Omega} g_{\beta}(i_1 - I_1^{\sigma}(\mathbf{x})) I_2^{\sigma}(\mathbf{x} + \mathbf{h}(\mathbf{x})) d\mathbf{x}}{\int_{\Omega} g_{\beta}(i_1 - I_1^{\sigma}(\mathbf{x})) d\mathbf{x}} \right) di_1,$$

and, since

$$\begin{aligned} \frac{\int_{\Omega} g_{\beta}(i_1 - I_1^{\sigma}(\mathbf{x})) I_2^{\sigma}(\mathbf{x} + \mathbf{h}(\mathbf{x})) d\mathbf{x}}{\int_{\Omega} g_{\beta}(i_1 - I_1^{\sigma}(\mathbf{x})) d\mathbf{x}} &\leq \mathcal{A}, \\ \left| \frac{\partial d}{\partial z_1} \right| &\leq \frac{\mathcal{A}}{\beta} \int_{\mathbb{R}} |z_1 - i_1| g_{\beta}(z_1 - i_1) di_1. \end{aligned}$$

The right-hand side of this inequality is equal to $2\frac{A}{\beta} \int_0^{+\infty} z g_\beta(z) dz$, from where the conclusion follows. \square

This allows us to prove the following result.

Theorem 6.24 *The function $f_{CR}^g : \mathcal{H}_2 \rightarrow \mathbb{R}$ is Lipschitz continuous and bounded.*

Proof : The denominator $|\Omega|v_2(\mathbf{h})$ is > 0 and bounded (lemma 6.15), and Lipschitz continuous on H and K (corollary 6.19).

The numerator is bounded because $\mathbf{CR}(\mathbf{h})$ is bounded by 1, d is bounded (lemma 6.23) and $\mu_2(\mathbf{h})$ is bounded (lemma 6.15).

The product $\mu_2(\mathbf{h})\mathbf{CR}(\mathbf{h})$ is Lipschitz continuous on H as the product of two bounded Lipschitz continuous functions (proposition 6.1 and theorem 6.21). Hence we have proved the boundedness of f_{CR}^g .

The function $d : \mathcal{H}_1 \rightarrow \mathbb{R}$ is Lipschitz continuous (lemma 6.23).

The function $r : \mathcal{H}_1 \rightarrow \mathbb{R}$ defined by $(z_2, \mathbf{h}) \rightarrow z_2\mathbf{CR}(\mathbf{h})$ is Lipschitz continuous because of theorem 6.21 and

$$\begin{aligned} |z_2\mathbf{CR}(\mathbf{h}) - z'_2\mathbf{CR}(\mathbf{h}')| &= \\ |z_2(\mathbf{CR}(\mathbf{h}) - \mathbf{CR}(\mathbf{h}')) + \mathbf{CR}(\mathbf{h}')(z_2 - z'_2)| &\leq \\ \mathcal{A}|\mathbf{CR}(\mathbf{h}) - \mathbf{CR}(\mathbf{h}')| + |z_2 - z'_2|. & \end{aligned}$$

Hence the numerator is also Lipschitz continuous and, from proposition 6.1, so is f_{CR}^g . \square

We finally obtain the main result of this section.

Theorem 6.25 *The function $F_{CR}^g : H \rightarrow H$ defined by*

$$F_{CR}^g(\mathbf{h}) = f_{CR}^g(I_1^\sigma, I_2^\sigma(\mathbf{Id} + \mathbf{h}))\nabla I_2^\sigma(\mathbf{Id} + \mathbf{h})$$

is Lipschitz continuous and bounded.

Proof : The boundedness follows from theorem 6.24 and the fact that $|\nabla I_2^\sigma|$ is bounded. It implies that $F_{CR}^g(\mathbf{h}) \in H \forall \mathbf{h} \in H$. The rest of the proof follows exactly the same pattern as the proof of theorem 6.13 and uses theorem 6.24. \square

We finish this section with the following proposition.

Proposition 6.26 *The function $\Omega \rightarrow \mathbb{R}^n$ such that $\mathbf{x} \rightarrow F_{CR}^g(\mathbf{h}(\mathbf{x}))$ satisfies*

$$|F_{CR}^g(\mathbf{h}(\mathbf{x})) - F_{CR}^g(\mathbf{h}(\mathbf{y}))| \leq K(|\mathbf{x} - \mathbf{y}| + |\mathbf{h}(\mathbf{x}) - \mathbf{h}(\mathbf{y})|),$$

for some constant $K > 0$.

Proof : The proof is similar to that of proposition 6.14 and follows from theorem 6.24 and the fact that the functions I_1^σ , I_2^σ and all its derivatives, are Lipschitz continuous. \square

6.2.3 Cross Correlation

The goal of this section is to prove the Lipschitz-continuity of the function $F_{CC}^g(\mathbf{h})$ (equation (5.9)).

Lemma 6.27 *The following inequalities are verified for μ_1 and v_1 .*

$$\begin{aligned} 0 &\leq \mu_1 \leq \mathcal{A} \\ \beta &\leq v_1 \leq \beta + \mathcal{A}^2. \end{aligned}$$

Proof : We have

$$\mu_1 = \int_{\mathbb{R}^2} i_1 \left(\frac{1}{|\Omega|} \int_{\Omega} G_\beta(\mathbf{I}_h(\mathbf{x}) - \mathbf{i}) d\mathbf{x} \right) di_1 di_2 = \frac{1}{|\Omega|} \int_{\Omega} I_1^\sigma(\mathbf{x}) d\mathbf{x}, \quad (6.12)$$

and the first inequality follows from the fact that $I_1^\sigma(\mathbf{x}) \in [0, \mathcal{A}]$. Similarly, for v_1 we have

$$\begin{aligned} v_1 &= \int_{\mathbb{R}^2} i_1^2 \left(\frac{1}{|\Omega|} \int_{\Omega} G_\beta(\mathbf{I}_h(\mathbf{x}) - \mathbf{i}) d\mathbf{x} \right) di_1 di_2 - \mu_1^2 = \\ &\quad \beta + \frac{1}{|\Omega|} \int_{\Omega} I_1^\sigma(\mathbf{x})^2 d\mathbf{x} - \left(\frac{1}{|\Omega|} \int_{\Omega} I_1^\sigma(\mathbf{x}) d\mathbf{x} \right)^2, \quad (6.13) \end{aligned}$$

from which the second inequality follows. \square

Proposition 6.28 *The function $H \rightarrow \mathbb{R}$ defined by $\mathbf{h} \mapsto v_{1,2}(\mathbf{h})$ is bounded and Lipschitz continuous.*

Proof : We have

$$\begin{aligned} v_{1,2}(\mathbf{h}) &= \int_{\mathbb{R}^2} (i_1 - \mu_1)(i_2 - \mu_2(\mathbf{h})) \frac{1}{|\Omega|} \int_{\Omega} G_\beta(\mathbf{I}_h(\mathbf{x}) - \mathbf{i}) d\mathbf{x} di_1 di_2 \\ &= \frac{1}{|\Omega|} \int_{\Omega} \left(\int_{\mathbb{R}} (i_1 - \mu_1) g_\beta(I_1^\sigma(\mathbf{x}) - i_1) di_1 \right. \\ &\quad \left. \int_{\mathbb{R}} (i_2 - \mu_2(\mathbf{h})) g_\beta(I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) - i_2) di_2 \right) d\mathbf{x} \\ &= \frac{1}{|\Omega|} \int_{\Omega} (I_1^\sigma(\mathbf{x}) - \mu_1)(I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) - \mu_2(\mathbf{h})) d\mathbf{x}, \end{aligned}$$

that is

$$v_{1,2}(\mathbf{h}) = \frac{1}{|\Omega|} \int_{\Omega} I_1^\sigma(\mathbf{x}) I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) d\mathbf{x} - \mu_1 \mu_2(\mathbf{h}). \quad (6.14)$$

Hence, we have $|v_{1,2}(\mathbf{h})| \leq \mathcal{A}^2$, which proves the first part of the proposition. For the second part, since $\mu_2(\mathbf{h})$ is Lipschitz continuous (lemma 6.18), it suffices to show the Lipschitz continuity of the first term in the right-hand side. For this term, we have

$$\begin{aligned} & \left| \frac{1}{|\Omega|} \left| \int_{\Omega} I_1^\sigma(\mathbf{x}) I_2^\sigma(\mathbf{x} + \mathbf{h}_1(\mathbf{x})) d\mathbf{x} - \int_{\Omega} I_1^\sigma(\mathbf{x}) I_2^\sigma(\mathbf{x} + \mathbf{h}_2(\mathbf{x})) d\mathbf{x} \right| \right| \\ & \leq \frac{1}{|\Omega|} \int_{\Omega} |I_1^\sigma(\mathbf{x})| |I_2^\sigma(\mathbf{x} + \mathbf{h}_1(\mathbf{x})) - I_2^\sigma(\mathbf{x} + \mathbf{h}_2(\mathbf{x}))| d\mathbf{x} \\ & \leq \frac{\mathcal{A} \text{Lip}(I_2^\sigma)}{|\Omega|} \int_{\Omega} |\mathbf{h}_1(\mathbf{x}) - \mathbf{h}_2(\mathbf{x})| d\mathbf{x} \end{aligned}$$

Hence (Cauchy-Schwarz)

$$\begin{aligned} & \left| \frac{1}{|\Omega|} \left| \int_{\Omega} I_1^\sigma(\mathbf{x}) I_2^\sigma(\mathbf{x} + \mathbf{h}_1(\mathbf{x})) d\mathbf{x} - \int_{\Omega} I_1^\sigma(\mathbf{x}) I_2^\sigma(\mathbf{x} + \mathbf{h}_2(\mathbf{x})) d\mathbf{x} \right| \right| \\ & \leq \frac{\mathcal{A} \text{Lip}(I_2^\sigma)}{|\Omega|^{1/2}} \|\mathbf{h}_1 - \mathbf{h}_2\|_H. \end{aligned}$$

□

Theorem 6.29 *The function $H \rightarrow \mathbb{R}$ defined by $\mathbf{h} \mapsto \mathbf{CC}^g(\mathbf{h})$ is bounded and Lipschitz continuous.*

Proof : $\mathbf{CC}^g(\mathbf{h})$ is bounded by 1. The mapping $\mathbf{h} \mapsto \mathbf{CC}^g(\mathbf{h})$ is Lipschitz because we have

$$\mathbf{CC}^g(\mathbf{h}) = \frac{v_{1,2}(\mathbf{h})^2}{v_1 v_2(\mathbf{h})}, \quad (6.15)$$

with $v_1 \geq \beta > 0$ (proposition 6.27), $v_{1,2}(\mathbf{h})$ bounded and Lipschitz (proposition 6.28), $v_2(\mathbf{h})$ Lipschitz (corollary 6.19) and $v_2(\mathbf{h}) \geq \beta$ (lemma 6.15), which allows us to apply proposition 6.1. □

Theorem 6.30 *The function $f_{CC}^g : \mathcal{H}_2 \rightarrow \mathbb{R}$ defined by*

$$(z_1, z_2, \mathbf{h}) \mapsto \frac{-2}{|\Omega|} \left[\frac{v_{1,2}(\mathbf{h})}{v_2(\mathbf{h})} \left(\frac{z_1 - \mu_1}{v_1} \right) - \mathbf{CC}^g(\mathbf{h}) \left(\frac{z_2 - \mu_2(\mathbf{h})}{v_2(\mathbf{h})} \right) \right].$$

is bounded and Lipschitz continuous.

Proof : Taking into account the properties mentioned in the proof of the previous proposition, plus the boundedness and Lipschitz continuity of $\mathbf{CC}^g(\mathbf{h})$ (proposition 6.29) and of μ_2 (lemma 6.18), we see that $f_{\mathbf{CC}}^g$ may be written as

$$f_{\mathbf{CC}}^g(z_1, z_2, \mathbf{h}) = f_1(\mathbf{h}) z_1 + f_2(\mathbf{h}) z_2 + f_3(\mathbf{h}),$$

where the functions $H \rightarrow \mathbb{R}$ f_1, f_2 and f_3 are Lipschitz continuous and bounded (proposition 6.1). The result readily follows since

$$\begin{aligned} |z_i f_i(\mathbf{h}) - z'_i f_i(\mathbf{h}')| &= |z_i (f_i(\mathbf{h}) - f_i(\mathbf{h}')) + f_i(\mathbf{h}') (z_i - z'_i)| \\ &\leq \mathcal{A} |f_i(\mathbf{h}) - f_i(\mathbf{h}')| + \|f_i\|_\infty |z_i - z'_i| \\ &\leq \mathcal{A} \text{Lip}(f_i) \|\mathbf{h} - \mathbf{h}'\|_H + \|f_i\|_\infty |z_i - z'_i| \quad i = 1, 2. \end{aligned}$$

□

Theorem 6.31 *The function $F_{\mathbf{CC}}^g : H \rightarrow H$, defined by*

$$F_{\mathbf{CC}}^g(\mathbf{h}) = f_{\mathbf{CC}}^g(I_1^\sigma, I_2^\sigma(\mathbf{Id} + \mathbf{h})) \nabla I_2^\sigma(\mathbf{Id} + \mathbf{h})$$

is Lipschitz continuous and bounded.

Proof : The boundedness follows from theorem 6.30 and the fact that $|\nabla F_2^g|$ is bounded. It implies that $F_{\mathbf{CC}}^g(\mathbf{h}) \in H \forall \mathbf{h} \in H$. The rest of the proof follows exactly the same pattern as the proof of theorem 6.13 and uses theorem 6.30. □

We finish this section with the following proposition.

Proposition 6.32 *The function $\Omega \rightarrow \mathbb{R}^n$ such that $\mathbf{x} \rightarrow F_{\mathbf{CC}}^g(\mathbf{h}(\mathbf{x}))$ satisfies*

$$|F_{\mathbf{CC}}^g(\mathbf{h}(\mathbf{x})) - F_{\mathbf{CC}}^g(\mathbf{h}(\mathbf{y}))| \leq K(|\mathbf{x} - \mathbf{y}| + |\mathbf{h}(\mathbf{x}) - \mathbf{h}(\mathbf{y})|),$$

for some constant $K > 0$.

Proof : The proof is similar to that of proposition 6.14 and follows from theorem 6.30 and the fact that the functions I_1^σ, I_2^σ and all its derivatives, are Lipschitz continuous.

□

6.3 Local Criteria

We now study the Lipschitz continuity of the gradients of the local criteria, as defined in (5.14).

The analysis of the local criteria follows pretty much directly from the analysis of the global ones and from theorem 5.2. The main difference with the global case is that we have an extra spatial dependency. In the next lemma we introduce a constant that is needed in the sequel.

Lemma 6.33 *Let $\text{diam}(\Omega)$ be the diameter of the open bounded set Ω :*

$$\text{diam}(\Omega) = \sup_{\mathbf{x}, \mathbf{y} \in \Omega} \|\mathbf{x} - \mathbf{y}\|.$$

We note $G_\gamma(\text{diam}(\Omega))$ the value $\inf_{\mathbf{x}, \mathbf{y} \in \Omega} G_\gamma(\mathbf{x} - \mathbf{y})$ and define

$$K_\Omega = \frac{G_\gamma(\mathbf{0})}{G_\gamma(\text{diam}(\Omega))}. \quad (6.16)$$

We say that

$$\int_{\Omega} G_\gamma(\mathbf{x} - \mathbf{x}_0) \frac{1}{\mathcal{G}_\gamma(\mathbf{x}_0)} d\mathbf{x}_0 \leq K_\Omega \quad \forall \mathbf{x} \in \Omega.$$

Proof : Since $\mathcal{G}_\gamma(\mathbf{x}_0) = \int_{\Omega} G_\gamma(\mathbf{y} - \mathbf{x}_0) d\mathbf{y}$, we have $\mathcal{G}_\gamma(\mathbf{x}_0) \geq |\Omega|G_\gamma(\text{diam}(\Omega))$. Therefore

$$\begin{aligned} \int_{\Omega} G_\gamma(\mathbf{x} - \mathbf{x}_0) \frac{1}{\mathcal{G}_\gamma(\mathbf{x}_0)} d\mathbf{x}_0 &\leq \frac{1}{|\Omega|G_\gamma(\text{diam}(\Omega))} \int_{\Omega} G_\gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x}_0 \leq \\ &\frac{1}{|\Omega|G_\gamma(\text{diam}(\Omega))} \times |\Omega|G_\gamma(\mathbf{0}) = K_\Omega. \end{aligned}$$

□

We note $k_\Omega \equiv |\Omega|G_\gamma(\text{diam}(\Omega))$, i.e. the constant such that $\forall \mathbf{x} \in \Omega, G_\gamma(\mathbf{x}) \geq k_\Omega$.

6.3.1 Mutual Information

The functions $q_{\mathbf{h}}$ and $Q_{\mathbf{h}}$ defined in propositions 6.4 and 6.5 are now functions of \mathbf{x}_0 but the propositions are unchanged. The function $a(i_2, \mathbf{x}_0, \mathbf{h})$, the local version of equation (6.1), is given by

$$a(i_2, \mathbf{x}_0, \mathbf{h}) = \frac{1}{\beta} \frac{\int_{\Omega} I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) g_\beta(I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) - i_2) G_\gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x}}{\int_{\Omega} g_\beta(I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) - i_2) G_\gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x}}, \quad (6.17)$$

while the function $A(\mathbf{i}, \mathbf{h}, \mathbf{x}_0)$, the local version of equation (6.2), is given by

$$A(\mathbf{i}, \mathbf{h}, \mathbf{x}_0) = \frac{1}{\beta} \frac{\int_{\Omega} I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) G_\beta(\mathbf{I}_{\mathbf{h}}(\mathbf{x}) - \mathbf{i}) G_\gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x}}{\int_{\Omega} G_\beta(\mathbf{I}_{\mathbf{h}}(\mathbf{x}) - \mathbf{i}) G_\gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x}}. \quad (6.18)$$

Similarly, the function $L_{\text{MI}, \mathbf{h}}^g(\mathbf{i})$ defined by equation (6.3) must be modified as follows:

$$L_{\text{MI}, \mathbf{h}}^l(\mathbf{i}, \mathbf{x}_0) = -\frac{1}{\mathcal{G}_\gamma(\mathbf{x}_0)} (A(\mathbf{i}, \mathbf{x}_0, \mathbf{h}) - a(i_2, \mathbf{x}_0, \mathbf{h})), \quad (6.19)$$

as well as the function b of equation (6.4):

$$b(z_2, \mathbf{h}, \mathbf{x}) = (G_\gamma \star g_\beta \star \frac{a}{\mathcal{G}_\gamma})(z_2, \mathbf{h}, \mathbf{x}) = \int_{\Omega} \int_{\mathbb{R}} G_\gamma(\mathbf{x} - \mathbf{x}_0) g_\beta(z_2 - i_2) \frac{1}{\mathcal{G}_\gamma(\mathbf{x}_0)} a(i_2, \mathbf{x}_0, \mathbf{h}) di_2 d\mathbf{x}_0, \quad (6.20)$$

and the function B of equation (6.5):

$$B(\mathbf{z}, \mathbf{h}, \mathbf{x}) = (G_\gamma \star G_\beta \star \frac{A}{\mathcal{G}_\gamma})(\mathbf{z}, \mathbf{h}, \mathbf{x}) = \int_{\Omega} \int_{\mathbb{R}^2} G_\gamma(\mathbf{x} - \mathbf{x}_0) G_\beta(\mathbf{z} - \mathbf{i}) \frac{1}{\mathcal{G}_\gamma(\mathbf{x}_0)} A(\mathbf{i}, \mathbf{x}_0, \mathbf{h}) di d\mathbf{x}_0. \quad (6.21)$$

This being done, propositions 6.6 and 6.7 can be adapted to the present case as follows.

Proposition 6.34 *The function $\mathbb{R} \rightarrow \mathbb{R}^+$ defined by $z_2 \rightarrow b(z_2, \mathbf{h}, \mathbf{x})$ is Lipschitz continuous with a Lipschitz constant L_b^l which is independent of \mathbf{h} and \mathbf{x} . Moreover, it is bounded by $\frac{A}{\beta}$.*

Proof : We give the proof in this particular simple case to give the flavor of the ideas which extend to the more complicated cases that come later.

The second part of the proposition follows from the fact that $0 \leq a(i_2, \mathbf{h}, \mathbf{x}) \leq \frac{A}{\beta} \forall i_2 \in \mathbb{R}$ and $\forall \mathbf{h} \in H$ (local version of proposition 6.4) and lemma 6.33.

In order to prove the first part, we prove that the magnitude of the derivative of the function is bounded independently of \mathbf{h} and \mathbf{x} . Indeed

$$\left| \frac{\partial b(z_2, \mathbf{h}, \mathbf{x})}{\partial z_2} \right| = \frac{1}{\beta} \left| \int_{\Omega} \int_{\mathbb{R}} (i_2 - z_2) G_\gamma(\mathbf{x} - \mathbf{x}_0) \frac{1}{\mathcal{G}_\gamma(\mathbf{x}_0)} g_\beta(z_2 - i_2) a(i_2, \mathbf{x}_0, \mathbf{h}) di_2 d\mathbf{x}_0 \right| \leq \frac{AK_\Omega}{\beta} \int_{\mathbb{R}} |z_2 - i_2| g_\beta(z_2 - i_2) di_2.$$

In order to derive the last inequality we have used the local version of proposition 6.4 and lemma 6.33. The function on the right-hand side of the last inequality is independent of \mathbf{h} and \mathbf{x} and continuous on $[0, \mathcal{A}]$, therefore bounded. \square

Proposition 6.35 *The function $H \rightarrow \mathbb{R}$ defined by $\mathbf{h} \rightarrow b(z_2, \mathbf{h}, \mathbf{x})$ is Lipschitz continuous with a Lipschitz constant L_b^l which is independent of $z_2 \in \mathbb{R}$ and $\mathbf{x} \in \Omega$.*

Proof : The proof is similar to that of proposition 6.7. \square

We also have the following.

Proposition 6.36 *The function $\Omega \rightarrow \mathbb{R}$ defined by $\mathbf{x} \rightarrow b(z_2, \mathbf{h}, \mathbf{x})$ is Lipschitz continuous uniformly on \mathcal{H}_1 .*

Proof : Because of equation (6.20) and proposition 6.34 we have

$$|b(z_2, \mathbf{h}, \mathbf{x}) - b(z_2, \mathbf{y}, \mathbf{h})| \leq \frac{\mathcal{A}}{\beta} \int_{\Omega} \int_{\mathbb{R}} \frac{|G_{\gamma}(\mathbf{x} - \mathbf{x}_0) - G_{\gamma}(\mathbf{y} - \mathbf{x}_0)|}{G_{\gamma}(\mathbf{x}_0)} g_{\beta}(z_2 - i_2) di_2 d\mathbf{x}_0.$$

The proof of lemma 6.33 allows us to write

$$|b(z_2, \mathbf{h}, \mathbf{x}) - b(z_2, \mathbf{y}, \mathbf{h})| \leq \frac{\mathcal{A}}{\beta|\Omega|G_{\gamma}(\text{diam}(\Omega))} \int_{\Omega} \int_{\mathbb{R}} |G_{\gamma}(\mathbf{x} - \mathbf{x}_0) - G_{\gamma}(\mathbf{y} - \mathbf{x}_0)| g_{\beta}(z_2 - i_2) di_2 d\mathbf{x}_0,$$

and therefore

$$|b(z_2, \mathbf{h}, \mathbf{x}) - b(z_2, \mathbf{y}, \mathbf{h})| \leq \frac{\mathcal{A} \text{Lip}(G_{\gamma})}{\beta G_{\gamma}(\text{diam}(\Omega))} |\mathbf{x} - \mathbf{y}|.$$

□

As a consequence of lemma 6.3 we can state the following.

Proposition 6.37 *The function $b : \mathcal{H}_1 \times \Omega \rightarrow \mathbb{R}$ is Lipschitz continuous on \mathcal{H}_1 uniformly on Ω .*

Proof : The proof follows from lemma 6.3 and proposition 6.34 and 6.35. □

Similarly we have the

Proposition 6.38 *The function $\Omega \rightarrow \mathbb{R}$ defined by $\mathbf{x} \rightarrow B(z_1, z_2, \mathbf{h}, \mathbf{x})$ is Lipschitz continuous uniformly on \mathcal{H}_2 .*

Proof : The proof is similar to that of proposition 6.36. □

The following proposition is also needed.

Proposition 6.39 *The function $B : \mathcal{H}_2 \times \Omega \rightarrow \mathbb{R}$ is Lipschitz continuous on \mathcal{H}_2 uniformly on Ω . It is bounded by $\frac{\mathcal{A}}{\beta}$.*

Proof : The proof is similar to that of proposition 6.37. □

And therefore

Corollary 6.40 *The function $f_{MI}^l : \mathcal{H}_2 \times \Omega \longrightarrow \mathbb{R}$ defined by*

$$(z_1, z_2, \mathbf{h}, \mathbf{x}) \longrightarrow -(B(z_1, z_2, \mathbf{h}, \mathbf{x}) - b(z_2, \mathbf{h}, \mathbf{x}))$$

is Lipschitz continuous on \mathcal{H}_2 uniformly on Ω and bounded.

From this follows the local version of theorem 6.13

Theorem 6.41 *The function $F_{MI}^l : H \longrightarrow H$ defined by*

$$\begin{aligned} F_{MI}^l(\mathbf{h}) &= f_{MI}^l(I_1^\sigma, I_2^\sigma(\mathbf{Id} + \mathbf{h}), \mathbf{Id}, \mathbf{h}) \nabla I_2^\sigma(\mathbf{Id} + \mathbf{h}) = \\ &\quad - (B(I_1^\sigma, I_2^\sigma(\mathbf{Id} + \mathbf{h}), \mathbf{Id}, \mathbf{h}) - b(I_2^\sigma(\mathbf{Id} + \mathbf{h}), \mathbf{Id}, \mathbf{h})) \nabla I_2^\sigma(\mathbf{Id} + \mathbf{h}) \end{aligned}$$

is Lipschitz continuous and bounded.

Proof : Boundedness follows from corollary 6.40 and the fact that $|\nabla I_2^\sigma|$ is bounded. It implies that $F_{MI}^l(\mathbf{h}) \in H \forall \mathbf{h} \in H$.

We next consider the i th component $F_{MI}^{l,i}$ of F_{MI}^l :

$$F_{MI}^{l,i}(\mathbf{h}_1)(\mathbf{x}) - F_{MI}^{l,i}(\mathbf{h}_2)(\mathbf{x}) = S_1 T_1 - S_2 T_2,$$

with

$$\begin{aligned} S_j &= -(B(I_1^\sigma(\mathbf{x}), I_2^\sigma(\mathbf{x} + \mathbf{h}_j(\mathbf{x})), \mathbf{h}_j, \mathbf{x}) - b(I_2^\sigma(\mathbf{x} + \mathbf{h}_j(\mathbf{x})), \mathbf{h}_j, \mathbf{x})) \\ T_j &= \partial_i I_2^\sigma(\mathbf{x} + \mathbf{h}_j(\mathbf{x})) \quad j = 1, 2, \end{aligned}$$

$j = 1, 2$. We continue with

$$|F_{MI}^{l,i}(\mathbf{h}_1)(\mathbf{x}) - F_{MI}^{l,i}(\mathbf{h}_2)(\mathbf{x})| \leq |S_1 - S_2| |T_1| + |S_2| |T_1 - T_2|$$

Because $\partial_i I_2^\sigma$ is bounded, $|T_j| \leq \|\partial_i I_2^\sigma\|_\infty$. Because of propositions 6.34 and 6.39, $|S_2| \leq 2 \frac{A}{\beta}$. Because $\partial_i I_2^\sigma$ is Lipschitz continuous $|T_1 - T_2| \leq Lip(\partial_i I_2^\sigma) |\mathbf{h}_1(\mathbf{x}) - \mathbf{h}_2(\mathbf{x})|$. Finally, because of corollary 6.40 and the fact that I_2^σ is Lipschitz continuous,

$$|S_1 - S_2| \leq Lip(f_{MI}^l) (Lip(I_2^\sigma) |\mathbf{h}_1(\mathbf{x}) - \mathbf{h}_2(\mathbf{x})| + \|\mathbf{h}_1 - \mathbf{h}_2\|_H).$$

The conclusion of the theorem follows from these inequalities through the same procedures as in the proof of theorem 6.13. \square

We finish this section with the

Proposition 6.42 *The function $\Omega \rightarrow \mathbb{R}^n$ such that $\mathbf{x} \rightarrow F_{MI}^l(\mathbf{h}(\mathbf{x}))$ satisfies*

$$|F_{MI}^l(\mathbf{h}(\mathbf{x})) - F_{MI}^l(\mathbf{h}(\mathbf{y}))| \leq K (|\mathbf{x} - \mathbf{y}| + |\mathbf{h}(\mathbf{x}) - \mathbf{h}(\mathbf{y})|),$$

for some constant $K > 0$.

Proof : The proof is similar to that of proposition 6.14 and uses propositions 6.36 and 6.38. \square

6.3.2 Correlation Ratio

In this case also, the proofs follow pretty much the same pattern as those in the global case. In detail, the analog of lemma 6.15 is the

Lemma 6.43

$$0 \leq \mu_2(\mathbf{h}, \mathbf{x}_0) = \frac{1}{\mathcal{G}_\gamma(\mathbf{x}_0)} \int_{\Omega} I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) G_\gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x} \leq \mathcal{A} \quad (6.22)$$

$$\beta \leq v_2(\mathbf{h}, \mathbf{x}_0) = \beta + \theta(\mathbf{h}, \mathbf{x}_0) \leq \beta + \mathcal{A}^2, \quad (6.23)$$

where

$$\theta(\mathbf{h}, \mathbf{x}_0) = \frac{1}{\mathcal{G}_\gamma(\mathbf{x}_0)} \int_{\Omega} I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x}))^2 G_\gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x} - \left(\frac{1}{\mathcal{G}_\gamma(\mathbf{x}_0)} \int_{\Omega} I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) G_\gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x} \right)^2. \quad (6.24)$$

Proof : Because of equation (4.16) we have

$$\begin{aligned} \mu_2(\mathbf{h}, \mathbf{x}_0) &= \int_{\mathbb{R}} i_2 \left(\frac{1}{\mathcal{G}_\gamma(\mathbf{x}_0)} \int_{\Omega} g_\beta(i_2 - I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x}))) G_\gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x} \right) di_2 = \\ &= \frac{1}{\mathcal{G}_\gamma(\mathbf{x}_0)} \int_{\Omega} I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) G_\gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x}. \end{aligned}$$

This yields the first part of the lemma. For the second part, we use equation (4.17):

$$v_2(\mathbf{h}, \mathbf{x}_0) = \int_{\mathbb{R}} i_2^2 \left(\frac{1}{\mathcal{G}_\gamma(\mathbf{x}_0)} \int_{\Omega} g_\beta(i_2 - I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x}))) G_\gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x} \right) di_2 - \mu_2(\mathbf{x}_0, \mathbf{h})^2,$$

and hence

$$v_2(\mathbf{h}, \mathbf{x}_0) = \beta + \frac{1}{\mathcal{G}_\gamma(\mathbf{x}_0)} \int_{\Omega} I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x}))^2 G_\gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x} - \left(\frac{1}{\mathcal{G}_\gamma(\mathbf{x}_0)} \int_{\Omega} I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) G_\gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x} \right)^2,$$

from which the upper and lower bounds of the lemma follow. \square

We next take care of $\mathbf{E}[\mathbf{Var}[X_{I_2^\sigma, \mathbf{h}} | X_{I_1^\sigma}](\mathbf{x}_0)]$ with the analog of lemma 6.16

Lemma 6.44

$$\mathbf{E}[\mathbf{Var}[X_{I_2^\sigma, \mathbf{h}} | X_{I_1^\sigma}](\mathbf{x}_0)] = \beta + \frac{1}{\mathcal{G}_\gamma(\mathbf{x}_0)} \int_{\Omega} I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x}))^2 G_\gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x} - M[\mathbf{x}_0, \mathbf{h}],$$

where

$$M[\mathbf{x}_0, \mathbf{h}] = \int_{\Omega \times \Omega} f(\mathbf{x}_0, \mathbf{x}, \mathbf{x}') I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) I_2^\sigma(\mathbf{x}' + \mathbf{h}(\mathbf{x}')) d\mathbf{x} d\mathbf{x}',$$

and

$$f(\mathbf{x}_0, \mathbf{x}, \mathbf{x}') = \frac{G_\gamma(\mathbf{x} - \mathbf{x}_0) G_\gamma(\mathbf{x}' - \mathbf{x}_0)}{\mathcal{G}_\gamma(\mathbf{x}_0)^2} \int_{\mathbb{R}} \frac{g_\beta(i_1 - I_1^\sigma(\mathbf{x})) g_\beta(i_1 - I_1^\sigma(\mathbf{x}'))}{p(i_1, \mathbf{x}_0)} di_1,$$

is such that

$$\begin{aligned} \int_{\Omega} f(\mathbf{x}_0, \mathbf{x}, \mathbf{x}') d\mathbf{x} &= \frac{G_\gamma(\mathbf{x}' - \mathbf{x}_0)}{\mathcal{G}_\gamma(\mathbf{x}_0)} \\ \int_{\Omega} f(\mathbf{x}_0, \mathbf{x}, \mathbf{x}') d\mathbf{x}' &= \frac{G_\gamma(\mathbf{x} - \mathbf{x}_0)}{\mathcal{G}_\gamma(\mathbf{x}_0)}. \end{aligned}$$

and

$$\int_{\Omega \times \Omega} f(\mathbf{x}_0, \mathbf{x}, \mathbf{x}') d\mathbf{x} d\mathbf{x}' = 1.$$

Proof : According to Table 4.2 we have $v_{2|1}(i_1, \mathbf{h}, \mathbf{x}_0) = S(i_1, \mathbf{x}_0) - T^2(i_1, \mathbf{x}_0)$ and

$$\mathbf{E}[\mathbf{Var}[X_{I_2^\sigma, \mathbf{h}} | X_{I_1^\sigma}](\mathbf{x}_0)] = \int_{\mathbb{R}} v_{2|1}(i_1, \mathbf{h}, \mathbf{x}_0) p(i_1, \mathbf{x}_0) di_1,$$

with

$$S(i_1, \mathbf{x}_0) = \int_{\mathbb{R}} i_2^2 \frac{P_{\mathbf{h}}(i_1, i_2, \mathbf{x}_0)}{p(i_1, \mathbf{x}_0)} di_2,$$

and

$$T(i_1, \mathbf{x}_0) = \int_{\mathbb{R}} i_2 \frac{P_{\mathbf{h}}(i_1, i_2, \mathbf{x}_0)}{p(i_1, \mathbf{x}_0)} di_2 = \mu_{2|1}(i_1, \mathbf{h}, \mathbf{x}_0).$$

It is straightforward to show that

$$\int_{\mathbb{R}} S(i_1, \mathbf{x}_0) p(i_1, \mathbf{x}_0) di_1 = \beta + \frac{1}{\mathcal{G}_\gamma(\mathbf{x}_0)} \int_{\Omega} I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x}))^2 G_\gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x}.$$

It is also straightforward to show that

$$T(i_1, \mathbf{x}_0) = \frac{1}{\mathcal{G}_\gamma(\mathbf{x}_0) p(i_1, \mathbf{x}_0)} \int_{\Omega} g_\beta(i_1 - I_1^\sigma(\mathbf{x})) I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) G_\gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x},$$

and hence that

$$\begin{aligned} \int_{\mathbb{R}} T^2(i_1, \mathbf{x}_0) p(i_1, \mathbf{x}_0) di_1 &= \\ \frac{1}{\mathcal{G}_\gamma(\mathbf{x}_0)^2} \int_{\mathbb{R}} \frac{1}{p(i_1)} \left(\int_{\Omega} g_\beta(i_1 - I_1^\sigma(\mathbf{x})) I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) G_\gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x} \right)^2 di_1. \end{aligned} \quad (6.25)$$

We next write

$$\begin{aligned} \left(\int_{\Omega} g_{\beta}(i_1 - I_1^{\sigma}(\mathbf{x})) I_2^{\sigma}(\mathbf{x} + \mathbf{h}(\mathbf{x})) G_{\gamma}(\mathbf{x} - \mathbf{x}_0) d\mathbf{x} \right)^2 = \\ \int_{\Omega \times \Omega} g_{\beta}(i_1 - I_1^{\sigma}(\mathbf{x})) I_2^{\sigma}(\mathbf{x} + \mathbf{h}(\mathbf{x})) G_{\gamma}(\mathbf{x} - \mathbf{x}_0) \\ g_{\beta}(i_1 - I_1^{\sigma}(\mathbf{x}')) I_2^{\sigma}(\mathbf{x}' + \mathbf{h}(\mathbf{x}')) G_{\gamma}(\mathbf{x}' - \mathbf{x}_0) d\mathbf{x} d\mathbf{x}', \end{aligned}$$

commute the integration with respect to i_1 with that with respect to \mathbf{x} and \mathbf{x}' to obtain the result. \square

We continue with the analog of lemma 6.22:

Lemma 6.45 *The function $h_{CR}^l = G_{\beta} \star L_{CR}^l : \mathcal{H}_2 \times \Omega \rightarrow \mathbb{R}$, where L_{CR}^l is given by equation (5.12), is equal to the following expression*

$$h_{CR}^l(z_1, z_2, \mathbf{h}, \mathbf{x}_0) = \frac{2}{\mathcal{G}_{\gamma}(\mathbf{x}_0) v_2(\mathbf{h}, \mathbf{x}_0)} \left(\mu_2(\mathbf{h}, \mathbf{x}_0) - d(z_1, \mathbf{h}, \mathbf{x}_0) + \mathbf{CR}^l(\mathbf{h}, \mathbf{x}_0)(z_2 - \mu_2(\mathbf{h}, \mathbf{x}_0)) \right),$$

where

$$d(z_1, \mathbf{h}, \mathbf{x}_0) = \int_{\mathbb{R}} g_{\beta}(z_1 - i_1) \left(\frac{\int_{\Omega} g_{\beta}(i_1 - I_1^{\sigma}(\mathbf{x})) I_2^{\sigma}(\mathbf{x} + \mathbf{h}(\mathbf{x})) G_{\gamma}(\mathbf{x} - \mathbf{x}_0) d\mathbf{x}}{\int_{\Omega} g_{\beta}(i_1 - I_1^{\sigma}(\mathbf{x})) G_{\gamma}(\mathbf{x} - \mathbf{x}_0) d\mathbf{x}} \right) di_1.$$

Proof : We use equation (5.12) and apply the convolution to it. The value of d is obtained from:

$$\begin{aligned} d(z_1, \mathbf{h}, \mathbf{x}_0) &= \int_{\mathbb{R}} g_{\beta}(z_1 - i_1) \mu_{2|1}(i_1, \mathbf{h}, \mathbf{x}_0) di_1 = \\ &= \int_{\mathbb{R}} \frac{g_{\beta}(z_1 - i_1)}{p(i_1, \mathbf{x}_0)} \left(\int_{\mathbb{R}} i_2 P_{\mathbf{h}}(i_1, i_2, \mathbf{x}_0) di_2 \right) di_1 = \\ &= \int_{\mathbb{R}} \frac{g_{\beta}(z_1 - i_1)}{\mathcal{G}_{\gamma}(\mathbf{x}_0) p(i_1, \mathbf{x}_0)} \left(\int_{\Omega} I_2^{\sigma}(\mathbf{x} + \mathbf{h}(\mathbf{x})) g_{\beta}(i_1 - I_1^{\sigma}(\mathbf{x})) G_{\gamma}(\mathbf{x} - \mathbf{x}_0) d\mathbf{x} \right) di_1 \end{aligned}$$

from where the result follows. \square

Our goal is to prove that $f_{CR}^l = G_{\gamma} \star h_{CR}^l$, a function from $\mathcal{H}_2 \times \Omega$ in \mathbb{R} is Lipschitz continuous in \mathcal{H}_2 uniformly on Ω .

In order to prove this, it is sufficient to prove that the numerator and the denominator of h_{CR}^l are bounded and Lipschitz continuous in \mathcal{H}_2 uniformly on Ω , and that the denominator is strictly positive.

Indeed, we have the following

Lemma 6.46 *Let $N : \mathcal{H}_2 \times \Omega \rightarrow \mathbb{R}$ be bounded and Lipschitz continuous in \mathcal{H}_2 uniformly on Ω . Let also $D : \mathcal{H}_2 \times \Omega \rightarrow \mathbb{R}^+$ be bounded, strictly positive, and Lipschitz continuous in \mathcal{H}_2 uniformly on Ω . Then the function $G_\gamma \star \frac{N}{D} : \mathcal{H}_2 \times \Omega \rightarrow \mathbb{R}$ is Lipschitz continuous in \mathcal{H}_2 uniformly on Ω .*

Proof : We form

$$\begin{aligned} \left| \int_{\Omega} G_\gamma(\mathbf{x} - \mathbf{x}_0) \left(\frac{N(\mathbf{x}_0)[\mathbf{z}', \mathbf{h}']}{D(\mathbf{x}_0)[\mathbf{z}', \mathbf{h}']} - \frac{N(\mathbf{x}_0)[\mathbf{z}, \mathbf{h}]}{D(\mathbf{x}_0)[\mathbf{z}, \mathbf{h}]} \right) d\mathbf{x}_0 \right| \leq \\ G_\gamma(\mathbf{0}) \int_{\Omega} \frac{|N(\mathbf{x}_0)[\mathbf{z}', \mathbf{h}'] - N(\mathbf{x}_0)[\mathbf{z}, \mathbf{h}]|}{D(\mathbf{x}_0)[\mathbf{z}', \mathbf{h}']} d\mathbf{x}_0 + \\ G_\gamma(\mathbf{0}) \int_{\Omega} \frac{|N(\mathbf{x}_0)[\mathbf{z}, \mathbf{h}]| |D(\mathbf{x}_0)[\mathbf{z}', \mathbf{h}'] - D(\mathbf{x}_0)[\mathbf{z}, \mathbf{h}]|}{D(\mathbf{x}_0)[\mathbf{z}', \mathbf{h}'] D(\mathbf{x}_0)[\mathbf{z}, \mathbf{h}]} d\mathbf{x}_0 \end{aligned}$$

According to the hypotheses, there exists $a > 0$ such that $a \leq D(\mathbf{x}_0)[\mathbf{z}, \mathbf{h}] \forall \mathbf{z}, \mathbf{x}_0, \mathbf{h}$, there exists $K_N \geq 0$ such that $|N(\mathbf{x}_0)[\mathbf{z}, \mathbf{h}]| \leq K_N \forall \mathbf{z}, \mathbf{x}_0, \mathbf{h}$, and there exists L_N and L_D such that

$$\begin{aligned} |N(\mathbf{x}_0)[\mathbf{z}', \mathbf{h}'] - N(\mathbf{x}_0)[\mathbf{z}, \mathbf{h}]| &\leq L_N (|\mathbf{z} - \mathbf{z}'| + \|\mathbf{h} - \mathbf{h}'\|_H) \quad \text{and} \\ |D(\mathbf{x}_0)[\mathbf{z}', \mathbf{h}'] - D(\mathbf{x}_0)[\mathbf{z}, \mathbf{h}]| &\leq L_D (|\mathbf{z} - \mathbf{z}'| + \|\mathbf{h} - \mathbf{h}'\|_H) \quad \forall \mathbf{z}, \mathbf{z}', \mathbf{x}_0, \mathbf{h}, \mathbf{h}'. \end{aligned}$$

We therefore have through Cauchy-Schwarz inequality:

$$\left| \int_{\Omega} G_\gamma(\mathbf{x} - \mathbf{x}_0) \left(\frac{N(\mathbf{x}_0)[\mathbf{z}', \mathbf{h}']}{D(\mathbf{x}_0)[\mathbf{z}', \mathbf{h}']} - \frac{N(\mathbf{x}_0)[\mathbf{z}, \mathbf{h}]}{D(\mathbf{x}_0)[\mathbf{z}, \mathbf{h}]} \right) d\mathbf{x}_0 \right| \leq C (|\mathbf{z} - \mathbf{z}'| + \|\mathbf{h} - \mathbf{h}'\|_H)$$

for some positive constant C . \square

We prove these properties for the numerator and the denominator of h_{CR}^l .

Lemma 6.47 *We have*

$$|\Omega| \text{diam}(\Omega) \beta \leq \mathcal{G}_\gamma(\mathbf{x}_0) v_2(\mathbf{h}, \mathbf{x}_0) \leq |\Omega| G_\gamma(\mathbf{0}) (\beta + \mathcal{A}^2)$$

Proof : The proof is a direct consequence of the definition of $\mathcal{G}(\mathbf{x}_0)$ and of lemma 6.43. \square

We then prove the following

Lemma 6.48 *The function $H \times \Omega \rightarrow \mathbb{R}^+$ such that $(\mathbf{h}, \mathbf{x}_0) \rightarrow \mathcal{G}_\gamma(\mathbf{x}_0) v_2(\mathbf{h}, \mathbf{x}_0)$ is Lipschitz continuous in H uniformly in Ω .*

Lemma 6.49 *The function $\mathcal{H}_2 \times \Omega \rightarrow \mathbb{R}$ such that*

$$(\mathbf{z}, \mathbf{h}, \mathbf{x}_0) \rightarrow \mu_2(\mathbf{h}, \mathbf{x}_0) - d(z_1, \mathbf{h}, \mathbf{x}_0) + \mathbf{CR}^l(\mathbf{h}, \mathbf{x}_0)(z_2 - \mu_2(\mathbf{h}, \mathbf{x}_0)),$$

is bounded by $3\mathcal{A}$.

Proof : This is because $0 \leq \mathbf{CR}(\mathbf{h}, \mathbf{x}_0) \leq 1$, $0 \leq \mu_2(\mathbf{h}, \mathbf{x}_0) \leq \mathcal{A}$, and, according to lemma 6.45, because $0 \leq d(z_1, \mathbf{h}, \mathbf{x}_0) \leq \mathcal{A}$. \square

At this point we can prove the following

Proposition 6.50 *The function $\Omega \rightarrow \mathbb{R}^n$ such that $\mathbf{x} \rightarrow f_{CR}^l(\mathbf{z}, \mathbf{h}, \mathbf{x})$ is Lipschitz continuous uniformly in \mathcal{H}_2 .*

Proof : With the notations of lemma 6.46 we have

$$|G_\gamma \star h_{CR}^l(\mathbf{x}) - G_\gamma \star h_{CR}^l(\mathbf{y})| \leq \int_{\Omega} |G_\gamma(\mathbf{x} - \mathbf{x}_0) - G_\gamma(\mathbf{y} - \mathbf{x}_0)| \frac{|N(\mathbf{x}_0)[\mathbf{z}, \mathbf{h}]|}{D(\mathbf{x}_0)[\mathbf{z}, \mathbf{h}]} d\mathbf{x}_0$$

Because of lemmas 6.47 and 6.49 we have

$$|G_\gamma \star h_{CR}^l(\mathbf{x}) - G_\gamma \star h_{CR}^l(\mathbf{y})| \leq \frac{3\mathcal{A}}{|\Omega| \text{diam}(\Omega) \beta} \int_{\Omega} |G_\gamma(\mathbf{x} - \mathbf{x}_0) - G_\gamma(\mathbf{y} - \mathbf{x}_0)| d\mathbf{x}_0 \leq \frac{3\mathcal{A} \text{Lip}(G_\gamma)}{\text{diam}(\Omega) \beta} |\mathbf{x} - \mathbf{y}|,$$

hence the result. \square

We also have the analog of lemmas 6.23, 6.18, and theorem 6.21.

Lemma 6.51 *The function $d : \mathcal{H}_1 \times \Omega \rightarrow \mathbb{R}$ is bounded and Lipschitz continuous in \mathcal{H}_1 uniformly in Ω .*

Lemma 6.52 *The functions $H \times \Omega \rightarrow \mathbb{R}$ defined by*

$$(\mathbf{h}, \mathbf{x}_0) \rightarrow \frac{1}{\mathcal{G}_\gamma(\mathbf{x}_0)} \int_{\Omega} I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) G_\gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x}$$

and

$$(\mathbf{h}, \mathbf{x}_0) \rightarrow \frac{1}{\mathcal{G}_\gamma(\mathbf{x}_0)} \int_{\Omega} I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x}))^2 G_\gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x}$$

are bounded and Lipschitz continuous in H uniformly in Ω .

Theorem 6.53 *The function $H \times \Omega \rightarrow \mathbb{R}$ defined by*

$$(\mathbf{h}, \mathbf{x}_0) \rightarrow \frac{\mathbf{E}[\mathbf{Var}[X_{I_2^\sigma, \mathbf{h}} | X_{I_1^\sigma}^l](\mathbf{x}_0)]}{v_2(\mathbf{h}, \mathbf{x}_0)}$$

is Lipschitz continuous in H uniformly in Ω .

From which we deduce the

Theorem 6.54 *The function $\mathcal{H}_2 \times \Omega \rightarrow \mathbb{R}$ such that*

$$(\mathbf{z}, \mathbf{h}, \mathbf{x}_0) \rightarrow \mu_2(\mathbf{h}, \mathbf{x}_0) - d(z_1, \mathbf{h}, \mathbf{x}_0) + \mathbf{CR}^l(\mathbf{h}, \mathbf{x}_0)(z_2 - \mu_2(\mathbf{h}, \mathbf{x}_0)),$$

is Lipschitz continuous in \mathcal{H}_2 uniformly in Ω .

We can now prove the

Theorem 6.55 *The function $f_{CR}^l : \mathcal{H}_2 \times \Omega \rightarrow \mathbb{R}$ such that $(\mathbf{z}, \mathbf{h}, \mathbf{x}) \rightarrow f_{CR}^l(\mathbf{z}, \mathbf{h}, \mathbf{x})$ is Lipschitz continuous in \mathcal{H}_2 uniformly in Ω .*

Proof : The proof is just an application of lemma 6.46 to f_{CR}^l . \square

The combination of proposition 6.50 and theorem 6.55 yields the following

Theorem 6.56 *The function $f_{CR}^l : \mathcal{H}_2 \times \Omega \rightarrow \mathbb{R}$ such that $(\mathbf{z}, \mathbf{h}, \mathbf{x}) \rightarrow f_{CR}^l(\mathbf{z}, \mathbf{h}, \mathbf{x})$ is Lipschitz continuous.*

And we can conclude with the following theorem and proposition.

Theorem 6.57 *The function $F_{CR}^l : H \rightarrow H$ defined by*

$$F_{CR}^l(\mathbf{h}) = f_{CR}^l(I_1^\sigma, I_2^\sigma(\mathbf{Id} + \mathbf{h}), \mathbf{Id}) \nabla I_2^\sigma(\mathbf{Id} + \mathbf{h})$$

is Lipschitz continuous and bounded.

Proof : The proof follows exactly the same pattern as the proof of theorem 6.41 and uses theorem 6.55. \square

Proposition 6.58 *The function $\Omega \rightarrow \mathbb{R}^n$ such that $\mathbf{x} \rightarrow F_{CR}^l(\mathbf{h}(\mathbf{x}))$ satisfies*

$$|F_{CR}^l(\mathbf{h}(\mathbf{x})) - F_{CR}^l(\mathbf{h}(\mathbf{y}))| \leq K(|\mathbf{x} - \mathbf{y}| + |\mathbf{h}(\mathbf{x}) - \mathbf{h}(\mathbf{y})|),$$

for some constant $K > 0$.

Proof : The proof is similar to that of proposition 6.26 and follows from theorem 6.56 and the fact that the functions I_1^σ , I_2^σ and all its derivatives, are Lipschitz continuous.

\square

6.3.3 Cross Correlation

In this section we prove the Lipschitz-continuity of the mapping $H \rightarrow H$ defined by $F_{CC}^l(\mathbf{h})$ (equation (5.14)). The reasoning is completely analog to that of the global case. We start with some estimates which guaranty that the variance of $X_{I_1^\sigma, \mathbf{x}_0}$ is strictly positive.

Lemma 6.59 $\forall \mathbf{x}_0 \in \Omega$, the following inequalities are verified

$$\begin{aligned} 0 &\leq \mu_1(\mathbf{x}_0) \leq \mathcal{A}, \\ \beta &\leq v_1(\mathbf{x}_0) \leq \beta + \mathcal{A}^2. \end{aligned}$$

Proof :

$$\begin{aligned} \mu_1(\mathbf{x}_0) &= \int_{\mathbb{R}^2} i_1 \left(\frac{1}{\mathcal{G}_\gamma(\mathbf{x}_0)} \int_{\Omega} G_\beta(\mathbf{I}_h(\mathbf{x}) - \mathbf{i}) G_\gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x} \right) di_1 di_2 = \\ &\quad \frac{1}{\mathcal{G}_\gamma(\mathbf{x}_0)} \int_{\Omega} I_1^\sigma(\mathbf{x}) G_\gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x} \quad (6.26) \end{aligned}$$

and the first inequalities follow from the fact that $I_1^\sigma(\mathbf{x}) \in [0, \mathcal{A}]$. Similarly, for $v_1(\mathbf{x}_0)$, we have

$$\begin{aligned} v_1(\mathbf{x}_0) &= \int_{\mathbb{R}^2} i_1^2 \left(\frac{1}{\mathcal{G}_\gamma(\mathbf{x}_0)} \int_{\Omega} G_\beta(\mathbf{I}_h(\mathbf{x}) - \mathbf{i}) G_\gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x} \right) di_1 di_2 - \mu_1^2(\mathbf{x}_0) = \\ &\quad \beta + \frac{1}{\mathcal{G}_\gamma(\mathbf{x}_0)} \int_{\Omega} I_1^\sigma(\mathbf{x})^2 G_\gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x} - \\ &\quad \left(\frac{1}{\mathcal{G}_\gamma(\mathbf{x}_0)} \int_{\Omega} I_1^\sigma(\mathbf{x}) G_\gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x} \right)^2, \quad (6.27) \end{aligned}$$

from which the second inequalities follow. \square

We now show the boundedness and Lipschitz-continuity of the covariance of $X_{I_1^\sigma, \mathbf{x}_0}$ and $X_{I_2^\sigma, \mathbf{x}_0, \mathbf{h}}$.

Proposition 6.60 The function $H \times \Omega \rightarrow \mathbb{R}$ defined by $(\mathbf{h}, \mathbf{x}_0) \mapsto v_{1,2}(\mathbf{h}, \mathbf{x}_0)$ is bounded and Lipschitz continuous in H , uniformly in Ω .

Proof : Indeed, we have

$$\begin{aligned} v_{1,2}(\mathbf{h}, \mathbf{x}_0) &= \int_{\mathbb{R}^2} (i_1 - \mu_1(\mathbf{x}_0)) (i_2 - \mu_2(\mathbf{h}, \mathbf{x}_0)) \\ &\quad \frac{1}{\mathcal{G}_\gamma(\mathbf{x}_0)} \int_{\Omega} G_\beta(\mathbf{I}_h(\mathbf{x}) - \mathbf{i}) G_\gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x} di_1 di_2. \end{aligned}$$

Hence

$$\begin{aligned} v_{1,2}(\mathbf{h}, \mathbf{x}_0) &= \frac{1}{\mathcal{G}_\gamma(\mathbf{x}_0)} \int_{\Omega} \left(\int_{\mathbb{R}} (i_1 - \mu_1(\mathbf{x}_0)) g_\beta(I_1^\sigma(\mathbf{x}) - i_1) di_1 \right) \\ &\quad \left(\int_{\mathbb{R}} (i_2 - \mu_2(\mathbf{h}, \mathbf{x}_0)) g_\beta(I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) - i_2) di_2 \right) G_\gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x}, \end{aligned}$$

that is

$$\begin{aligned}
v_{1,2}(\mathbf{h}, \mathbf{x}_0) &= \frac{1}{\mathcal{G}_\gamma(\mathbf{x}_0)} \\
&\int_{\Omega} (I_1^\sigma(\mathbf{x}) - \mu_1(\mathbf{x}_0)) (I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) - \mu_2(\mathbf{h}, \mathbf{x}_0)) G_\gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x} \\
&= \frac{1}{\mathcal{G}_\gamma(\mathbf{x}_0)} \int_{\Omega} I_1^\sigma(\mathbf{x}) I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) G_\gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x} - \mu_1(\mathbf{x}_0) \mu_2(\mathbf{h}, \mathbf{x}_0). \quad (6.28)
\end{aligned}$$

Thus we have, $\forall \mathbf{x}_0 \in \Omega$, $|v_{1,2}(\mathbf{h}, \mathbf{x}_0)| \leq \mathcal{A}^2$, which proves the first part of the proposition. For the second part, since $\mu_2(\mathbf{h}, \mathbf{x}_0)$ is Lipschitz continuous uniformly in Ω (lemma 6.52), it suffices to show the Lipschitz continuity of the first term in the right-hand side. For this term we have,

$$\begin{aligned}
&\frac{1}{\mathcal{G}_\gamma(\mathbf{x}_0)} \left| \int_{\Omega} I_1^\sigma(\mathbf{x}) I_2^\sigma(\mathbf{x} + \mathbf{h}_1(\mathbf{x})) G_\gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x} - \int_{\Omega} I_1^\sigma(\mathbf{x}) I_2^\sigma(\mathbf{x} + \mathbf{h}_2(\mathbf{x})) G_\gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x} \right| \\
&\leq \frac{1}{\mathcal{G}_\gamma(\mathbf{x}_0)} \int_{\Omega} |I_1^\sigma(\mathbf{x})| |I_2^\sigma(\mathbf{x} + \mathbf{h}_1(\mathbf{x})) - I_2^\sigma(\mathbf{x} + \mathbf{h}_2(\mathbf{x}))| G_\gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x} \\
&\leq \mathcal{A} \text{Lip}(I_2^\sigma) G_\gamma(0) k_\Omega \int_{\Omega} |\mathbf{h}_1(\mathbf{x}) - \mathbf{h}_2(\mathbf{x})| d\mathbf{x}.
\end{aligned}$$

Hence (by Cauchy-Schwarz):

$$\begin{aligned}
&\frac{1}{\mathcal{G}_\gamma(\mathbf{x}_0)} \left| \int_{\Omega} I_1^\sigma(\mathbf{x}) I_2^\sigma(\mathbf{x} + \mathbf{h}_1(\mathbf{x})) G_\gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x} - \int_{\Omega} I_1^\sigma(\mathbf{x}) I_2^\sigma(\mathbf{x} + \mathbf{h}_2(\mathbf{x})) G_\gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x} \right| \leq \mathcal{L} \|\mathbf{h}_1 - \mathbf{h}_2\|_H,
\end{aligned}$$

where the constant \mathcal{L} is independent of $\mathbf{x}_0 \in \Omega$. \square

Theorem 6.61 *The function $H \times \Omega \rightarrow \mathbb{R}$ defined by $(\mathbf{h}, \mathbf{x}) \mapsto \mathbf{CC}^l(\mathbf{h}, \mathbf{x})$ is bounded and Lipschitz continuous in H , uniformly in Ω .*

Proof : The cross-correlation function \mathbf{CC}^l is bounded by 1. Moreover, we have

$$\mathbf{CC}^l(\mathbf{h}, \mathbf{x}) = \frac{v_{1,2}(\mathbf{h}, \mathbf{x})^2}{v_1(\mathbf{x}) v_2(\mathbf{h}, \mathbf{x})}, \quad (6.29)$$

with:

- $v_{1,2}(\mathbf{h}, \mathbf{x})$ bounded and Lipschitz-continuous in H uniformly in Ω (proposition 6.60).

- $v_2(\mathbf{h}, \mathbf{x})$ bounded and Lipschitz-continuous in H , uniformly in Ω (readily seen from lemmas 6.15 and 6.52).
- $v_2(\mathbf{h}, \mathbf{x}) > 0$ (lemma 6.15).
- $v_1(\mathbf{x})$ bounded and > 0 (lemma 6.59).

We may therefore apply proposition 6.1. \square

Theorem 6.62 *The function $\mathcal{H}_2 \times \Omega \longrightarrow \mathbb{R}$ defined by*

$$(z_1, z_2, \mathbf{h}, \mathbf{x}) \mapsto L_{\mathbf{CC}, \mathbf{h}}^l(z_1, z_2, \mathbf{x})$$

is bounded and Lipschitz continuous in \mathcal{H}_2 , uniformly in Ω .

Proof : We have

$$L_{\mathbf{CC}, \mathbf{h}}^l(z_1, z_2, \mathbf{x}) = \frac{-2}{\mathcal{G}_\gamma(\mathbf{x})} \left[\frac{v_{1,2}(\mathbf{h}, \mathbf{x})}{v_2(\mathbf{h}, \mathbf{x})} \left(\frac{z_1 - \mu_1(\mathbf{x})}{v_1(\mathbf{x})} \right) - \mathbf{CC}^g(\mathbf{h}, \mathbf{x}) \left(\frac{z_2 - \mu_2(\mathbf{h}, \mathbf{x})}{v_2(\mathbf{h}, \mathbf{x})} \right) \right].$$

Taking into account the properties mentioned in the proof of proposition 6.61, the boundedness and Lipschitz continuity of \mathbf{CC}^l (proposition 6.61) and of μ_2 (lemma 6.52), plus the fact that $\mathcal{G}_\gamma(\mathbf{x}) > 0$, we see that $L_{\mathbf{CC}, \mathbf{h}}^l$ may be written as

$$L_{\mathbf{CC}, \mathbf{h}}^l(z_1, z_2, \mathbf{x}) = f_1(\mathbf{h}, \mathbf{x}) z_1 + f_2(\mathbf{h}, \mathbf{x}) z_2 + f_3(\mathbf{h}, \mathbf{x}),$$

where the functions $H \times \Omega \longrightarrow \mathbb{R}$ f_1, f_2 and f_3 are bounded and Lipschitz continuous in H uniformly in Ω , from where the result readily follows. \square

Theorem 6.63 *The function $f_{\mathbf{CC}}^l : \mathcal{H}_2 \times \Omega \longrightarrow \mathbb{R}$, defined by*

$$f_{\mathbf{CC}}^l(\mathbf{z}, \mathbf{x}) = (G_\gamma \star L_{\mathbf{CC}, \mathbf{h}}^l)(\mathbf{z}, \mathbf{x})$$

is bounded and Lipschitz continuous.

Proof : Indeed, let $f : \mathcal{H}_2 \times \Omega \longrightarrow \mathbb{R}$ be bounded by B_f and Lipschitz continuous in \mathcal{H}_2 uniformly in Ω , of constant L_f (these hypotheses are verified by $L_{\mathbf{CC}, \mathbf{h}}^l$ according to theorem 6.62). We have

$$\left| \int_{\Omega} G_\gamma(\mathbf{x} - \mathbf{x}_0) (f(\mathbf{z}, \mathbf{h}, \mathbf{x}_0) - f(\mathbf{z}', \mathbf{h}', \mathbf{x}_0)) d\mathbf{x}_0 \right| \leq G_\gamma(0) |\Omega| L_f (|\mathbf{z} - \mathbf{z}'| + \|\mathbf{h} - \mathbf{h}'\|_H),$$

and

$$|G_\gamma \star f(\mathbf{z}, \mathbf{h}, \mathbf{x}) - G_\gamma \star f(\mathbf{z}, \mathbf{h}, \mathbf{y})| \leq \int_{\Omega} |G_\gamma(\mathbf{x} - \mathbf{x}_0) - G_\gamma(\mathbf{y} - \mathbf{x}_0)| |f(\mathbf{z}, \mathbf{h}, \mathbf{x}_0)| d\mathbf{x}_0 \leq B_f \text{Lip}(G_\gamma) |\Omega| |\mathbf{x} - \mathbf{y}|,$$

hence the result.

□

Theorem 6.64 *The function $F_{CC}^l : H \rightarrow H$ defined by*

$$F_{CC}^l(\mathbf{h}) = f_{CC}^l(I_1^\sigma, I_2^\sigma(\mathbf{Id} + \mathbf{h}), \mathbf{Id}) \nabla I_2^\sigma(\mathbf{Id} + \mathbf{h})$$

is Lipschitz continuous and bounded.

Proof : The proof follows exactly the same pattern as the proof of theorem 6.41 and uses theorem 6.63. □

Proposition 6.65 *The function $\Omega \rightarrow \mathbb{R}^n$ such that $\mathbf{x} \rightarrow F_{CC}^l(\mathbf{h}(\mathbf{x}))$ satisfies*

$$|F_{CC}^l(\mathbf{h}(\mathbf{x})) - F_{CC}^l(\mathbf{h}(\mathbf{y}))| \leq K(|\mathbf{x} - \mathbf{y}| + |\mathbf{h}(\mathbf{x}) - \mathbf{h}(\mathbf{y})|),$$

for some constant $K > 0$.

Proof : The proof is similar to that of proposition 6.26 and follows from theorem 6.63 and the fact that the functions I_1^σ, I_2^σ and all its derivatives, are Lipschitz continuous.

□

Part III

Implementation Aspects

Chapter 7

Numerical Schemes

This chapter describes the numerical schemes employed in discretizing the continuous evolution equations (see equation (1.2) on page 39) for the different matching and regularization operators. We use spatial difference schemes to evaluate differential operators and an explicit forward discretization in time (Euler method). Parzen window estimates are computed by recursive filtering [31] of the discrete joint intensity histogram. These elements are detailed in the following sections.

7.1 Regularization Operators

We begin by describing the schemes for the differential operators used for regularization. We use a schematic notation for the description of the finite-difference schemes. For instance, let us denote by $L_p^{i,j,k}$ and $h_p^{i,j,k}$, the components ($p = 1, 2, 3$) of $\Delta \mathbf{h}$ and \mathbf{h} at a grid point (i, j, k) in the discrete image domain. The voxel size in all directions is assumed to be equal to one. A possible scheme for $\alpha \Delta \mathbf{h}$ would be

$$L_p^{i,j} = \alpha \left(h_p^{i+1,j} + h_p^{i-1,j} + h_p^{i,j-1} + h_p^{i,j+1} - 4 h_p^{i,j} \right), \quad (7.1)$$

in the 2D case ($p = 1, 2$) and

$$L_p^{i,j,k} = \alpha \left(h_p^{i+1,j,k} + h_p^{i-1,j,k} + h_p^{i,j-1,k} + h_p^{i,j+1,k} + h_p^{i,j,k-1} + h_p^{i,j,k+1} - 6 h_p^{i,j,k} \right), \quad (7.2)$$

in the 3D case ($p = 1, 2, 3$), which we write schematically as

$$L_p^{i,j} = \begin{array}{|c|c|c|} \hline & & \\ \hline & 1 & \\ \hline 1 & -4 & 1 \\ \hline & 1 & \\ \hline \end{array} \quad \text{and} \quad L_p^{i,j,k} = \begin{array}{|c|c|c|} \hline & & \\ \hline & 1 & \\ \hline 1 & -6 & 1 \\ \hline & 1 & \\ \hline & & \\ \hline & 1 & \\ \hline & & \\ \hline \end{array}$$

$h_p \times \alpha$ $h_p \times \alpha$

respectively. In this notation, the tables represent the discrete grid and contain the weights associated to each pixel (voxel). The weight is zero if the voxel is empty. The function to which the grid corresponds is written at the bottom, together with any global weight (for example $h_p \times \alpha$ above means the grid is that of h_p , weighted globally by α). In each table, the index i is assumed to increase from left to right and the index j from top to bottom, the center being the coordinates (i, j) . In the 3D case, each of the three stacked tables represents a different index k , which increases in the bottom-up direction, the one in the middle being that of index k . Although less compact than the notations (7.1) and (7.2), these schematic representations have the advantage of clearly showing the position of the weights within the neighborhood of each voxel.

7.1.1 The Linearized Elasticity Operator

In this section, we describe our numerical schemes for computing the linearized elasticity operator:

$$A = c (\xi \Delta \mathbf{h} + (1 - \xi) \nabla (\nabla \cdot \mathbf{h})). \quad (7.3)$$

Our schemes are based on a first order Taylor expansion of (7.3). For $n = 2$, it yields the following scheme:

$$A_1^{i,j} = \begin{array}{|c|c|c|} \hline & 1 & \\ \hline 1 & -4 & 1 \\ \hline & 1 & \\ \hline \end{array} + \begin{array}{|c|c|c|} \hline & & \\ \hline 1 & -2 & 1 \\ \hline & & \\ \hline \end{array} + \begin{array}{|c|c|c|} \hline 1 & & -1 \\ \hline & & \\ \hline -1 & & 1 \\ \hline \end{array},$$

$h_1 \times \xi$ $h_1 \times (1 - \xi)$ $h_2 \times \frac{1}{4}(1 - \xi)$

$$A_2^{i,j} = \begin{array}{|c|c|c|} \hline & 1 & \\ \hline 1 & -4 & 1 \\ \hline & 1 & \\ \hline \end{array} + \begin{array}{|c|c|c|} \hline & 1 & \\ \hline & -2 & \\ \hline & 1 & \\ \hline \end{array} + \begin{array}{|c|c|c|} \hline 1 & & -1 \\ \hline & & \\ \hline -1 & & 1 \\ \hline \end{array}.$$

$h_2 \times \xi$ $h_2 \times (1 - \xi)$ $h_1 \times \frac{1}{4}(1 - \xi)$

7.1.2 The Nagel-Enkelmann Operator

2D case

Our implementation corresponds to the scheme proposed by Alvarez et al [5]. Let now $A = \text{div}(\mathbf{T}_{I_1^\sigma} Dh)$, where

$$\mathbf{T}_{I_1^\sigma} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

The scheme is the following, for $p = 1, 2$:

$$\begin{aligned}
 A_p^{i,j} = & \begin{array}{c} \begin{array}{|c|c|c|} \hline & & \\ \hline & 1 & 1 \\ \hline & & \\ \hline \end{array} * \begin{array}{|c|c|c|} \hline & & \\ \hline & -1 & 1 \\ \hline & & \\ \hline \end{array} + \begin{array}{|c|c|c|} \hline & & \\ \hline 1 & 1 & \\ \hline & & \\ \hline \end{array} * \begin{array}{|c|c|c|} \hline & & \\ \hline 1 & -1 & \\ \hline & & \\ \hline \end{array} \\
 & \frac{1}{2}a \qquad h_p \qquad \frac{1}{2}a \qquad h_p \\
 + & \begin{array}{|c|c|c|} \hline & 1 & \\ \hline & 1 & \\ \hline & & \\ \hline \end{array} * \begin{array}{|c|c|c|} \hline & 1 & \\ \hline & -1 & \\ \hline & & \\ \hline \end{array} + \begin{array}{|c|c|c|} \hline & & \\ \hline & 1 & \\ \hline & & \\ \hline \end{array} * \begin{array}{|c|c|c|} \hline & & \\ \hline & -1 & \\ \hline & & \\ \hline \end{array} \\
 & \frac{1}{2}c \qquad h_p \qquad \frac{1}{2}c \qquad h_p \\
 + & \begin{array}{|c|c|c|} \hline & & \\ \hline & 1 & \\ \hline & & 1 \\ \hline \end{array} * \begin{array}{|c|c|c|} \hline & & \\ \hline & -1 & \\ \hline & & 1 \\ \hline \end{array} + \begin{array}{|c|c|c|} \hline & & \\ \hline 1 & & \\ \hline & 1 & \\ \hline \end{array} * \begin{array}{|c|c|c|} \hline & & \\ \hline 1 & & \\ \hline & -1 & \\ \hline \end{array} \\
 & \frac{1}{4}b \qquad h_p \qquad \frac{1}{4}b \qquad h_p \\
 - & \begin{array}{|c|c|c|} \hline & & 1 \\ \hline & 1 & \\ \hline & & \\ \hline \end{array} * \begin{array}{|c|c|c|} \hline & & 1 \\ \hline & & -1 \\ \hline & & \\ \hline \end{array} - \begin{array}{|c|c|c|} \hline & & \\ \hline & & 1 \\ \hline & & \\ \hline \end{array} * \begin{array}{|c|c|c|} \hline & & \\ \hline & -1 & \\ \hline & & \\ \hline \end{array} \\
 & \frac{1}{4}b \qquad h_p \qquad \frac{1}{4}b \qquad h_p
 \end{aligned}$$

3D case

This scheme generalizes readily to the 3D case. In order to write explicitly the 3D scheme in a compact way, we take profit of the very simple form of this scheme to introduce a more compact notation. We shall write

$$S_{\frac{1}{2}}(a, x^+) \equiv \begin{array}{|c|c|c|} \hline & & \\ \hline & 1 & 1 \\ \hline & & \\ \hline \end{array} * \begin{array}{|c|c|c|} \hline & & \\ \hline & -1 & 1 \\ \hline & & \\ \hline \end{array},$$

$\frac{1}{2}a$
 h_p

where x^+ indicates the direction defined by the voxels with non-null weights, starting at the center. With this notation, we write the 2D Nagel-Enkelmann operator above as

$$\begin{aligned}
 A_p^{i,j} = & S_{\frac{1}{2}}(a, x^+) + S_{\frac{1}{2}}(a, x^-) + S_{\frac{1}{2}}(c, y^+) + S_{\frac{1}{2}}(c, y^-) \\
 & + S_{\frac{1}{4}}(b, x^+y^+) + S_{\frac{1}{4}}(b, x^-y^-) - S_{\frac{1}{4}}(b, x^+y^-) - S_{\frac{1}{4}}(b, x^-y^+).
 \end{aligned}$$

In the 3D case, we have

$$\mathbf{T}_{I_1^g} = \begin{pmatrix} a & b & c \\ b & d & e \\ c & e & f \end{pmatrix},$$

and the corresponding scheme is, for $p = 1, 2, 3$,

$$\begin{aligned} A_p^{i,j,k} = & S_{\frac{1}{2}}(a, x^+) + S_{\frac{1}{2}}(a, x^-) + S_{\frac{1}{2}}(d, y^+) + S_{\frac{1}{2}}(d, y^-) + S_{\frac{1}{2}}(f, z^+) + \\ & S_{\frac{1}{2}}(f, z^-) + S_{\frac{1}{4}}(b, x^+ y^+) + S_{\frac{1}{4}}(b, x^- y^-) - S_{\frac{1}{4}}(b, x^+ y^-) - S_{\frac{1}{4}}(b, x^- y^+) \\ & + S_{\frac{1}{4}}(c, x^+ z^+) + S_{\frac{1}{4}}(c, x^- z^-) - S_{\frac{1}{4}}(c, x^+ z^-) - S_{\frac{1}{4}}(c, x^- z^+) \\ & + S_{\frac{1}{4}}(e, y^+ z^+) + S_{\frac{1}{4}}(e, y^- z^-) - S_{\frac{1}{4}}(e, y^+ z^-) - S_{\frac{1}{4}}(e, y^- z^+). \end{aligned}$$

7.2 Dissimilarity Terms

This section discusses implementation issues concerning the three global matching functions $F_{\text{MI}}^g(\mathbf{h})$, $F_{\text{CR}}^g(\mathbf{h})$ and $F_{\text{CC}}^g(\mathbf{h})$ as defined in (5.9) on page 85, and the local matching function $F_{\text{CC}}^l(\mathbf{h})$ as defined in (5.14) on page 86. The remaining two functions will be treated in section 7.3. The reason for this separation is that the global functions and $F_{\text{CC}}^l(\mathbf{h})$ can be computed as direct estimations of their respective definitions, while $F_{\text{MI}}^l(\mathbf{h})$ and $F_{\text{CR}}^l(\mathbf{h})$ require two convolutions, which makes them very hard to implement with limited memory.

By observing their respective definitions, it appears that some specific operations are required to implement the global functions and $F_{\text{CC}}^l(\mathbf{h})$. These operations are the following.

- *Convolutions*

The convolutions by a Gaussian kernel are approximated by recursive filtering using the smoothing operator introduced by Deriche [31]. Given a discrete 1D input sequence $x(n)$, $n = 1, \dots, M$, its convolution by the smoothing operator $S_\alpha(n) = k (\alpha|n| + 1) e^{-\alpha|n|}$ is calculated efficiently as (see [31]):

$$y(n) = (S_\alpha \star x)(n) = y_1(n) + y_2(n),$$

where

$$\begin{cases} y_1(n) = k (x(n) + e^{-\alpha} (\alpha - 1) x(n-1)) \\ \quad \quad \quad + 2 e^{-\alpha} y_1(n-1) - e^{-2\alpha} y_1(n-2), \\ y_2(n) = k (e^{-\alpha} (\alpha + 1) x(n+1) - e^{-2\alpha} x(n+2)) \\ \quad \quad \quad + 2 e^{-\alpha} y_2(n+1) - e^{-2\alpha} y_2(n+2). \end{cases}$$

The normalization constant k is chosen by requiring that $\int_{\mathbb{R}} S_\alpha(t) dt = 1$, which yields $k = \alpha/4$. This scheme is very efficient since the number of operations

required is independent of the smoothing parameter α . The smoothing filter can be readily generalized to n dimensions by defining the separable filter $T_\alpha(\mathbf{x}) = \prod_{i=1}^n S_\alpha(x_i)$.

We can determine an optimal value of α to approximate a Gaussian of variance β . Requiring that $G_\beta(t)$ and $S_\alpha(t)$ have the same L^2 norm yields the relation

$$\alpha = \frac{16}{5\sqrt{\pi}\beta} \simeq 1.8/\sqrt{\beta},$$

while minimizing the L^2 norm of $(G_\beta(t) - S_\alpha(t))$ yields

$$\alpha \simeq 1.695/\sqrt{\beta}.$$

As showed by Alvarez et al. in [2], the recursive procedure above can be seen as a numerical implementation of the heat equation. The convolution by S_α presents the advantage of being computable exactly by a recursive filter of order two, giving very precise results and fast computations.

By computing the derivative of $S_\alpha(t)$,

$$S'_\alpha(t) = -\frac{1}{4} \alpha^3 t e^{-\alpha|t|},$$

we obtain a derivative filter of which a recursive realization can be similarly obtained:

$$y(n) = (S'_\alpha \star x)(n) = (S_\alpha \star x')(n) = -\frac{1}{4} \alpha^3 e^{-\alpha} (y_1(n) + y_2(n)),$$

where now

$$\begin{cases} y_1(n) &= x(n-1) + 2 e^{-\alpha} y_1(n-1) - e^{-2\alpha} y_1(n-2), \\ y_2(n) &= x(n+1) + 2 e^{-\alpha} y_2(n+1) - e^{-2\alpha} y_2(n+2). \end{cases}$$

- *Interpolation*

Terms of the form $\nabla I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x}))$ are calculated as follows. Convolution of I_2 with the derivative filter is used to compute the components of ∇I_2^σ on $\Omega^{i,j,k}$. Then their value for an arbitrary position (not necessarily on the grid) is computed using a trilinear interpolation scheme, defined as follows. Let f be the function to be evaluated at $(i+x, j+y, k+z)$, where (i, j, k) is a point on the grid and $(x, y, z) \in [0, 1]^3$. We set

$$\begin{aligned} V_1 &= f(i, j, k), & V_2 &= f(i+1, j, k), \\ V_3 &= f(i+1, j+1, k), & V_4 &= f(i, j+1, k), \\ V_5 &= f(i, j, k+1), & V_6 &= f(i+1, j, k+1), \\ V_7 &= f(i+1, j+1, k+1), & V_8 &= f(i, j+1, k+1). \end{aligned}$$

Then the value $f(i + x, j + y, k + z)$ is estimated as

$$\begin{aligned} V = & V_1 (1 - x)(1 - y)(1 - z) + V_2 x(1 - y)(1 - z) \\ & V_3 xy(1 - z) + V_4 (1 - x)y(1 - z) + \\ & V_5 (1 - x)(1 - y)z + V_6 x(1 - y)z + \\ & V_7 xyz + V_8 (1 - x)yz. \end{aligned}$$

The same interpolation scheme is used for estimating the value of $\mathcal{I}_2^{\mathcal{F}}(\mathbf{x} + \mathbf{h}(\mathbf{x}))$.

- *Density estimation*

Parzen density estimates are obtained by smoothing the discrete joint histogram of intensities. To describe this procedure, we define the piecewise constant function $\mathbf{v} : \Omega \rightarrow [0, N]^2 \subset \mathbb{N}^2$ by quantification of $\mathbf{I}_{\mathbf{h}}(\mathbf{x})$ into $N + 1$ intensity levels (bins):

$$\mathbf{v}(\mathbf{x}) = \begin{pmatrix} \lfloor \zeta I_1^\sigma(\mathbf{x}) \rfloor \\ \lfloor \zeta I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) \rfloor \end{pmatrix} = \begin{cases} (0, 0)^T & \text{on } \Omega_{0,0} \\ \vdots \\ (N, N)^T & \text{on } \Omega_{N,N}, \end{cases}$$

where $\zeta = N/\mathcal{A}$, $\lfloor \cdot \rfloor$ denotes the floor operator in \mathbb{R}^+ , i.e. the function $\mathbb{R}^+ \rightarrow \mathbb{N}$ such that $\lfloor x \rfloor = \max\{n \in \mathbb{N} : n \leq x\}$, and $\{\Omega_{k,l}\}_{(k,l) \in [0,N]^2}$ is a partition of Ω . We then compute, setting $\beta' = \zeta^2 \beta$,

$$\begin{aligned} P_{\mathbf{h}}(\mathbf{i}) &= \frac{1}{|\Omega|} \int_{\Omega} G_{\beta}(\mathbf{I}_{\mathbf{h}}(\mathbf{x}) - \mathbf{i}) \, d\mathbf{x} \\ &= \frac{\zeta}{|\Omega|} \int_{\Omega} G_{\beta'}(\zeta(\mathbf{I}_{\mathbf{h}}(\mathbf{x}) - \mathbf{i})) \, d\mathbf{x} \simeq \frac{\zeta}{|\Omega|} \int_{\Omega} G_{\beta'}(\mathbf{v}(\mathbf{x}) - \zeta \mathbf{i}) \, d\mathbf{x} \\ &= \frac{\zeta}{|\Omega|} \sum_{k=0}^N \sum_{l=0}^N \int_{\Omega_{k,l}} G_{\beta'}(k - \zeta i_1, l - \zeta i_2) \, d\mathbf{x} \\ &= \zeta \sum_{k=0}^N \sum_{l=0}^N \underbrace{|\Omega_{k,l}|/|\Omega|}_{H(k,l)} G_{\beta'}(k - \zeta i_1, l - \zeta i_2) = \zeta (H \star G_{\beta'})(\zeta \mathbf{i}), \end{aligned}$$

H being the discrete joint histogram. The convolution is performed by recursive filtering as described above. Note that this way of computing $P_{\mathbf{h}}$ is quite efficient since only one pass on the images is required, followed by the convolution.

These basic tools being described, we now state pseudo-algorithms of the way the different matching functions are computed.

Algorithm 7.1 ($F_{\text{MI}}^g(\mathbf{h})$)

- Estimate $P_{\mathbf{h}}(\mathbf{i})$ and its marginals.
- Estimate $L_{\text{MI},\mathbf{h}}^g(\mathbf{i})$ (equation (5.6) on page 84) using centered finite-differences for the derivatives.
- Estimate $f_{\text{MI}}^g(\mathbf{i}, \mathbf{h}) = G_{\beta} \star L_{\text{MI},\mathbf{h}}^g(\mathbf{i})$ by recursive smoothing.
- Estimate $F_{\text{MI}}^g(\mathbf{h})(\mathbf{x}) = f_{\text{MI}}^g(\mathbf{I}_{\mathbf{h}}(\mathbf{x}), \mathbf{h}) \nabla I_2^g(\mathbf{x} + \mathbf{h}(\mathbf{x}))$.

Algorithm 7.2 ($F_{\text{CR}}^g(\mathbf{h})$)

- Estimate $P_{\mathbf{h}}(\mathbf{i})$ and its marginals.
- Estimate $\mu_2(\mathbf{h}), v_2(\mathbf{h}), \mu_{2|1}(i_1, \mathbf{h})$ and $\text{CR}^g(\mathbf{h})$ using equations (4.3), (4.4), (4.5), (4.10) and table 4.1, on pages 66–68. Here the integrals are estimated by finite sums on the interval $[0, \mathcal{A}]$.
- Estimate $L_{\text{CR},\mathbf{h}}^g(\mathbf{i})$ (equation (5.7) on page 84).
- Estimate $f_{\text{CR}}^g(\mathbf{i}, \mathbf{h}) = G_{\beta} \star L_{\text{CR},\mathbf{h}}^g(\mathbf{i})$ by recursive smoothing.
- Estimate $F_{\text{CR}}^g(\mathbf{h})(\mathbf{x}) = f_{\text{CR}}^g(\mathbf{I}_{\mathbf{h}}(\mathbf{x}), \mathbf{h}) \nabla I_2^g(\mathbf{x} + \mathbf{h}(\mathbf{x}))$.

Algorithm 7.3 ($F_{\text{CC}}^g(\mathbf{h})$)

- Estimate the values $\mu_2(\mathbf{h})$ and $v_2(\mathbf{h})$ using equations (6.9), (6.10) and (6.11) on page 97 and the values $\mu_1, v_1, v_{1,2}(\mathbf{h})$ and $\text{CC}^g(\mathbf{h})$ using equations (6.12), (6.13), (6.14) and (6.15) on pages 103–104.
- Estimate $L_{\text{CC},\mathbf{h}}^g(\mathbf{i})$ (equation (5.8) on page 85).
- Estimate $F_{\text{CC}}^g(\mathbf{h})(\mathbf{x}) = L_{\text{CC},\mathbf{h}}^g(\mathbf{I}_{\mathbf{h}}(\mathbf{x})) \nabla I_2^g(\mathbf{x} + \mathbf{h}(\mathbf{x}))$.

Algorithm 7.4 ($F_{\text{CC}}^l(\mathbf{h})$)

- Estimate the function $\mathcal{G}_{\gamma}(\mathbf{x})$ using equation (4.13) on page 68, the functions $\mu_2(\mathbf{h}, \mathbf{x})$ and $v_2(\mathbf{h}, \mathbf{x})$ using equations (6.22), (6.23) and (6.24) on page 110 and the functions $\mu_1(\mathbf{x}), v_1(\mathbf{x}), v_{1,2}(\mathbf{h}, \mathbf{x})$ and $\text{CC}^l(\mathbf{h}, \mathbf{x})$ using equations (6.26), (6.27), (6.28) and (6.29) on pages 116–117. These estimations are computed by convolutions using recursive filtering.

- Estimate $(G_\gamma \star L_{\text{CC},\mathbf{h}}^l)(\mathbf{i}, \mathbf{x})$ (see equation (5.13) on page 85) as:

$$(G_\gamma \star L_{\text{CC},\mathbf{h}}^l)(\mathbf{i}, \mathbf{x}) = (G_\gamma \star f_1)(\mathbf{x}) i_1 + (G_\gamma \star f_2)(\mathbf{x}) i_2 + (G_\gamma \star f_3)(\mathbf{x}),$$

where

$$\begin{cases} f_1(\mathbf{x}) &= \frac{-2 v_{1,2}(\mathbf{h}, \mathbf{x})}{\mathcal{G}_\gamma(\mathbf{x}) v_1(\mathbf{x}) v_2(\mathbf{h}, \mathbf{x})}, \\ f_2(\mathbf{x}) &= \frac{2 \text{CC}^l(\mathbf{h}, \mathbf{x})}{\mathcal{G}_\gamma(\mathbf{x}) v_2(\mathbf{h}, \mathbf{x})}, \\ f_3(\mathbf{x}) &= -\left(f_1(\mathbf{x}) \mu_1(\mathbf{x}) + f_2(\mathbf{x}) \mu_2(\mathbf{h}, \mathbf{x})\right). \end{cases}$$

- Estimate $F_{\text{CC}}^l(\mathbf{h})(\mathbf{x}) = (G_\gamma \star L_{\text{CC},\mathbf{h}}^l)(\mathbf{I}_h(\mathbf{x}), \mathbf{x}) \nabla I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x}))$.

Note that algorithm 7.4 is similar to the algorithm proposed by Cachier and Pennec in [21], modulo the adding of a positive multiple of β in all the denominators. This factor is crucial for the Lipschitz-continuity of $F_{\text{CC}}^l(\mathbf{h})$.

7.3 Approximate Implementations of $F_{\text{MI}}^l(\mathbf{h})$ and $F_{\text{CR}}^l(\mathbf{h})$

This section discusses the implementation of the functions $F_{\text{MI}}^l(\mathbf{h})$ and $F_{\text{CR}}^l(\mathbf{h})$, defined in theorem 5.2 on page 86. These two functions are much more difficult to compute than those of the previous section. The reason for this is that they involve two convolutions, one with respect to the intensity variable \mathbf{i} and the other with respect to the space variable \mathbf{x} . A possible way to implement them would be to estimate the functions $L_{\text{MI},\mathbf{h}}^l$ and $L_{\text{CR},\mathbf{h}}^l$, and then “smooth” these functions, e.g. by recursive filtering. The problem is that this would require a dense data structure of dimension $(n+1)$ for $L_{\text{CR},\mathbf{h}}^l$ and $(n+2)$ for $L_{\text{MI},\mathbf{h}}^l$. With 3D images, it becomes extremely difficult to maintain these four and five-dimensional structures due to memory space limitations, not counting the computational effort of smoothing them, which has to be done at each iteration of the minimization flow. Our implementations rely on using directly the “unsmoothed” versions of $L_{\text{MI},\mathbf{h}}^l$ and $L_{\text{CR},\mathbf{h}}^l$, i.e. on eliminating both convolutions. We note the two functions obtained $\tilde{F}_{\text{MI}}^l(\mathbf{h})$ and $\tilde{F}_{\text{CR}}^l(\mathbf{h})$. Their implementation is described in more detail in the following sections.

7.3.1 Mutual Information

We define $\tilde{F}_{\text{MI}}^l(\mathbf{h})$ by eliminating the convolutions in the definition of $F_{\text{MI}}^l(\mathbf{h})$. It remains to estimate the function $L_{\text{MI},\mathbf{h}}^l(I_1^\sigma, I_2^\sigma(\mathbf{Id} + \mathbf{h}), \mathbf{Id})$, which is done using equations (6.17), (6.18) and (6.19), starting on page 106. Note that clearly, since all the denominators involved are strictly positive and bounded, the function $\mathbf{h} \rightarrow L_{\text{MI},\mathbf{h}}^l(I_1^\sigma, I_2^\sigma(\mathbf{Id} + \mathbf{h}), \mathbf{Id})$ is also Lipschitz-continuous, and therefore so is $\tilde{F}_{\text{MI}}^l(\mathbf{h})$.

Note however that the integrals involved can no longer be approximated by recursive filtering. The filtering becomes non-stationary. We therefore propose the following algorithm.

Algorithm 7.5 ($\tilde{F}_{\text{MI}}^l(\mathbf{h})$)

- Estimate for each \mathbf{y} on the discrete grid the value of $a(I_2^\sigma(\mathbf{y} + \mathbf{h}(\mathbf{y})), \mathbf{y}, \mathbf{h})$ using equation (6.17) on page 106, approximating the integrals by a finite sum on a sufficiently large¹ neighborhood around \mathbf{y} .
- Estimate in a similar way the value of $A(I_1^\sigma(\mathbf{y}), I_2^\sigma(\mathbf{y} + \mathbf{h}(\mathbf{y})), \mathbf{y}, \mathbf{h})$ using equation (6.18) on page 106, of $\mathcal{G}_\gamma(\mathbf{x})$ using equation (4.13) on page 68 and of $L_{\text{MI},\mathbf{h}}^l(I_1^\sigma(\mathbf{y}), I_2^\sigma(\mathbf{y} + \mathbf{h}(\mathbf{y})), \mathbf{y})$ using equation (6.19) on page 106.
- Estimate $\tilde{F}_{\text{MI}}^l(\mathbf{h})(\mathbf{y}) = L_{\text{MI},\mathbf{h}}^l(I_1^\sigma(\mathbf{y}), I_2^\sigma(\mathbf{y} + \mathbf{h}(\mathbf{y})), \mathbf{y}) \nabla I_2^\sigma(\mathbf{y} + \mathbf{h}(\mathbf{y}))$.

7.3.2 Correlation Ratio

Similarly to the previous case, we eliminate the two convolutions in the definition of $F_{\text{CR}}^l(\mathbf{h})$. It remains to estimate the function $L_{\text{CR},\mathbf{h}}^l(I_1^\sigma, I_2^\sigma(\text{Id} + \mathbf{h}), \text{Id})$, where $\mu_{2|1}(i_1, \mathbf{h}, \mathbf{x}_0)$ is given by (see lemma 6.45 on page 112):

$$\mu_{2|1}(i_1, \mathbf{h}, \mathbf{x}_0) = \frac{\int_{\Omega} g_\beta(i_1 - I_1^\sigma(\mathbf{x})) I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) G_\gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x}}{\int_{\Omega} g_\beta(i_1 - I_1^\sigma(\mathbf{x})) G_\gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x}}. \quad (7.4)$$

However, there is still another difficulty in this case. The value of $\mathbf{CR}^l(\mathbf{h}, \mathbf{x})$, which is needed in the computation of $L_{\text{CR},\mathbf{h}}^l$, requires itself a convolution with respect to the intensity variable (see e.g. the proof of lemma 6.44 and in particular equation (6.25) on page 111). For this reason, we make another approximation, namely that of replacing

$$\mathbf{CR}^l(\mathbf{h}, \mathbf{x}) = \int_{\mathbb{R}} \frac{v_{2|1}(i_1, \mathbf{h}, \mathbf{x})}{v_2(\mathbf{h}, \mathbf{x})} p(i_1) di_1$$

in the definition of $L_{\text{CR},\mathbf{h}}^l$, by the function

$$\theta_1(i_1, \mathbf{h}, \mathbf{x}) = \frac{v_{2|1}(i_1, \mathbf{h}, \mathbf{x})}{v_2(\mathbf{h}, \mathbf{x})}, \quad (7.5)$$

where the function $v_{2|1}(i_1, \mathbf{h}, \mathbf{x})$ is given by (see the proof of lemma 6.44 on page 111):

$$v_{2|1}(i_1, \mathbf{h}, \mathbf{x}) = S(i_1, \mathbf{h}, \mathbf{x}) - \mu_{2|1}(i_1, \mathbf{h}, \mathbf{x})^2. \quad (7.6)$$

¹In practice we use $\gamma = 5$ and a neighborhood of size 19×19 .

The function $S(i_1, \mathbf{h}, \mathbf{x})$ is given by

$$\begin{aligned} S(i_1, \mathbf{h}, \mathbf{x}_0) &= \frac{1}{p(i_1, \mathbf{x}_0)} \int_{\mathbb{R}} i_2^2 P_{\mathbf{h}}(i) di_2 \\ &= \beta + \frac{\int_{\Omega} g_{\beta}(i_1 - I_1^{\sigma}(\mathbf{x})) I_2^{\sigma}(\mathbf{x} + \mathbf{h}(\mathbf{x}))^2 G_{\gamma}(\mathbf{x} - \mathbf{x}_0) d\mathbf{x}}{\int_{\Omega} g_{\beta}(i_1 - I_1^{\sigma}(\mathbf{x})) G_{\gamma}(\mathbf{x} - \mathbf{x}_0) d\mathbf{x}}. \end{aligned} \quad (7.7)$$

With these approximations, we define

$$\begin{aligned} \tilde{L}_{\text{CR},\mathbf{h}}^l(i, \mathbf{x}) &= \frac{1}{\frac{1}{2} \mathcal{G}_{\gamma}(\mathbf{x}) v_2(\mathbf{h}, \mathbf{x})} \\ &\quad \left(\mu_2(\mathbf{h}, \mathbf{x}) - \mu_{2|1}(i_1, \mathbf{h}, \mathbf{x}) + \theta_1(i_1, \mathbf{h}, \mathbf{x}) (i_2 - \mu_2(\mathbf{h}, \mathbf{x})) \right), \end{aligned} \quad (7.8)$$

and

$$\tilde{F}_{\text{CR}}^l(\mathbf{h}) = \tilde{L}_{\text{CR},\mathbf{h}}^l(I_1^{\sigma}, I_2^{\sigma}(\mathbf{Id} + \mathbf{h}), \mathbf{Id}) \nabla I_2^{\sigma}(\mathbf{Id} + \mathbf{h}).$$

Again, we notice that the Lipschitz-continuity of $\tilde{F}_{\text{CR}}^l(\mathbf{h})$ is preserved. We also have in this case a non-stationary filtering process. We therefore propose the following algorithm.

Algorithm 7.6 ($\tilde{F}_{\text{CR}}^l(\mathbf{h})$)

- Estimate for each \mathbf{y} on the discrete grid the value of $\mathcal{G}_{\gamma}(\mathbf{y})$ using equation (4.13) on page 68, of $\mu_2(\mathbf{h}, \mathbf{y})$ and $v_2(\mathbf{h}, \mathbf{y})$ using equations (6.22), (6.23) and (6.24) on page 110, of $S(I_1^{\sigma}(\mathbf{y}), \mathbf{h}, \mathbf{y})$ using equation (7.7), and of $\mu_{2|1}(I_1^{\sigma}(\mathbf{y}), \mathbf{h}, \mathbf{y})$ using equation (7.4), approximating the integrals by a finite sum on a sufficiently large² neighborhood around \mathbf{y} .
- Estimate $v_{2|1}(I_1^{\sigma}(\mathbf{y}), \mathbf{h}, \mathbf{y})$ using equation (7.6), $\theta_1(I_1^{\sigma}(\mathbf{y}), \mathbf{h}, \mathbf{y})$ using equation (7.5), and $\tilde{L}_{\text{CR},\mathbf{h}}^l(I_1^{\sigma}(\mathbf{y}), I_2^{\sigma}(\mathbf{y} + \mathbf{h}(\mathbf{y})), \mathbf{y})$ using equation (7.8).
- Estimate $\tilde{F}_{\text{CR}}^l(\mathbf{h})(\mathbf{y}) = \tilde{L}_{\text{CR},\mathbf{h}}^l(I_1^{\sigma}(\mathbf{y}), I_2^{\sigma}(\mathbf{y} + \mathbf{h}(\mathbf{y})), \mathbf{y}) \nabla I_2^{\sigma}(\mathbf{y} + \mathbf{h}(\mathbf{y}))$.

7.3.3 Parallel Implementation

Algorithms 7.5 and 7.6 are very well adapted to parallelization, as they both describe a non-stationary filtering procedure. The operations required for computing $\tilde{F}_{\text{MI}}^l(\mathbf{h})$ and $\tilde{F}_{\text{CR}}^l(\mathbf{h})$ at each voxel are defined from the knowledge of the functions I_1^{σ} , I_2^{σ} and \mathbf{h} in a relatively large neighborhood around it.

²In practice we use $\gamma = 5$ and a neighborhood of size 19×19 .

We have implemented these two algorithms using the MPI library for parallel execution using a cluster of N_p processors. A master-slave architecture is used, i.e. one of the processors handles a certain amount of global work, and distributes local work to the remaining $N_p - 1$ processors. The distributed work is the computation of $\tilde{F}_{\text{MI}}^l(\mathbf{h})$ and $\tilde{F}_{\text{CR}}^l(\mathbf{h})$ at each iteration. The remaining operations (mainly the computation of the regularization term and the time-step update) are handled by the master processor. The execution flows for the master processor and for a slave processor are illustrated in figures 7.1 and 7.2. It is assumed that all the processors have access to the data corresponding to I_1^σ , I_2^σ and \mathbf{h} at each iteration. This is achieved in practice by special synchronization routines of the MPI library. With this architecture and using a cluster of 24 processors, we have achieved 15 times faster execution times than with the sequential version.

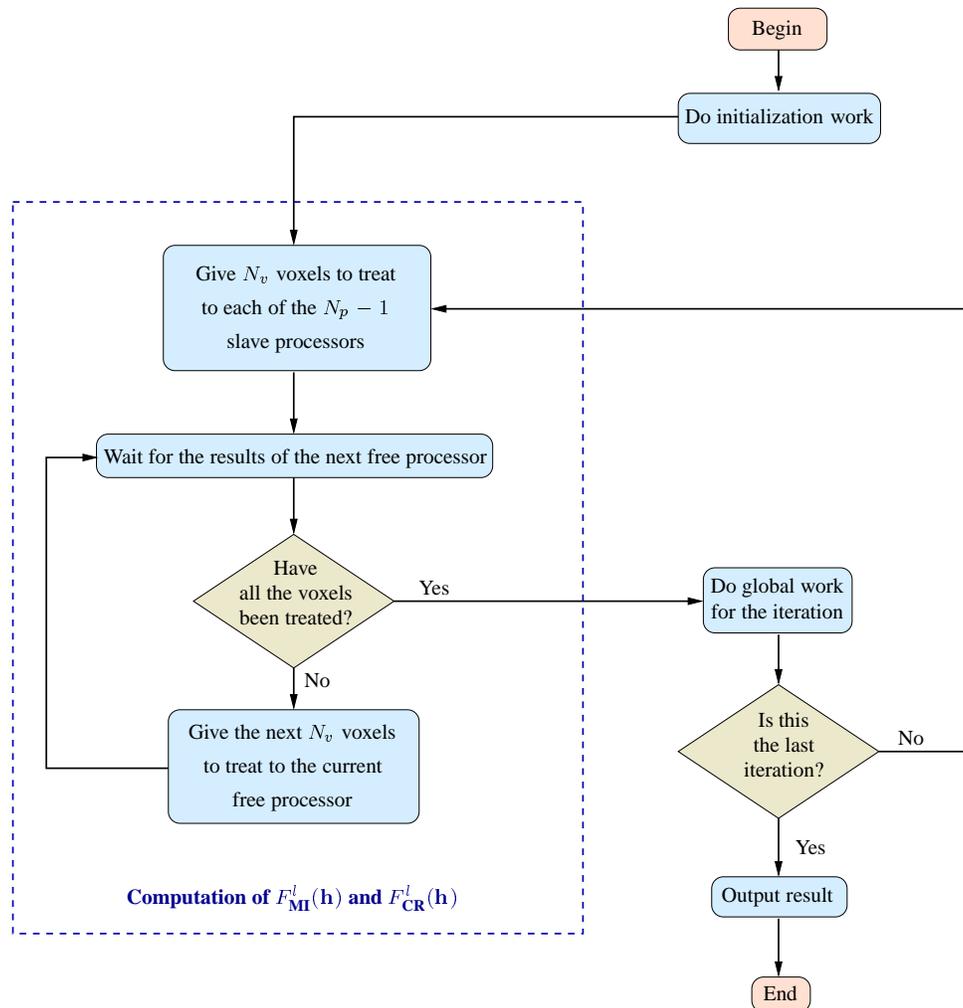


Figure 7.1: Execution flow for the master processor in the parallel implementation of the matching flow.

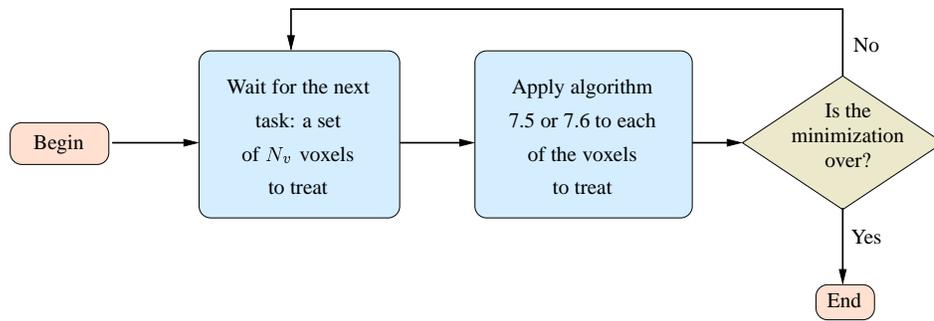


Figure 7.2: Execution flow for each slave processor in the parallel implementation of the matching flow.

Chapter 8

Determining Parameters

This chapter discusses the way in which the different parameters of the algorithms are determined, particularly the smoothing parameter for the Parzen window estimates.

The matching algorithms use the following parameters.

- γ : This parameter is fixed to 5 with a local window size of 19x19 for the mutual information and the correlation ratio. For the cross correlation, the value of this parameter does not affect the computation time. This important property is due to the fact that the local statistics are calculated using the recursive smoothing filter. Thanks to this property, we have conducted some experiments with different values of this parameter, which have shown that the algorithms are not very sensitive to it. This is the reason why we have fixed it for the mutual information and the correlation ratio. Qualitatively speaking, the local window has to be large enough for the statistics to be significant, and small enough to account for non-stationarities of the density. The value chosen has given good results in practice.
- β : This is the smoothing parameter for the Parzen estimates. Unlike the parameter γ , determining a good value for this parameter is crucial for the matching results. The determination of this parameter is discussed in more detail in the next section.
- α : This parameter determines the weight given to the regularization term in the energy functional. Since the range of the different matching functions varies considerably, we replace this parameter by another one, noted C such that

$$\alpha = C \kappa,$$

where κ is given by

$$\kappa = \|F(\mathbf{h}_0)\|^\infty,$$

\mathbf{h}_0 being the initial field and F any of the matching functions.

- σ : This is the scale parameter. We adopt a multi-resolution approach, smoothing the images at each stage by a small amount. Within each stage of the multi-resolution pyramid, the parameter σ is fixed to a small value, typically 0.25 voxels.

Besides these global parameters, one extra parameter is needed for each family of regularization operators.

- ξ : This is the parameter controlling the behavior of the linearized elasticity operator. For ξ close to 1, the Laplacian operator becomes dominant, while the operator $\nabla(\operatorname{div}(\mathbf{h}))$ becomes dominant for ξ close to zero. Two experiments in the next chapter show qualitatively the behavior of the elasticity operator with respect to that of the Laplacian. In practice, we fix the value of ξ to 0.5, giving thus the same weight to both operators.
- λ : This is the parameter controlling the anisotropic behavior of the Nagel-Enkelmann tensor. We adopt the method proposed by Alvarez et al. [6]. Given s , which in practice is fixed to 0.1, we take the value of λ such that

$$s = \int_0^\lambda \mathcal{H}_{|\nabla I_1^\sigma|}(z) dz$$

where $\mathcal{H}_{|\nabla I_1^\sigma|}(z)$ is the normalized histogram of $|\nabla I_1^\sigma|$.

8.1 Determining the Smoothing Parameter

The parameter β of the gaussian kernel is determined automatically. A very large amount of literature has been published on the problem of determining an adequate value for β (we refer to [17] for a recent comprehensive study on non-parametric density estimation, containing many references to the forementioned literature).

We adopt a cross-validation method technique based on an empirical maximum likelihood method. We note $\{\mathbf{i}_k\}$ a set of m intensity pair samples ($k = 1 \dots m$) and take the value of β which maximizes the empirical likelihood:

$$L(\beta) = \prod_{k=1}^m \hat{P}_{\beta,k}(\mathbf{i}_k)$$

where

$$\hat{P}_{\beta,k}(\mathbf{i}_k) = \frac{1}{m - n_k} \sum_{\{s: \mathbf{i}_s \neq \mathbf{i}_k\}} G_\beta(\mathbf{i}_k - \mathbf{i}_s)$$

and n_k is the number of data samples for which $i_s = i_k$.

We present two examples of the determination of β by maximization of the empirical likelihood. Figures 8.1 and 8.2 show the value of the empirical likelihood of the estimated density as a function of the parameter β for two different images. Note how different the optimal values are for these two examples.

The parameter β for the joint intensity function is taken as $\max(\beta_1, \beta_2)$, where β_1 (resp. β_2) is the optimal value obtained by maximization of the empirical likelihood for I_1^σ (resp. I_2^σ).

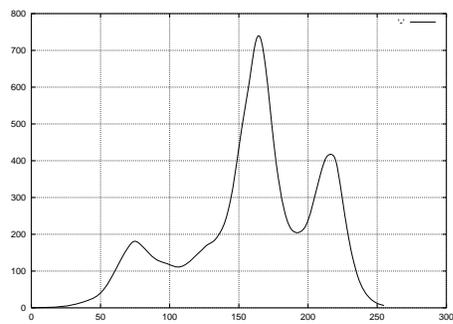
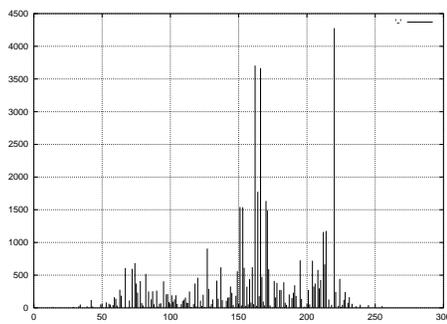
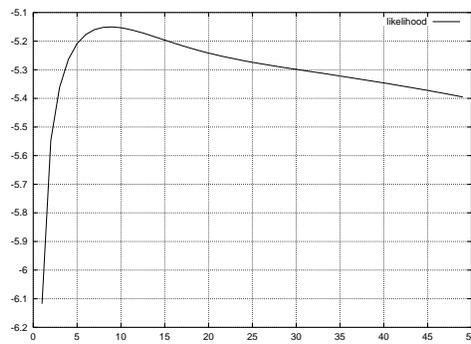


Figure 8.1: Density estimation: example of using maximization of the empirical likelihood. The curve in the second row shows the value of the empirical likelihood of the estimated density, as a function of the parameter β which attains a maximum for $\beta \simeq 8$. The bottom row presents the raw histogram of the image on the left, and its smoothed version with the optimal value of β on the right.

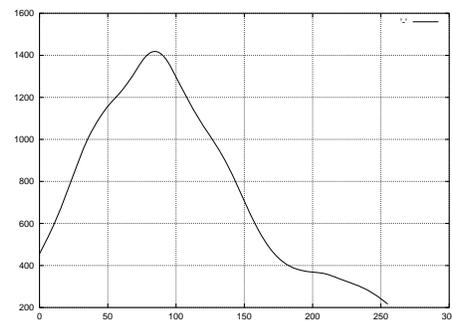
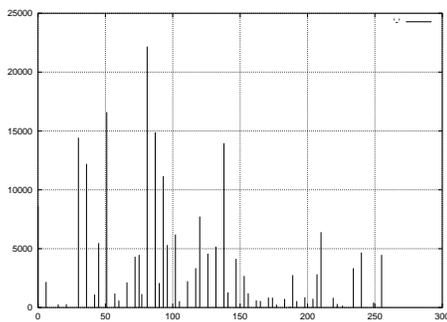
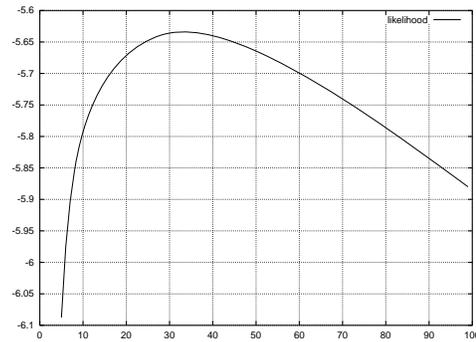


Figure 8.2: Density estimation: second example using maximization of the empirical likelihood. The curve in the second row shows the value of the empirical likelihood as a function of the parameter β , which reaches a maximum for $\beta \simeq 30$. The bottom row presents the raw histogram of the image on the left, and its smoothed version with the optimal value of β on the right.

Chapter 9

Experimental Results

This chapter presents experimental results for all the described algorithms using both real and synthetic data. Examples include 2D images for applications in computer vision and 3D images concerning different medical image modalities.

9.1 Classification

We present a total of eight experiments. The six dissimilarity criteria and the two regularization operators are tested. Both 2D and 3D problems are included. Some of the experiments are completely synthetic, some others are completely real and yet some others are partly synthetic. Table 9.1 on the following page summarizes how each of these categories are represented in the eight experiments shown. We use the following abbreviations to refer to the three global criteria: GMI, GCR, GCC, the three local criteria: LMI, LCR, LCC and the two regularization operators: LE (linearized elasticity) and AD (anisotropic diffusion using the Nagel-Enkelmann tensor).

9.2 Description of the Experiments

	Similarity measure used	: GMI.
	Intensity transformation	: known.
	Geometric transformation	: unknown.
	Regularization used	: LE, AD.
Experiment 9.1	Parameters	: $\alpha = 10$, number of scales = 3.
	Computation time	: $\simeq 3$ minutes.
	Matching program	: MatchPDE.
	Related figures	: 9.1 – 9.4.

Experiment:	9.1	9.2	9.3	9.4	9.5	9.6	9.7	9.8
2D	•	•	•				•	•
3D				•	•	•		
Known geometric transformation		•	•					
Known intensity transformation	•		•				•	
Global mutual information	•	•			•			
Global correlation ratio				•				•
Global cross correlation		•						•
Local mutual information			•					
Local correlation ratio			•					
Local cross correlation						•	•	
Linearized elasticity	•	•	•	•	•	•		
Anisotropic diffusion	•	•					•	•

Table 9.1: Summary of the characteristics of each of the experiments.

Comments:

This experiment shows the behavior of the two different families of regularization operators. Figure 9.1 shows on the first row the images I_1 (on the left) and I_2 (on the right), and on the second row the image $I_2 \circ (\mathbf{Id} + \mathbf{h}^*)$, where \mathbf{h}^* is the displacement field obtained with linearized elasticity (on the left) and anisotropic diffusion (on the right). The displacement fields are shown in figure 9.2. Figure 9.3 shows the result obtained with the linearized elasticity operator with a value of ξ close to $\frac{1}{2}$ on the left, and close to 1 on the right. Finally, 9.4 shows the determinant of the Jacobian of $(\mathbf{Id} + \mathbf{h}^*)$, where \mathbf{h}^* is the field of figure 9.3 on the left. The interest in this function is that if it is everywhere positive, then the transformation function: $\mathbf{x} \rightarrow \mathbf{Id} + \mathbf{h}^*(\mathbf{x})$ is invertible. This is the case for all the displacement fields shown in this experiment. This experiment shows

	Similarity measure used	: GCR.
	Intensity transformation	: unknown.
	Geometric transformation	: unknown.
Experiment 9.2	Regularization	: LE.
	Parameters	: $\alpha = 20$, number of scales: 3.
	Computation time	: $\simeq 10$ minutes.
	Matching program	: MatchPDE.
	Related figures	: 9.5–9.8.

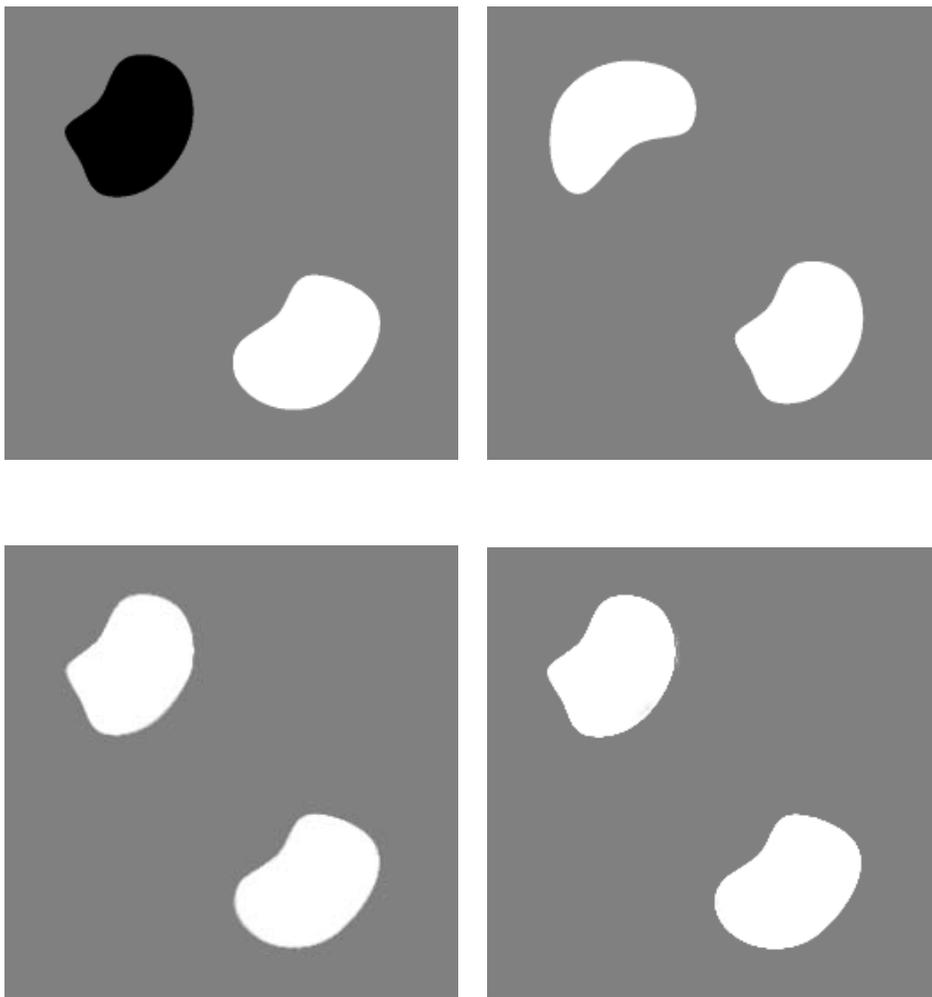


Figure 9.1: Experiment showing the behavior of the two regularization operators. See explanation of experiment 9.1 on page 143.

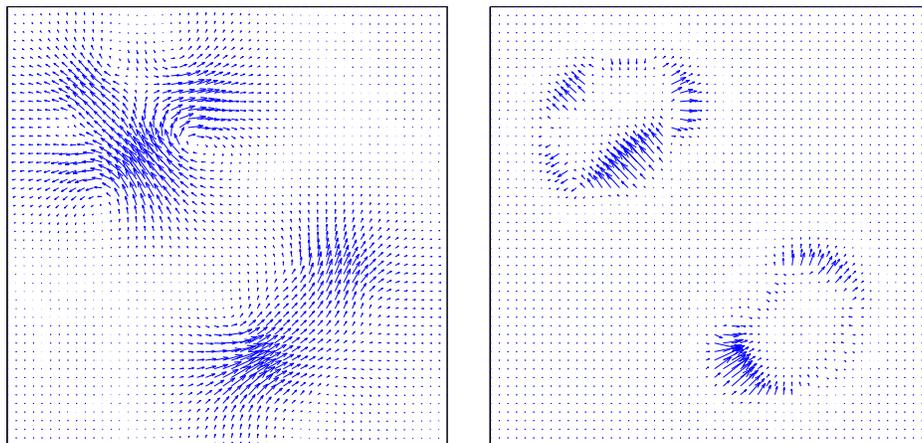


Figure 9.2: Displacement field obtained with linearized elasticity (left) and anisotropic diffusion (right).

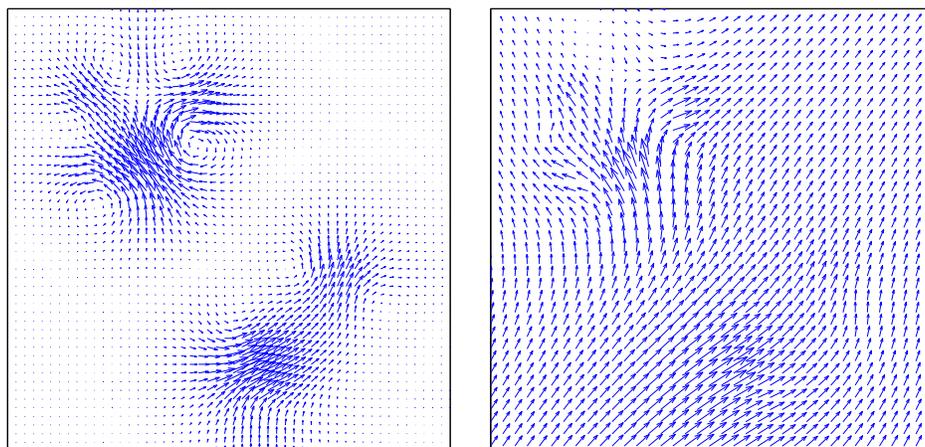


Figure 9.3: Displacement field obtained with linearized elasticity with a value of ξ close to $\frac{1}{2}$ (left), and close to one (right).

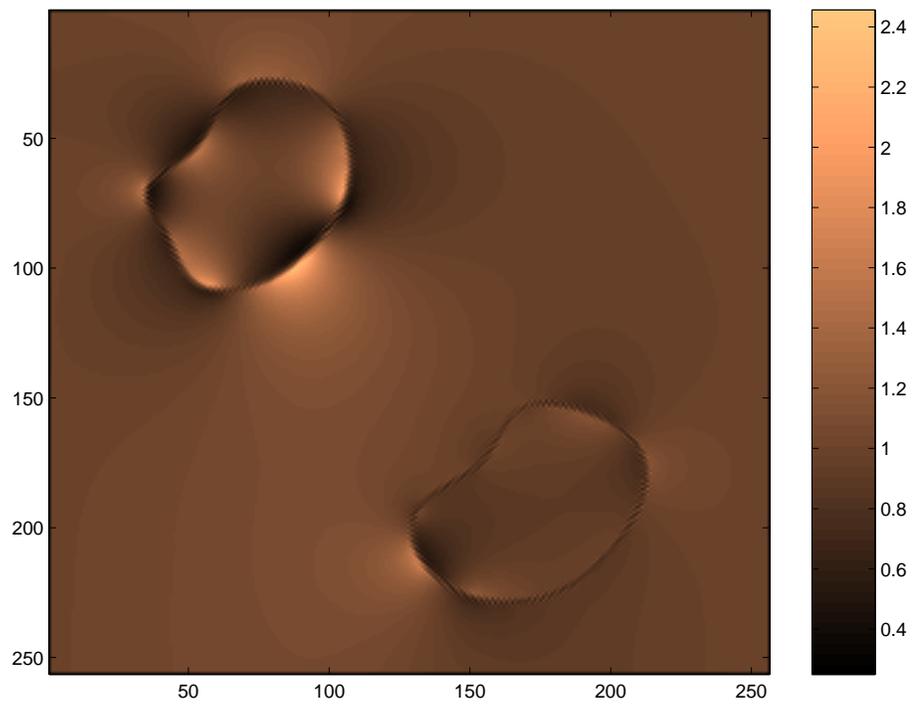


Figure 9.4: Determinant of the Jacobian of $(\text{Id} + \mathbf{h}^*)$, for the displacement field of figure 9.3 on the left.

Comments:

This experiment shows matching of a 2D plane extracted from a 3D proton-density image (PD) a similar T2-weighted 2D plane. An artificial warp was applied to the T2-weighted 2D plane. The deformation is well recovered using global correlation ratio.

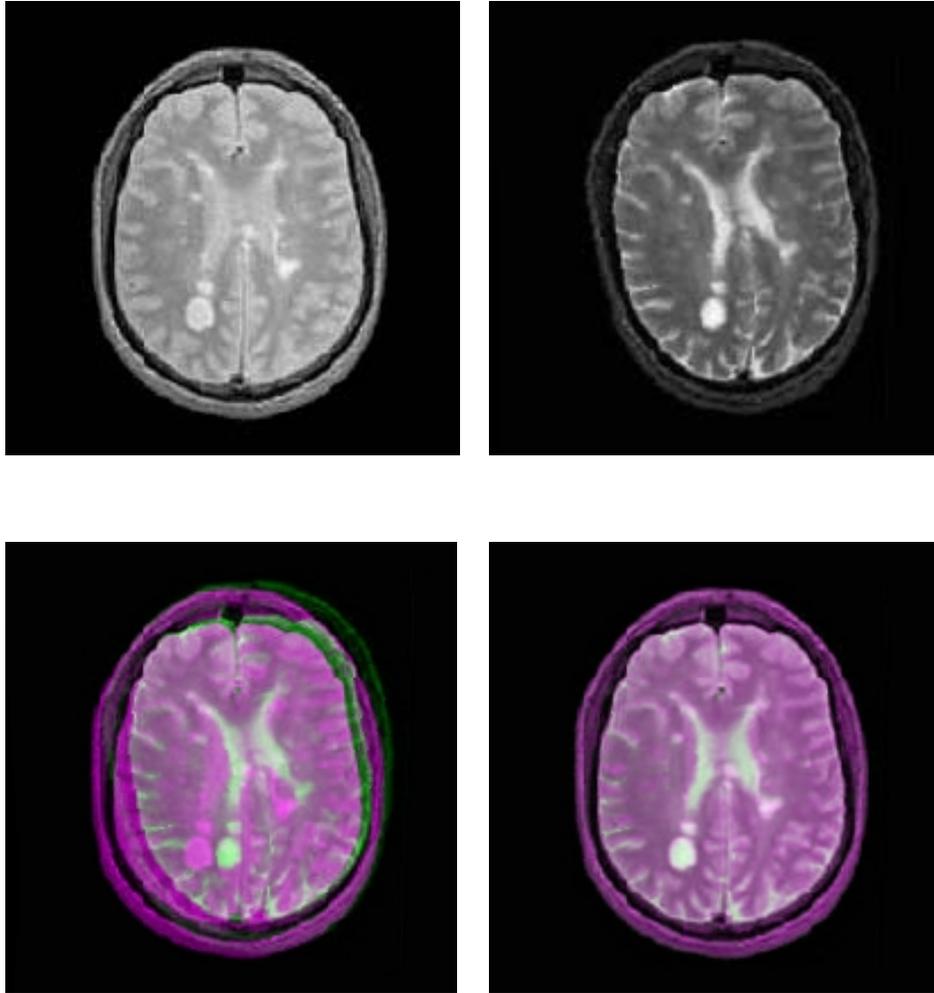


Figure 9.5: Proton density image matching against T2-weighted MRI.

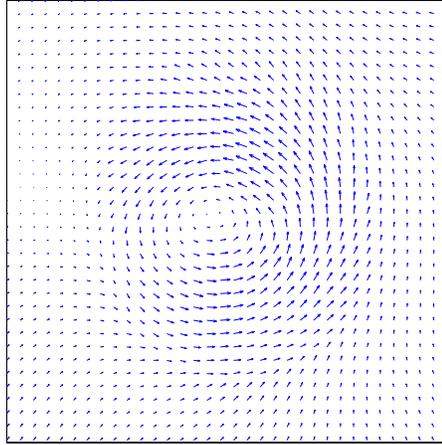


Figure 9.6: Deformation field recovered in the experiment of figure 9.5.

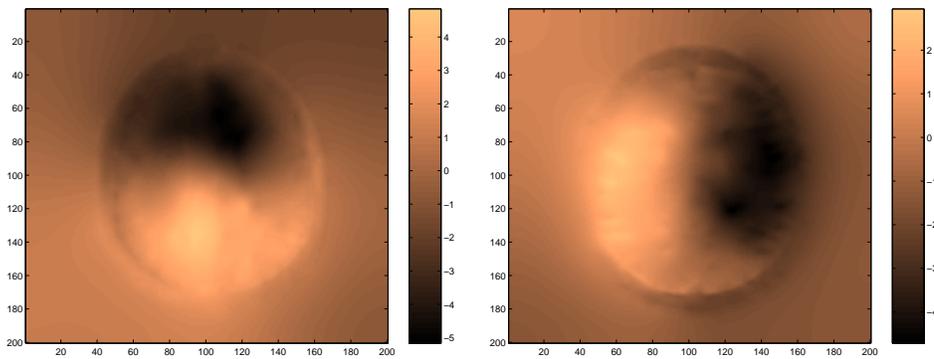


Figure 9.7: Components of the deformation field recovered in the experiment of figure 9.5.

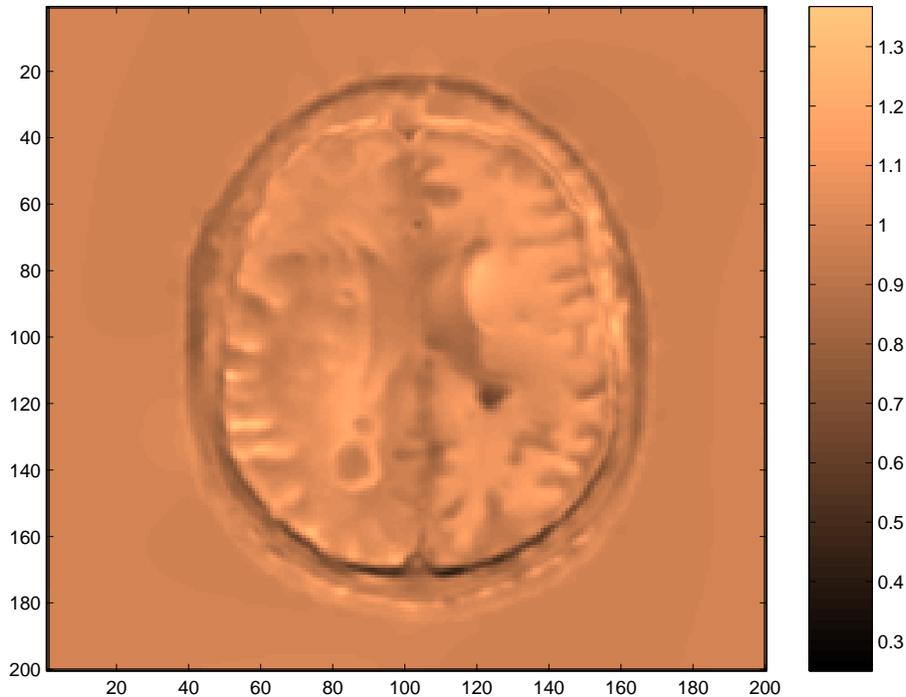


Figure 9.8: Determinant of the Jacobian of the deformation field recovered in the experiment of figure 9.5.

Experiment 9.3	Similarity measure used	: LMI, LCR.
	Intensity transformation	: known.
	Geometric transformation	: known.
	Regularization	: LE.
	Parameters	: $\alpha = 10$, number of scales: 2.
	Computation time	: $\simeq 30$ minutes (12 processors).
	Matching program	: mpi9pde
	Related figures	: 9.9–9.13.

Comments:

This experiment shows the result of the local mutual information and local correlation ratio on synthetic data. The reference and target image were both taken from the same 2D plane in a MRI data volume. The reference image J was then transformed in the following way ($|\Omega|_x$ is the size of the domain in the x direction):

$$J'(x, y) = \sin(2\pi J(x, y)) - \cos\left(\frac{2\pi}{|\Omega|} (x + y|\Omega|_x)\right)$$

and then linearly renormalized in $[0, 1]$. Notice that the effect of this manipulation produces a bias in the intensities of the reference image which resembles the real image modality of experiment 9.4, plus a sort of spatial bias. A non-rigid smooth deformation

was then applied to the target image. As expected, the global algorithms failed to align these two images, due to the severe non-stationarity in the intensity distributions.

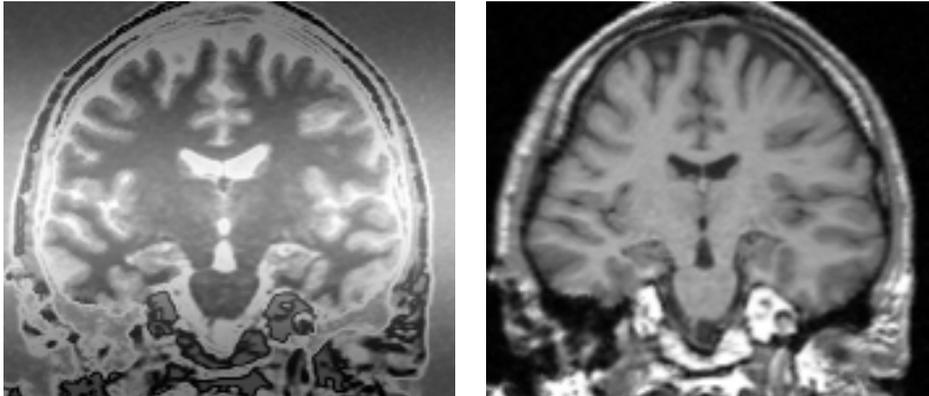


Figure 9.9: Matching with local mutual information and correlation ratio. Reference image (left), deformed image (right).

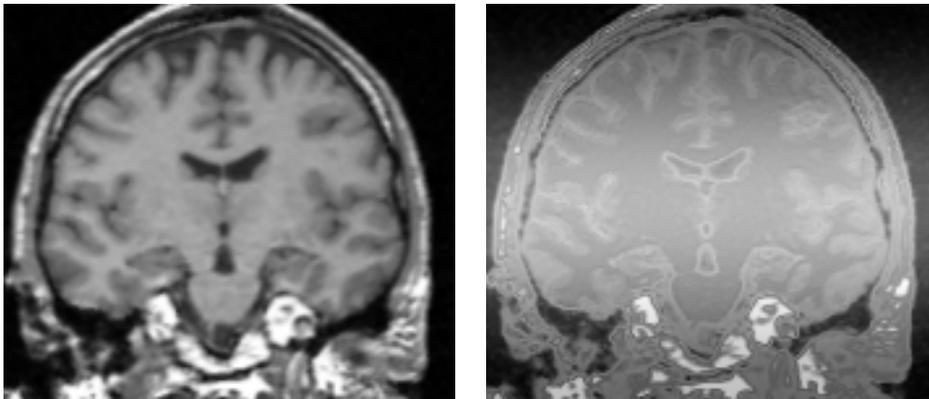


Figure 9.10: Realigned image and its superposition with the reference image in the experiment of figure 9.9.

Experiment 9.4	Similarity measure	: GMI
	Intensity transformation	: unknown
	Geometric transformation	: unknown
	Regularization	: LE.
	Parameters	: $\alpha = 10, N_s = 4$
	Computation time	: 25 minutes
	Command line	: MatchPDE3D
	Related figure	: 9.14.

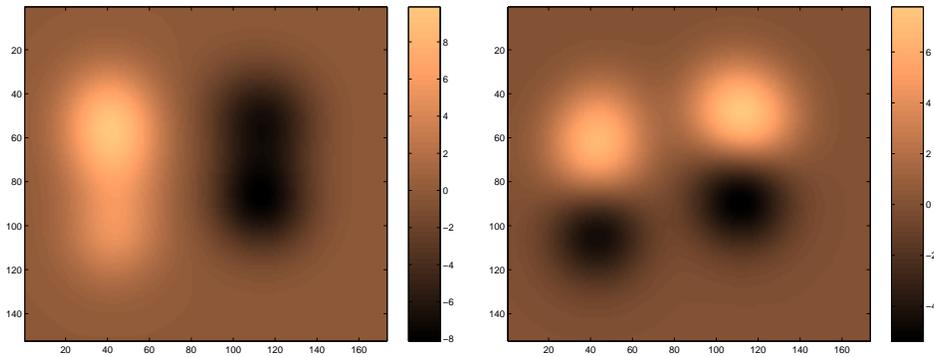


Figure 9.11: Components of the deformation field applied in the experiment of figure 9.9.

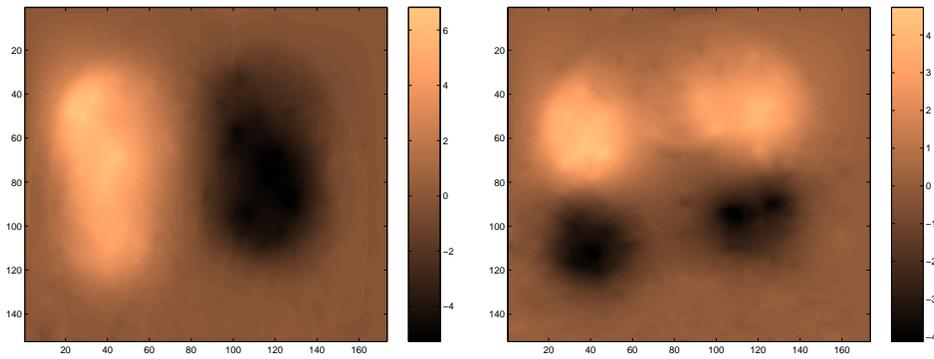


Figure 9.12: Components of the deformation field recovered in the experiment of figure 9.9.

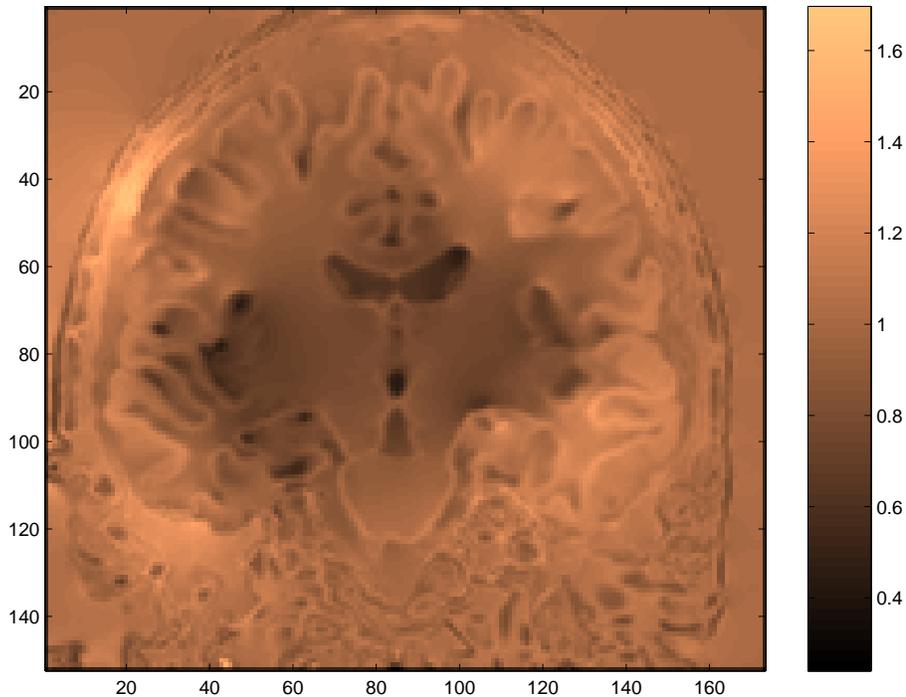


Figure 9.13: Determinant of the Jacobian for the deformation field recovered in the experiment of figure 9.9.

Comments:

This example shows an experiment with real MR data of the brain of a macaque monkey. The reference image is a T1-weighted anatomical volume and the target image is a functional, fMRI contrast MRI (fMRI). The contrast in this modality is related to blood oxygenation level. The figure shows the result of the global algorithms. Notice that the alignment of main axis of the volume has been corrected.

	Similarity measure	: LCC
	Intensity transformation	: unknown
	Geometric transformation	: unknown
	Regularization	: LE.
Experiment 9.5	Parameters	: $\alpha = 10$, number of scales: 4.
	Computation time	: 25 minutes
	Command line	: MatchPDE3D
	Related figure	: 9.15.

Comments:

Matching of T2-weighted anatomical MRI against EPI functional MRI, using local cross correlation. We obtain deformations of an amplitude up to 5 voxels, mostly in

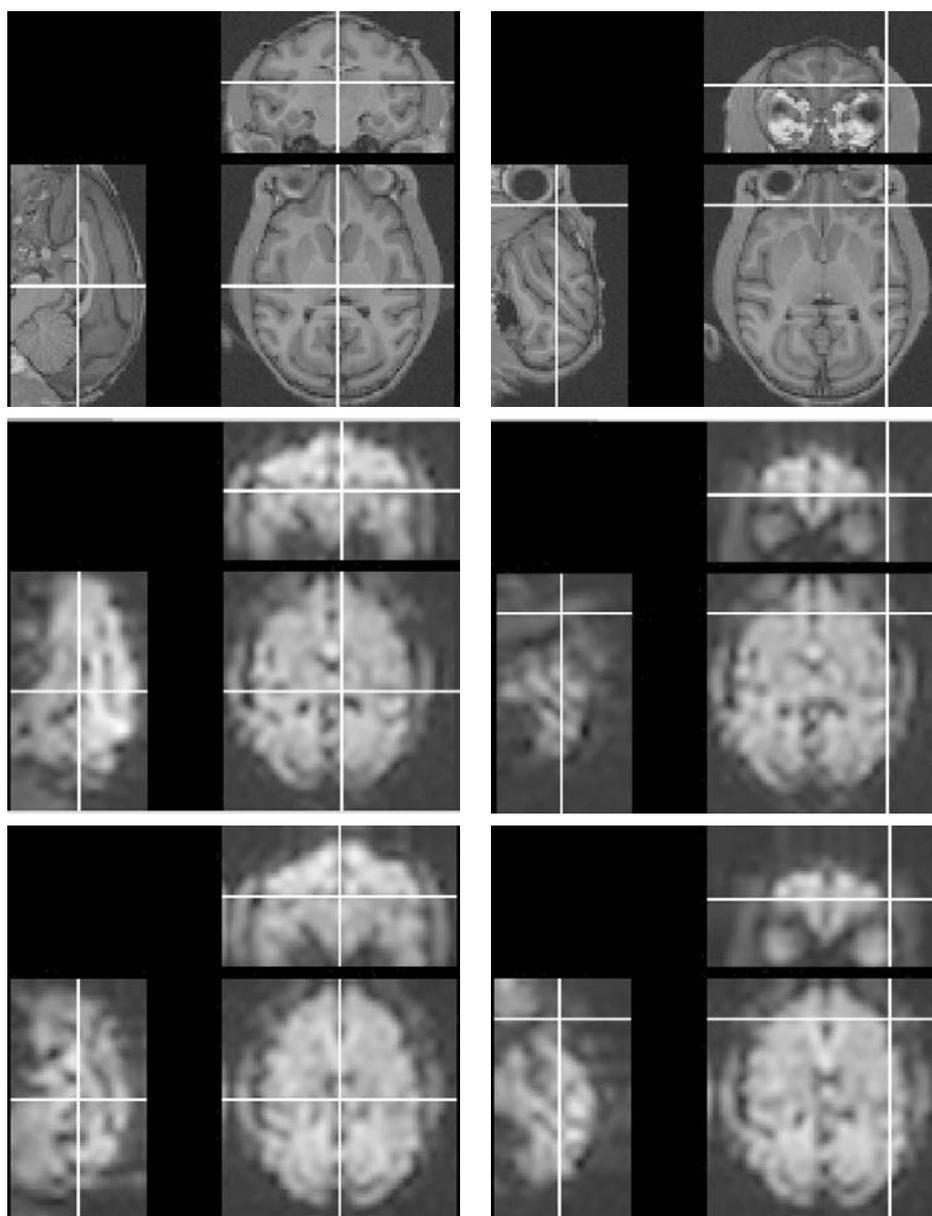


Figure 9.14: Global mutual information with fMRI data. Top row: reference anatomical MRI. Middle row: initial fMRI volume. Bottom row: final (corrected) fMRI volume. The two columns show two different points in the volumes.

the phase-encoding direction (horizontal axis of the lower-right view). This solution seems consistent with previous results obtained by Kybic, Thévenaz et al. [49] on the same dataset.

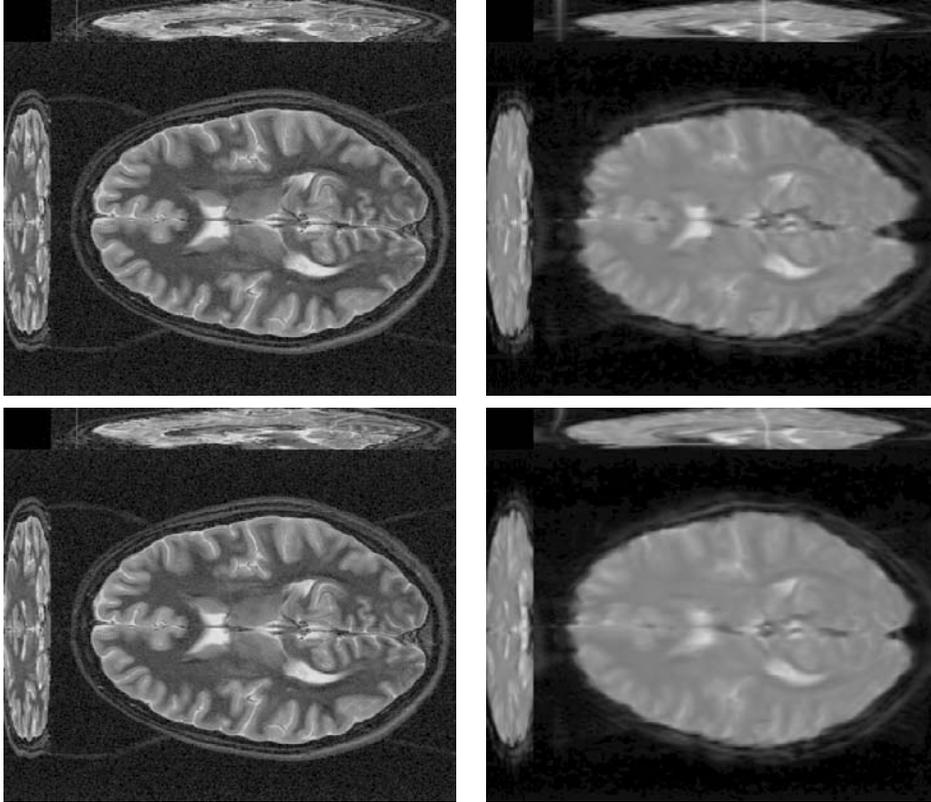


Figure 9.15: Matching of T2-weighted anatomical MRI against EPI functional MRI, using local cross correlation (images are courtesy of Jan Kybic, data provided by Arto Nirkko, Inselspital Bern). Top row: reference and deformed image. Bottom row : reference and realigned image.

Experiment 9.6	Similarity measure	: LMI, LCR
	Intensity transformation	: unknown
	Geometric transformation	: unknown
	Regularization	: LE.
	Parameters	: $\alpha = 10$, number of scales: 1.
	Computation time	: 30 minutes (24 processors).
	Command line	: mpi9pde.
	Related figure	: 9.16.

Comments:

Our next experiment shows the result of the local algorithms with real 3D MR data. The reference image is a T1 weighted anatomical MRI of a human brain. The target image is an MRI from the same patient which is acquired using a special magnetic field gradient as part of the process of obtaining an image of the water diffusion tensor at each point. Notice that the intensities in this modality are qualitatively close to our simulated experiment. The estimated deformation field has a dominant y component, a property which is physically coherent with the applied gradient. Both **MI** and **CR** yielded similar results in this case.

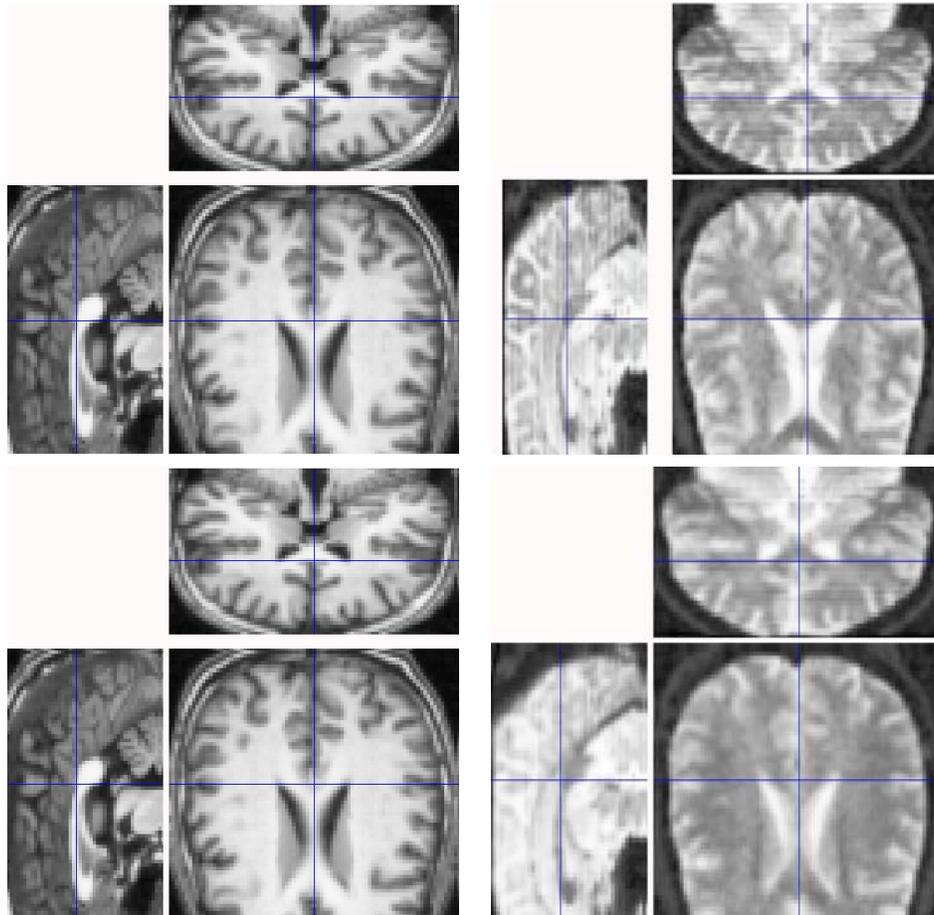


Figure 9.16: Matching of anatomical vs diffusion-tensor-related MRI, using local mutual information and correlation ratio.

	Similarity measure	: GMI
	Intensity transformation	: known.
	Geometric transformation	: unknown.
Experiment 9.7	Regularization	: LE, AD.
	Parameters	: $\alpha = 10$, number of scales: 5.
	Computation time	: 25 minutes.
	Command line	: MatchPDE.
	Related figures	: 9.17–9.20.

Comments:

This experiment shows a real stereo pair in which the intensities in one of the images were artificially transformed using a sine function. The matching is performed using global mutual information.



Figure 9.17: Sterao matching using global mutual information.

	Similarity measure	: LCC, GCC.
	Intensity transformation	: unknown.
	Geometric transformation	: unknown.
Experiment 9.8	Regularization	: LE, AD.
	Parameters	: $\alpha = 10$, number of scales: 3.
	Computation time	: 3 minutes.
	Command line	: MatchPDE
	Related figures	: 9.21, 9.25.

Comments:

This last experiment shows the use of the global and local cross correlation r criteria to perform template matching of human faces. In this case the illuminating conditions



Figure 9.18: Deformed image with the displacement field obtained, and its superposition with the reference image.

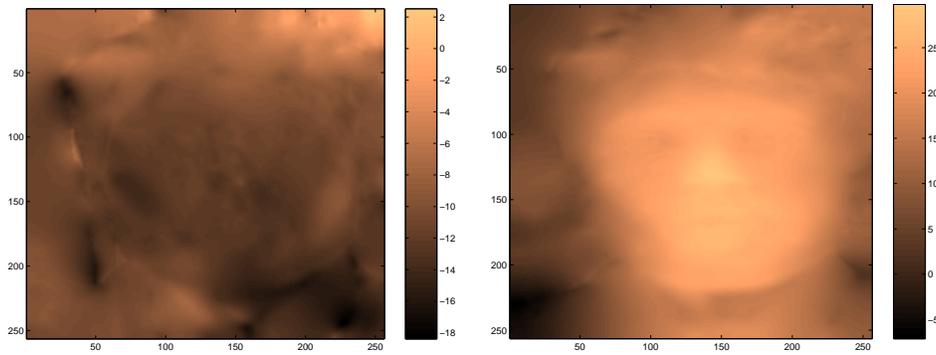


Figure 9.19: Components of the obtained deformation field in the experiment of figure 9.17.

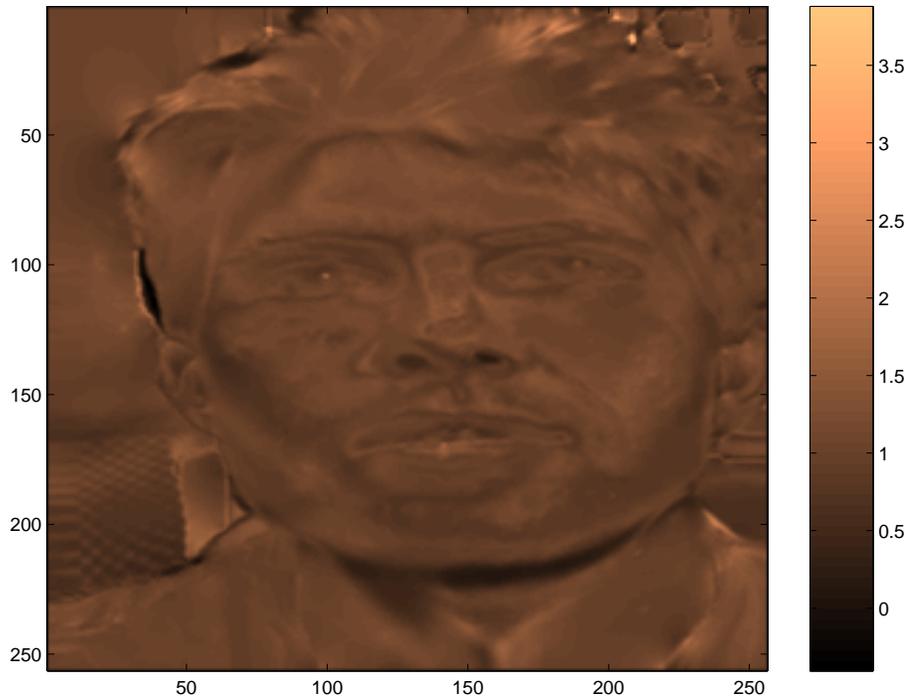


Figure 9.20: Determinant of the Jacobian for the obtained deformation field in the experiment of figure 9.17.

are the same in both photographs. If different, the local algorithms should be used. The different albedos of the two skins create a “multimodal” situation and the transformation is truly non rigid due to the different shapes of the noses and mouths. Notice the excellent matching of the different features. This result was obtained completely automatically with the same sets of parameters as the rest of the experiments, using global mutual information.

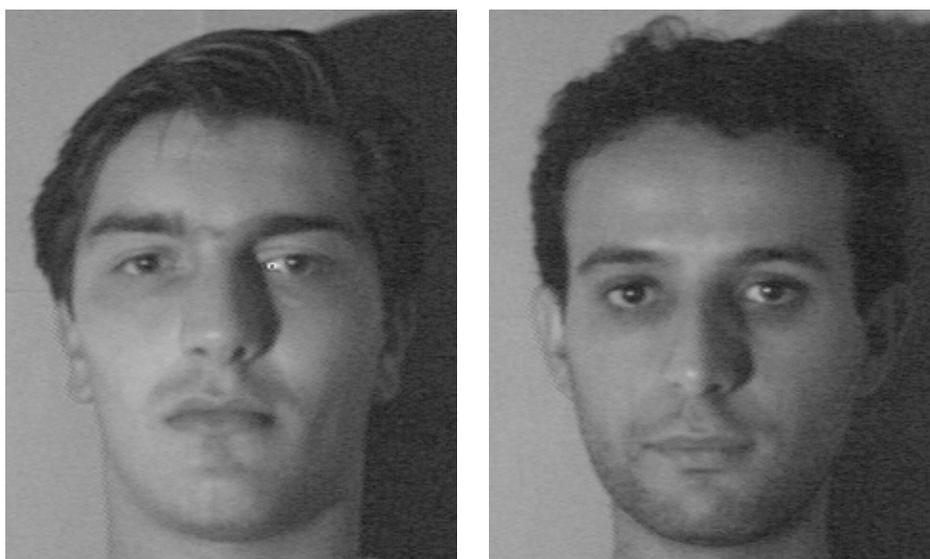


Figure 9.21: Human template matching. Reference (left) and target (right) images.

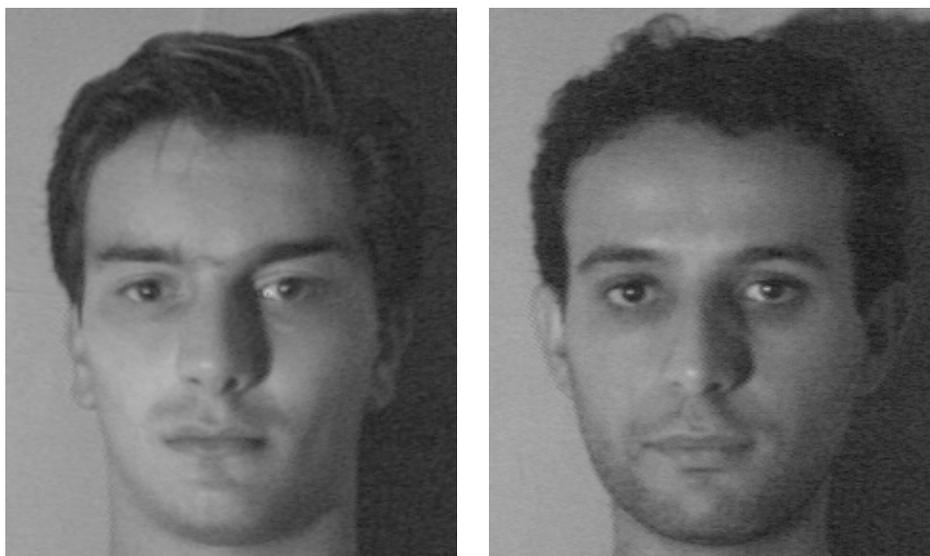


Figure 9.22: Reference image and deformed target image using the obtained deformation field in the experiment of figure 9.21.

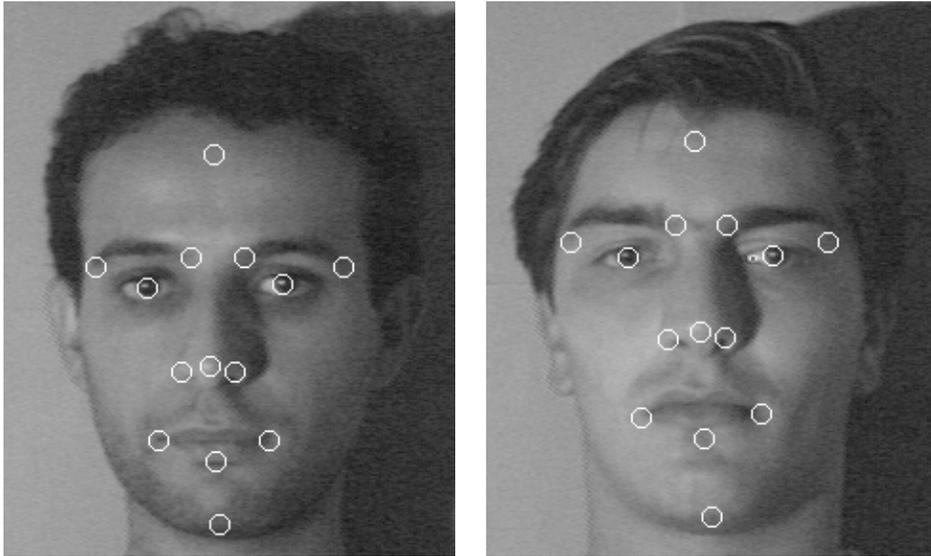


Figure 9.23: Some corresponding points using the obtained deformation field in the experiment of figure 9.21.

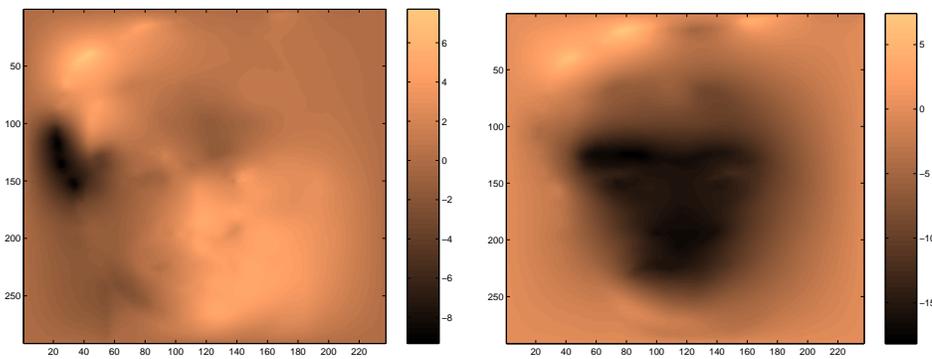


Figure 9.24: Components of the obtained deformation field in the experiment of figure 9.21.

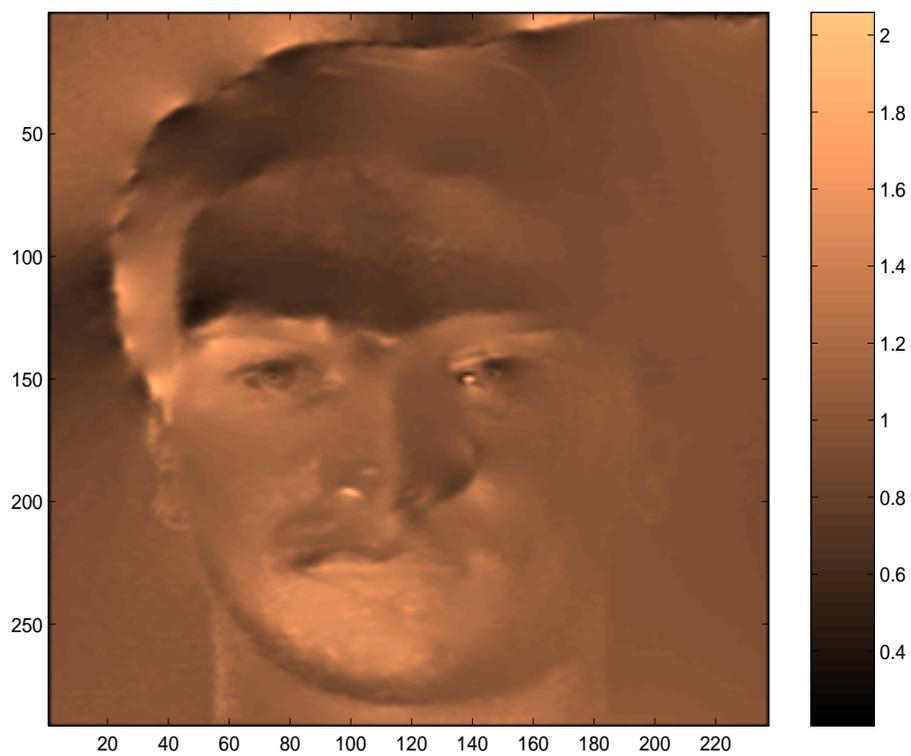


Figure 9.25: Determinant of the Jacobian for the obtained deformation field in the experiment of figure 9.21.

Appendix A

Other Applications

The gradients of the statistical similarity measures studied in the previous chapters may be used in other contexts where the same criteria are applicable. In this appendix we focus on two such applications, namely image segmentation by entropy minimization (Section A.1) and diffeomorphic matching (Section A.2). The goal is more to give the flavour of how the gradients are used in these contexts than to propose specific methods.

A.1 Entropy Minimization for Image Segmentation

The minimization of the entropy associated to an image yields an algorithm which can be viewed as a mean-shift process [29], useful for segmentation purposes. Although we restrict the discussion to global entropy in order to keep it as simple as possible, it can be generalized to the local case in a straightforward manner, yielding an algorithm which is very close to bilateral filtering [82]. The experimental results are shown using the local version of the energy.

Given an image (we restrict to the scalar case) $I : \Omega \rightarrow \mathbb{R}$, we may associate it a random variable X_I , whose values are noted i and whose samples are given by the values of $I(\mathbf{x})$, for $\mathbf{x} \in \Omega$. The probability density of X_I may be estimated by

$$P(i) = \frac{1}{|\Omega|} \int_{\Omega} G_{\beta}(I(\mathbf{x}) - i) di,$$

this definition being in agreement with the usual property

$$\int_{\mathbb{R}} P(i) di = \frac{1}{|\Omega|} \int_{\Omega} \underbrace{\int_{\mathbb{R}} G_{\beta}(I(\mathbf{x}) - i) di}_{1} d\mathbf{x} = 1.$$

The entropy of X_I is given by

$$\text{Ent}(X_I) = - \int_{\mathbb{R}} P(i) \log(P(i)) di.$$

This quantity is now viewed as a functional of I , and the segmentation problem may be defined as finding the function $I^* : \Omega \rightarrow \mathbb{R}$ satisfying

$$I^* = \arg \min_{I \in \mathcal{F}} \text{Ent}(I).$$

Clearly the global minimum is attained for a constant image, but rather than adding a data-attachment term, we use the non-convexity of this energy to find the closest local minimum, starting with the initial image I_0 as first estimate. Computing the first variation of $\text{Ent}(I)$, we have

$$\begin{aligned} \left. \frac{\partial \text{Ent}(I + \epsilon J)}{\partial \epsilon} \right|_{\epsilon=0} &= \int_{\mathbb{R}} (1 + \log(P(i))) \int_{\Omega} G'_{\beta}(I(\mathbf{x}) - i) J(\mathbf{x}) d\mathbf{x} di \\ &= \int_{\Omega} \underbrace{\left(G_{\beta} \star \frac{P'(i)}{P(i)} \right)}_{\nabla_H(\text{Ent}(I))} J(\mathbf{x}) d\mathbf{x}. \end{aligned}$$

At the steady state, $\nabla_H(\text{Ent}(I)) = 0$ and thus, necessarily, $\frac{P'(i)}{P(i)} = 0$ for all $i \in \mathbb{R}$. Now since we have

$$\frac{P'(i)}{P(i)} = \frac{1}{\beta^2} \left(i - \frac{\int_{\Omega} I(\mathbf{x}) G_{\beta}(I(\mathbf{x}) - i) d\mathbf{x}}{\int_{\Omega} G_{\beta}(I(\mathbf{x}) - i) d\mathbf{x}} \right),$$

we may try to solve the minimization problem by introducing time and a differentiable function $I : \mathbb{R}^+ \times \Omega \rightarrow \mathbb{R}$, and defining the solution as the steady state of the initial value problem

$$\begin{cases} \frac{\partial I}{\partial t}(t, \mathbf{y}) = I(t, \mathbf{y}) - \frac{\int_{\Omega} I(t, \mathbf{x}) G_{\beta}(I(t, \mathbf{x}) - I(t, \mathbf{y})) d\mathbf{x}}{\int_{\Omega} G_{\beta}(I(t, \mathbf{x}) - I(t, \mathbf{y})) d\mathbf{x}}, & \forall \mathbf{y} \in \Omega. \\ I(0, \cdot) = I_0(\cdot). \end{cases}$$

The second term on the right-hand side of the evolution equation is a weighted mean around the image intensity at \mathbf{y} . Using an explicit time discretization with time-step equal to one, the resulting algorithm describes a mean-shift process [29]. Figures A.1, A.2 and A.3 show the results obtained by applying this procedure to two different images. The results are shown with the local version of the algorithm so that two parameters must be specified: σ , controlling the size of the local window, and β , the smoothing parameter of the Parzen window estimate. The results shown are obtained after convergence for the specified parameters.

A.2 Diffeomorphic Matching

The contents of this section is described in more detail in Chef d'Hotel et al. [23]. The gradients of the statistical criteria computed in Chapter 5 may be used in the context of the template matching equations introduced by Christensen [26] and recently generalized by Trouvé [84]. (We also refer to the work of Miller and Younes [58], who

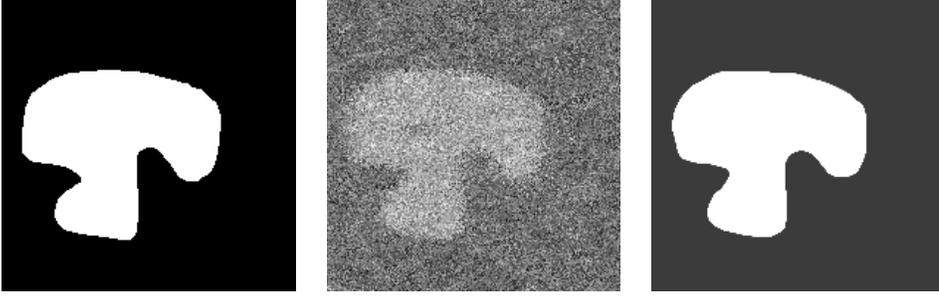


Figure A.1: Segmentation example: From left to right: original image, noisy image, result with $\sigma = 1$, $\beta = 5$.

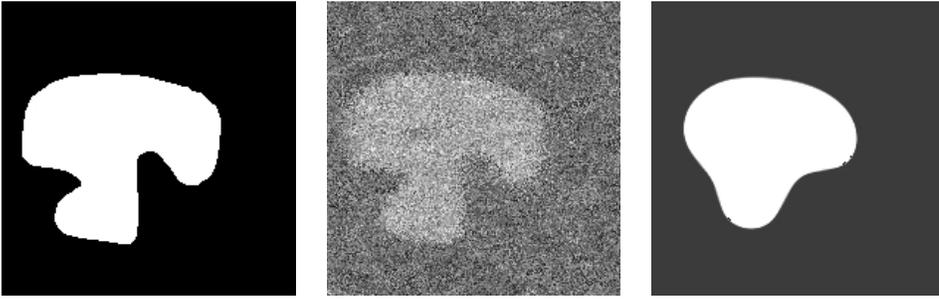


Figure A.2: Segmentation example: From left to right: original image, noisy image, result with $\sigma = 5$, $\beta = 5$.

describe a general framework related to this type of approach). In this context, the unknown is considered to be the transformation $\phi = \text{Id} + \mathbf{h}$, rather than the displacement field \mathbf{h} . The matching problem is then solved by constructing a one-parameter family of diffeomorphisms $\phi(t)$ ($0 \leq t \leq \infty$) and taking $\phi(\infty)$ as the solution. This family of diffeomorphisms is constructed as the solution to the initial value problem

$$\begin{cases} \frac{\partial \phi}{\partial t} = D\phi \cdot v, & \phi(0) = \text{Id}, \\ v(t) = \psi \star F_\phi(t), \end{cases} \quad (\text{A.1})$$

where ψ is a smoothing kernel which ensures the appropriate regularity of the time-dependent vector field v and F_ϕ is the gradient of the similarity criterion, i.e. may be replaced by one of the matching functions that we have studied (Equations (5.9) and (5.14) at the end of Chapter 5). Intuitively, one can construct this family by considering at each iteration k the deformed template $I_2^{\sigma'} \equiv I_2^\sigma \circ \phi^k$ and computing $F_\phi^k|_{\phi=\text{Id}}$ between I_1^σ and $I_2^{\sigma'}$. After the regularization which yields $v^k = \psi \star F_\phi^k$ and for a sufficiently small δt , the transformation $\text{Id} + \delta t v^k$ is a diffeomorphism. Composing ϕ^k (by right-composition) with this transformation gives an efficient scheme which is



Figure A.3: Segmentation example: From left to right: original image, result with $\sigma = 0.5$, $\beta = 5$.

consistent with the continuous flow A.1:

$$\phi^{k+1} = \phi^k \circ (\mathbf{Id} + \delta t v^k).$$

In summary, we have the following algorithm.

1. Set $k = 0$, $\phi^0 = \mathbf{Id}$.
2. Set $I_2^{\sigma k} = I_2^\sigma \circ \phi^k$.
3. Set $b^k = F_{\text{crit}}(I_1^\sigma, I_2^{\sigma k})|_{\mathbf{h}=0}$ for some statistical criterion “crit”.
4. Set $v^k = \psi \star b^k$ (letting for instance ψ be a Gaussian kernel).
5. Set $\phi^{k+1} = \phi^k \circ (\mathbf{Id} + \delta t v^k)$.
6. Set $k = k + 1$ and go to step 2.

Figures A.4 to A.7 show the result of applying this algorithm for two different criteria, namely local mutual information and local cross correlation. The first experiment (Figure A.4) is the same as Experiment 9.3, using local mutual information. As shown in the figure, the applied displacements are well recovered.

Figures A.5, A.6 and A.7 show results obtained with the local cross correlation criterion in conjunction with the algorithm described above. Three different random smooth deformations are applied to an inverse recovery echo planar image of the brain of a macaque monkey. This deformed image is superimposed in green over the anatomical MRI of the same monkey (in magenta). As shown in the figures, large displacements are consistently recovered with this approach.

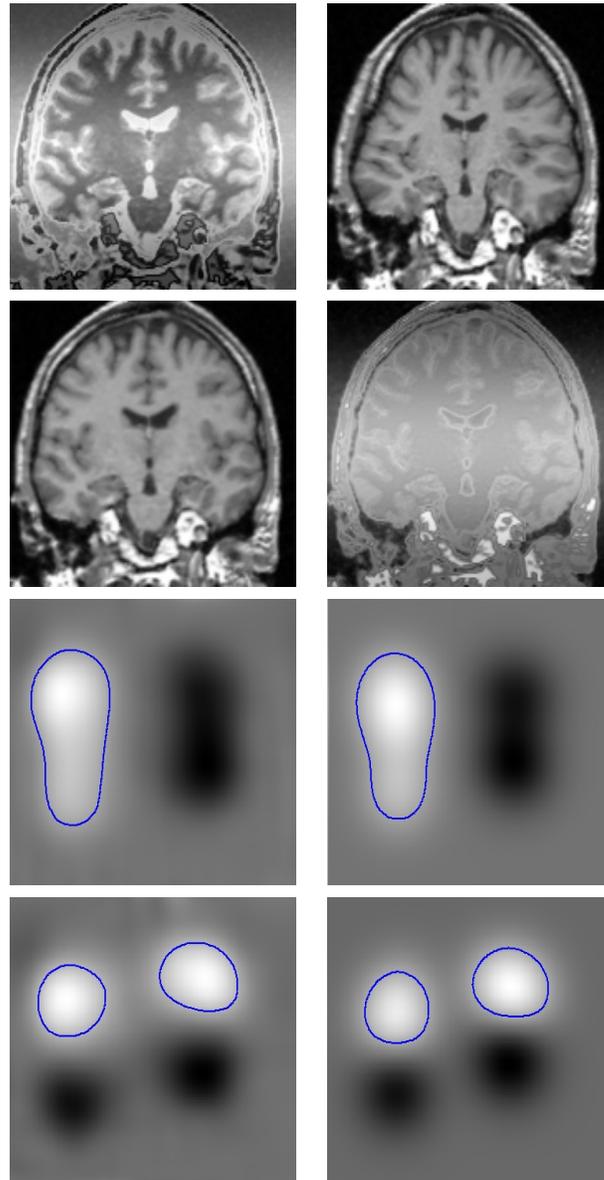


Figure A.4: Diffeomorphic matching using local mutual information. First row: reference and target images. Second row: corrected target image (left) and its superposition with the reference image (right). Third and fourth rows: horizontal and vertical components of the estimated (left) and true (right) deformations fields (iso-level 3.4 is outlined).

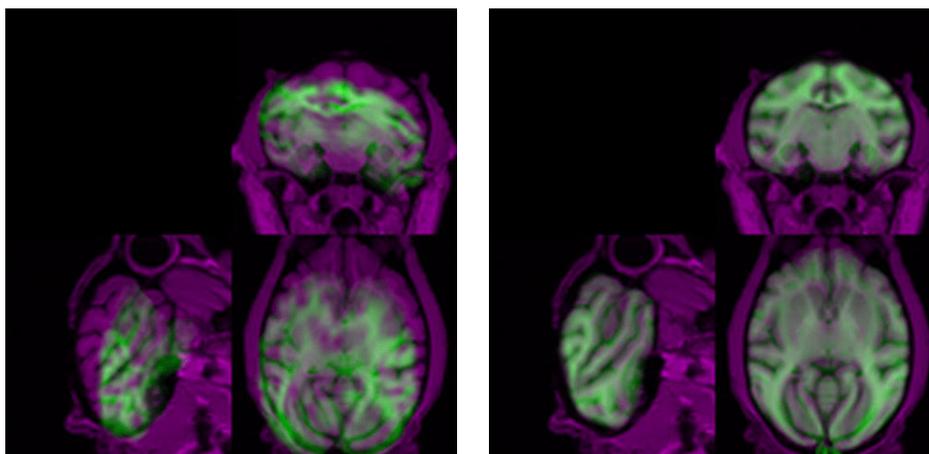


Figure A.5: Diffeomorphic matching using local cross correlation. Left: first random smooth deformation used as initial state. Right: deformed template after convergence of the algorithm.

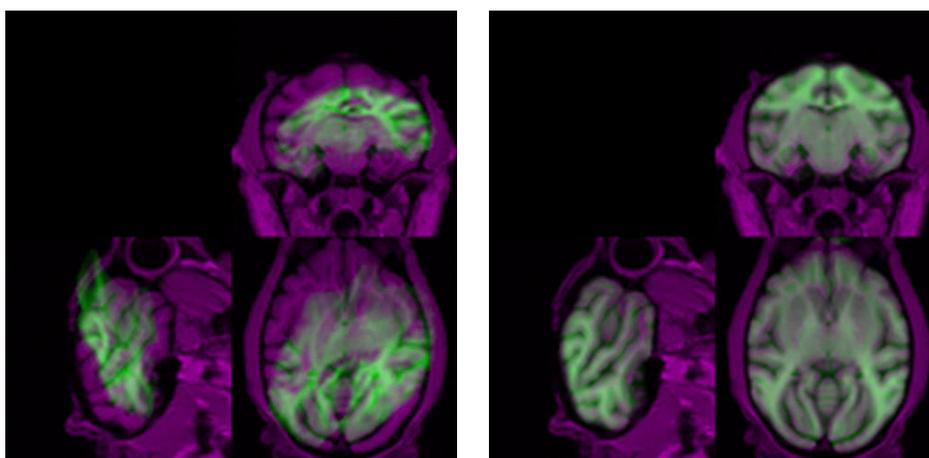


Figure A.6: Diffeomorphic matching using local cross correlation. Left: second random smooth deformation used as initial state. Right: deformed template after convergence of the algorithm.

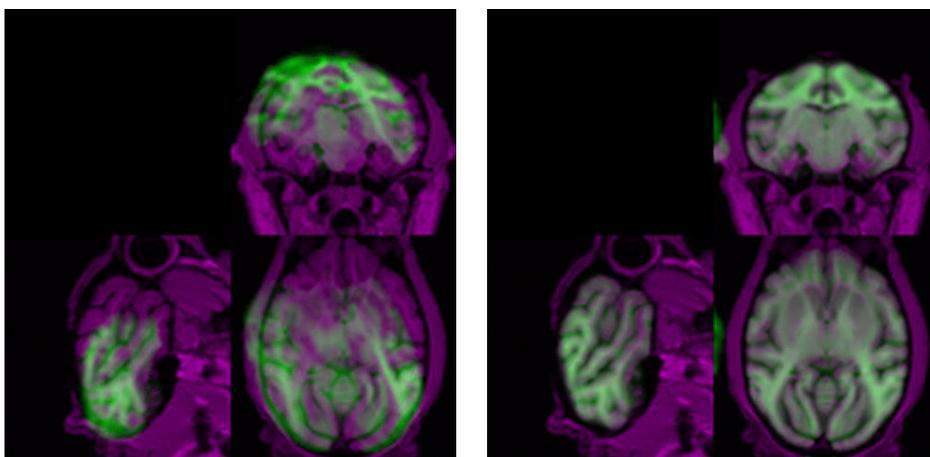


Figure A.7: Diffeomorphic matching using local cross correlation. Left: third random smooth deformation used as initial state. Right: deformed template after convergence of the algorithm.

Appendix B

Library Description

The algorithms described in this work are programmed in C++. They form a part of a complete library providing the basic tools and methods within a coherent framework. This chapter is intended as a description for this library which is composed of about 120 classes and a total of more than 15000 lines of code. The global philosophy of the entire library is to provide tools allowing the user to program a in a few lines a great variety of image processing applications. This is the reason why it is accompanied by a large numbers of small examples in the form of stand-alone executables, which at the same time illustrate the use of the specific classes they use and perform a useful task. The idea is that the same building blocks may be used in other applications, and actually the matching algorithms described in this thesis are high-level blocks using many of these low-level components.

B.1 General Remarks

All the source code is templated to abstract arrays and written in the source header. This presents the advantage of being completely accessible and at the same time guarantees optimum performance through in-lining. The only external library that is needed in X11, the graphics library in unix systems.

There are mainly three different types of containers, adapted to different uses. The first type is just a basic 3D array. The second type, derives from the first type and adds the capability to access any coordinates, so that it actually behaves like an infinite array. The class Image derives from this type of array, basically adding input-output routines and common basic image manipulation routines. The final type of array is an array adopted for PDE, and is actually a graph. The library contains a graphic visualization tool called Xhandler. The numerical schemes are isolated from all the aspects pertaining to the type of container in which the data is stored. They are in the

end the most important part of a PDE code so having the scheme isolated from all the rest presents the advantage of making it completely stand alone.

B.2 C++ Listing: Global Matching Functions

This section contains the listing for the computation of the three global matching functions. They use the basic tools provided by the C++ library described at the beginning of this appendix.

B.2.1 Mutual Information

TexC++Code/MutualInfodEdI:

```

/*
  Author : Gerardo Hermosillo
  Copyright (c) INRIA 1997 - 2002
*/
#ifndef MUTUALINFO_DE_DI_H
#define MUTUALINFO_DE_DI_H

#include <Image.H>
#include <GeneralMetric.H>
10

template<class array>
struct MutualInfodEdI : public GeneralMetric<array> {

    Image<float> dist, Dist;
    float criter, norm, beta; int NG;

    virtual ~MutualInfodEdI() {}

    MutualInfodEdI ( const array & I1, const array & I2,
                    const float betaArg = 10.0,
                    const int ng = 256 ) : beta(betaArg), NG(ng) {
        dist.SetSize(I1.dimx(),I1.dimy(),I1.dimz());
    }
20

    void Init ( const array & I1, const array & I2) {

        Dist.SetSize(NG,NG);

        Image<float, const NeumannBC> H, Hy, Dist, P, h, hy, h1, p;
        H.SetSize(NG,NG); Hy.SetSize(NG,NG); Dist.SetSize(NG,NG);
30
        P.SetSize(NG,NG);
        h.SetSize(NG); hy.SetSize(NG); h1.SetSize(NG);
        p.SetSize(NG);

```

```

H=double(0); Map(I1,x) H((int)I1[x],(int)I2[x])++;

H.SelfRecSmoothZeroBC(beta);

float sH=0.0; Map(H,x) sH += H[x]; H /= sH; 40

MapY(H,y) {
    float sum = 0;
    MapX(H,x) {
        sum += H(x,y);
    }
    h(y) = sum;
}

MapX(H,x) { 50
    float sum = 0;
    MapY(H,y) {
        sum += H(x,y);
    }
    h1(x) = sum;
}

MapXY(H,x,y) Hy(x,y) = (H(x,y+1)-H(x,y-1))/2.0;
Map(h,y) hy(y) = (h(y+1)-h(y-1))/2.0; 60

Map(P,x) P[x] = H[x] ? Hy[x] / H[x] : Hy[x];
Map(p,x) p[x] = h[x] ? hy[x] / h[x] : hy[x];

MapXY(Dist,x,y) Dist(x,y) = P(x,y) - p(y);

Dist.SelfRecSmoothZeroBC(beta);

float s1 = 0;
MapXY(H,x,y) if(H(x,y)) s1 += H(x,y) * log(H(x,y)/(h1(x)*h(y))); 70

criter = -s1;

norm = 0;
Map(I1,x) {
    const float val = Dist(I1[x],I2[x]) ;
    dist[x] = val;
    norm += val * val;
}

norm /= I1.size(); 80
dist /= -I1.size();
}

float operator() (const int x, const int y, const int z=0) const {

```

```

    return dist(x,y,z);
}

float operator[] (const int x) { return dist[x]; }
                                                                    90

float E() const { return criter; }

int SaveE(const char *name) const {

    Image<float> mi(dist.domain()); mi = criter;
    return mi.SaveINR(name);
}

float GetNorm() const { return norm; }
                                                                    100

double AbsMax() const { return dist.AbsMax(); }

Image<float> & Ima() { return dist; }
};
#endif

```

B.2.2 Correlation Ratio

TexC++Code/CorrelRatiodEdI:

```

/*
   Author : Gerardo Hermosillo
   Copyright (c) INRIA 1997 - 2002
*/
#ifndef CORRELRATIO_DE_DI_H
#define CORRELRATIO_DE_DI_H

#include <Image.H>
#include <SecondOrderRecFilter.H>
#include <ExponentialMask.H>
#include <GeneralMetric.H>
                                                                    10

template<class array>
struct CorrelRatiodEdI : public GeneralMetric<array> {

    Image<float> dist;
    float criter, norm; float beta; int NG;

    virtual ~CorrelRatiodEdI() {}
                                                                    20

    CorrelRatiodEdI ( const array & I1, const array & I2,

```

```

        const float betaArg = 10.0,
        const int ng = 256 ) : beta(betaArg), NG(ng) {

    dist.SetSize(I1.dimx(),I1.dimy(),I1.dimz());
}

void Init ( const array & I1, const array & I2) {

    Image<float> H,Dist,h,V,S0,S1,S2;
    H.SetSize(NG,NG); Dist.SetSize(NG,NG);
    h.SetSize(NG); V.SetSize(NG);
    S0.SetSize(NG); S1.SetSize(NG); S2.SetSize(NG);

    typedef SecondOrderRecFilter F;
    ExponentialMask<float> M(beta);

    H = double(0); Map(I1,x) H((int)I1[x],(int)I2[x])++;
    MapY(H,y) F::Apply(M,&H(0,y),&H(0,y),H.dimx(),1);
    MapX(H,x) F::Apply(M,&H(x,0),&H(x,0),H.dimy(),H.dimx());
    float sH=0.0; Map(H,x) sH += H[x]; H /= sH;

    MapX(H,x) {
        float s0=0.0,s1=0.0,s2=0.0;
        MapY(H,y) {
            s0 += H(x,y);
            s1 += H(x,y) * y;
            s2 += H(x,y) * y * y;
        }
        S0(x) = s0;
        S1(x) = s1;
        S2(x) = s2;
    }

    Map(h,x) {
        float mean = S0[x] ? S1[x]/S0[x] : 0.0;
        float var = mean ? S2[x]/S0[x] - mean*mean: 0.0;
        h[x] = mean;
        V[x] = var;
    }

    float I2mean=0; Map(h,x) I2mean += S1[x];
    float sum2 = 0; Map(S2,x) sum2 += S2[x];
    float VarI2 = sum2 - I2mean*I2mean;
    float EVarI2I1 = 0;
    Map(V,x) EVarI2I1 += V[x]*S0[x];
    float CR = 1.0 - EVarI2I1/VarI2;

    MapXY(Dist,x,y) Dist(x,y) = y-h(x) + (CR-1.0) * (y-I2mean);
    MapY(Dist,y) F::Apply(M,&Dist(0,y),&Dist(0,y),Dist.dimx(),1);
    MapX(Dist,x) F::Apply(M,&Dist(x,0),&Dist(x,0),

```

30

40

50

60

70

```

        Dist.dimy(),Dist.dimx());
Map(Dist,x) Dist[x] *= -2.0/VarI2;

criter = 0;
Map(I1,x) {
    const float val = Dist(I1[x],I2[x]);
    dist[x] = val;
    norm += val;
}
criter = -CR;
norm /= I1.size();
dist /= -I1.size();
}
80

template <class point>
float operator() (const point &m) const {

    return dist(m.x,m.y);
}
90

float operator() (const int x, const int y,const int z=0) const {

    return dist(x,y,z);
}

float E() const { return criter; }

int SaveE(const char *name) const {

    Image<float> cr(dist.domain()); cr = criter;
    return cr.SaveINR(name);
}
100

float GetNorm() const { return norm; }

double AbsMax() const { return dist.AbsMax(); }

Image<float> & Ima() { return dist; }
};
110
#endif

```

B.2.3 Cross Correlation

TexC++Code/CrossCorreIdEdI:

```

/*
  Author : Gerardo Hermosillo
  Copyright (c) INRIA 1997 - 2002
*/

#ifndef CCGCOMPARISON_H
#define CCGCOMPARISON_H

#include <Image.H>
#include <GeneralMetric.H> 10

template<class array>
struct CrossCorreIdEdI : public GeneralMetric<array> {

  Image<float> dist;
  float criter, norm;
  float beta;

public: 20

  virtual ~CrossCorreIdEdI() {}

  CrossCorreIdEdI ( const array & I1, const array & I2,
                   const float betaArg = 10) :
    beta(betaArg) {
    dist.SetSize(I1.dimx(),I1.dimy(),I1.dimz());
  }

  void Init ( const array & I1, const array & I2) { 30

    typedef float real;

    Domain Omega = I1.domain();

    float m1=0, v1=0, m2=0, v2=0, v12=0;
    Map(Omega,x) {
      const real i1 = I1[x];
      const real i2 = I2[x];
      m1 += i1; m2 += i2;
      v1 += i1 * i1; v2 += i2 * i2; 40
      v12 += i1 * i2;
    }

    m1 /= Omega.size(); m2 /= Omega.size();
    v1 = beta + v1 / Omega.size() - m1 * m1;
    v2 = beta + v2 / Omega.size() - m2 * m2;

```

```

v12 = v12 / Omega.size() - m1 * m2;

const float f1    = v12 / (v2 * v1);
const float f2    = -criter / v2;
const float f3    = -(f2 * m2 + f1 * m1);
                                                                    50

Map(Omega,x) {
    dist[x] = 2.0 * ( f1 * I1[x] + f2 * I2[x] + f3 );
}

criter = -criter;
dist /= -Omega.size();
                                                                    60
}

float operator() (const int x, const int y, const int z=0) const {
    return dist(x,y,z);
}

float E() const { return criter; }

int SaveE(const char *name) const {

    Image<float> cc(dist.domain()); cc = criter;
    return cc.SaveINR(name);
                                                                    70
}

double AbsMax() const { return dist.AbsMax(); }

Image<float> & Ima() { return dist; }
};
#endif

```

B.3 C++ Listing: Local Matching Functions

This section contains the listing for the computation of the three local matching functions. They use the basic tools provided by the C++ library described at the beginning of this appendix.

B.3.1 Mutual Information

TexC++Code/LocalMidEdI:

```

/*
  Author : Gerardo Hermosillo
  Copyright (c) INRIA 1997 - 2002
*/

#ifndef MILCOMPARISON_H
#define MILCOMPARISON_H

#include <Image.H>
#include <GeneralMetric.H>                                10

template<class array>
struct MILComparison : public GeneralMetric<array> {

    typedef Image<float> Function; Function dist,mi;
    float criter, norm; float sigma, beta;
    static const int ws = 19;
    static const int hws = 9;

public:                                                    20

    virtual ~MILComparison() {}

    MILComparison ( const array & I1, const array & I2,
                    const float sigmaArg = 5, const float betaArg = 10) :
        sigma(sigmaArg), beta(betaArg) {

        dist.SetSize(I1.dimx(),I1.dimy(),I1.dimz());
        mi.SetSize(I1.dimx(),I1.dimy(),I1.dimz());
    }                                                         30

    void Init ( const array & I1, const array & I2 ) {

        Domain Omega = I1.domain();

        Function GB(300), R(ws,ws);
        Map(GB,x) GB[x] = ( 1.0/sqrt(2*M_PI*beta*beta) *
                           exp(-x*x/(2*beta*beta)) );
        MapXY(R,x,y) R(x,y) = ( 1.0/(2*M_PI*beta*beta) *
                                exp(-((x-hws)*(x-hws) +
                                         (y-hws)*(y-hws))/(2*sigma*sigma)) );
                                40

        criter = 0; norm = 0; int pixel=0;
        MapY(Omega,y) {
            const int yy = y-hws;
            MapX(Omega,x) {

```

```

const int xx = x-hws;
const float ii = I1[pixel];
const float jj = I2[pixel];
float S1=0,W1=0,S2=0,W2=0,W3=0;
int count=0;
MapY(R,Y) {
    const int py = yy+Y;
    MapX(R,X) {
        const int px    = xx+X;
        const float i    = I1(px,py);
        const float j    = I2(px,py);
        const float w1    = GB[abs(int(i-ii))];
        const float w2    = GB[abs(int(j-jj))];
        const float wx    = R[count++];
        const float ww2   = w2 * wx;
        const float ww1   = ww2 * w1;
        S1 += j * ww1;
        S2 += j * ww2;
        W1 += ww1;
        W2 += ww2;
        W3 += w1 * wx;
    }
}
const float di = (S1 / W1 - S2 / W2) / W3;
const float mix = (W1 - W2) / W3;
mi[pixel]      = mix;
dist[pixel++] = -di;
criter        += mix;
norm          += di * di;
}
}
criter /= Omega.size();
norm   /= Omega.size();
}

```

50

60

70

80

```

float e (const int x, const int y, const int z=0) const {

    return mi(x,y,z);
}

float operator() (const int x, const int y, const int z=0) const {
    return dist(x,y,z);
}

float E() const { return criter; }

int SaveE(const char *name) const { return mi.SaveINR(name); }

double AbsMax() const { return dist.AbsMax(); }

```

90

```

    Function & Ima() { return dist; }
};
#endif

```

100

B.3.2 Correlation Ratio

TexC++Code/LocalCRdEdI:

```

/*
   Author : Gerardo Hermosillo
   Copyright (c) INRIA 1997 - 2002
*/

#ifndef CRLCOMPARISON_H
#define CRLCOMPARISON_H

#include <Image.H>
#include <GeneralMetric.H>

template<class array>
struct CRLComparison : public GeneralMetric<array> {

    typedef Image<float> Function; Function dist,cr;
    float criter, norm; float sigma, beta, Beta;
    static const int ws = 19;
    static const int hws = 9;

public:

    virtual ~CRLComparison() {}

    CRLComparison ( const array & I1, const array & I2,
                   const float sigmaArg = 5, const float betaArg = 20 ) :
        sigma(sigmaArg), beta(betaArg), Beta(betaArg*betaArg) {

        dist.SetSize(I1.dimx(),I1.dimy(),I1.dimz());
        cr.SetSize(I1.dimx(),I1.dimy(),I1.dimz());
    }

    void Init ( const array & I1, const array & I2 ) {

        Domain Omega = I1.domain();

        Function GB(300), R(ws,ws);

```

10

20

30

```

Map(GB,x) GB[x] = ( 1.0/sqrt(2*M_PI*beta*beta) *
                    exp(-x*x/(2*beta*beta)) );
MapXY(R,x,y) R(x,y) = ( 1.0/(2*M_PI*beta*beta) *
                        exp(-((x-hws)*(x-hws) +
                              (y-hws)*(y-hws))/(2*sigma*sigma)) );

criter = 0; norm = 0; int pixel=0;
MapY(Omega,y) {
  const int yy = y-hws;
  MapX(Omega,x) {
    const int xx = x-hws;
    const float ii = I1[pixel];
    const float jj = I2[pixel];
    float S1=0,W1=0,S2=0,W2=0,S3=0,W3=0,S4=0,W4=0;
    int count=0;
    MapY(R,Y) {
      const int py = yy+Y;
      MapX(R,X) {
        const int px = xx+X;
        const float i = I1(px,py);
        const float j = I2(px,py);
        const float w1 = R[count++];
        const float w2 = w1 * GB[abs(int(i-ii))];
        S1 += w1*j; W1 += w1;
        S2 += w1*j*j; W2 += w1;
        S3 += w2*j; W3 += w2;
        S4 += w2*j*j; W4 += w2;
      }
    }
    const float mu2 = S1/W1;
    const float var2 = Beta + S2/W2 - mu2 * mu2;
    const float mu21 = S3/W3;
    const float var21 = Beta + S4/W4 - mu21 * mu21;
    const float theta1 = var21 / var2;
    const float di = (
      2.0 / var2 * ( mu2 - mu21 + theta1 * (jj - mu2) ) );
    dist[pixel++] = di;
    criter += theta1;
    norm += di*di;
  }
}

criter /= Omega.size();
norm /= Omega.size();
}

float e (const int x, const int y, const int z=0) const {
  return cr(x,y,z);
}

```

```

    }
    float operator() (const int x, const int y, const int z=0) const {
        return dist(x,y,z);
    }

    float E() const { return criter; }

    int SaveE(const char *name) const { return cr.SaveINR(name); }

    double AbsMax() const { return dist.AbsMax(); }

    Function & Ima() { return dist; }
};
#endif

```

90

100

B.3.3 Cross Correlation

TexC++Code/AutoCorreldEdI:

```

/*
   Author : Gerardo Hermosillo
   Copyright (c) INRIA 1997 - 2002
*/

#ifndef CCLCOMPARISON_H
#define CCLCOMPARISON_H

#include <Image.H>
#include <GeneralMetric.H>

template<class array>
struct CLComparison : public GeneralMetric<array> {

    Image<float> dist,cc;
    float criter, norm;
    float sigma, beta;

public:
    virtual ~CLComparison() {}

    CLComparison(const array & I1, const array & I2,
                 const float sigmaArg = 20, const float betaArg = 1e-13) :
        sigma(sigmaArg), beta(betaArg) {

        dist.SetSize(I1.dimx(),I1.dimy(),I1.dimz());
        cc.SetSize(I1.dimx(),I1.dimy(),I1.dimz());
    }

```

10

20

```

}

void Init ( const array & I1, const array & I2) {                                     30

    typedef float real;

    Image<float>
        mu1(I1.domain()), mu2(I1.domain()),
        v1(I1.domain()), v2(I1.domain()),
        v12(I1.domain()), f1(I1.domain()),
        f2(I1.domain()), f3(I1.domain());

    Map(I1,x) {                                                                     40
        const real i1 = I1[x];
        const real i2 = I2[x];

        mu1[x] = i1; v1[x] = i1 * i1;
        mu2[x] = i2; v12[x] = i1 * i2;
        v2[x] = i2 * i2;
    }

    mu1.SelfRecSmoothZeroBC(sigma); v1.SelfRecSmoothZeroBC(sigma);
    mu2.SelfRecSmoothZeroBC(sigma); v2.SelfRecSmoothZeroBC(sigma);                 50
    v12.SelfRecSmoothZeroBC(sigma);

    criter = 0;
    Map(v1,x) {

        const real u1 = mu1[x];
        const real u2 = mu2[x];
        const real vv1 = v1[x] + beta - u1 * u1;
        const real vv2 = v2[x] + beta - u2 * u2;
        const real vv12 = v12[x] - u1 * u2;                                       60

        const real ff1 = vv12 / (vv1 * vv2);
        const real CC = vv12 * ff1;
        const real ff2 = - CC / vv2;
        const real ff3 = - (ff2 * u2 + ff1 * u1);

        f1[x] = ff1; f2[x] = ff2; f3[x] = ff3;
        cc[x] = -CC;
        criter += -CC;
    }                                                                               70

    f1.SelfRecSmoothZeroBC(sigma);
    f2.SelfRecSmoothZeroBC(sigma);
    f3.SelfRecSmoothZeroBC(sigma);

    norm = 0;
    Map(f1,x) {

```

```

        const float val = 2.0 * ( f1[x] * I1[x] +
                                f2[x] * I2[x] +
                                f3[x] );
        dist[x] = -val;
        norm += val * val;
    }
}

float e (const int x, const int y, const int z=0) const {

    return cc(x,y,z);
}

float operator() (const int x, const int y, const int z=0) const {
    return dist(x,y,z);
}

float E() const { return criter; }

int SaveE(const char *name) const { return cc.SaveINR(name); }

double AbsMax() const { return dist.AbsMax(); }

Image<float> & Ima() { return dist; }
};
#endif

```

B.4 C++ Listing: 2D Matching Flow

This section contains the listing for a generic 2D matching flow using the six matching functions described. Only the 2D case is illustrated.

TexC++Code/MatchFlow:

```

/*
   Author : Gerardo Hermosillo
   Copyright (c) INRIA 1997 - 2002
*/
#ifndef MATCH_FLOW_H
#define MATCH_FLOW_H

#include <Image.H>
#include <Xhandler.H>
#include <Schemes.H>
#include <ImageMetric.H>

```

```

typedef Image<float> image;

struct Handler {
    bool stop; Handler() : stop(false) {}
    void Stop(Xhandler *xw) { stop=true; }
};

template <class real, bool visual=false, bool verbose=false>
struct MatchFlow {
    Xhandler X; Handler H;
    MatchFlow() { X.SetButton(3,&H,&Handler::Stop); }

    void operator () ( const image &I1o, const image &I2o,
                      image &u, image &v,
                      const real sigma, const real alpha,
                      real dt, int iter, const char* metric ) {
        image I1, I2, I2w, dispX, dispY, I2x, I2y;
        I1o >> I1 >> I2 >> I2w >> dispX >> dispY >> I2x >> I2y;

        I1 = I1o; I1.Smooth(1.695/sigma);
        I2 = I2o; I2.Smooth(1.695/sigma);
        I2x = I2o; I2x.Smooth(1.695/sigma,1,0);
        I2y = I2o; I2y.Smooth(1.695/sigma,0,1);

        float dto=0, criter = 0, Oldcriter = 0;

        ImageMetric<image> D(I1,I2,metric);
        D.Init(I1,I2);

        const float lambda = D.AbsMax();
        float C = alpha * lambda;
        dto = dt / C;

        Rmap(it,iter) {
            MapXY(I2w,x,y) I2w(x,y) = I2(x+u(x,y),y+v(x,y));
            D.Init(I1,I2w);

            int p = 0;
            MapXY(dispX,x,y) {
                const double di = -D(x,y);
                const Schemes::Vector
                    Elas = Schemes::Elasticity(u,v,x,y,0.8);
                dispX[p] = di * I2x(x+u(x,y),y+v(x,y)) + C * Elas.x;
                dispY[p] = di * I2y(x+u(x,y),y+v(x,y)) + C * Elas.y;
                p++;
            }
        }
    }
};

```

20

30

40

50

60

```

Oldcriter = criter;
criter = 0;
Map(u,x) {
    const float du = disp[x] * dto;
    const float dv = dispy[x] * dto;
    float & uu = u[x];
    float & vv = v[x];
    criter += uu * uu + vv * vv;
    uu += du;
    vv += dv;
}
criter /= I1.size();

if(verbose) printf("%.13f\n",criter);
if(visual) X(I2w);
if(H.stop) return;
}
};
#endif

```

70

80

B.5 C++ Listing: Main Program and Multiscale Handling

This section contains the listing for the main matching program, including multiscale handling.

TexC++Code/MatchPDE:

```

/*
   Author : Gerardo Hermosillo
   Copyright (c) INRIA 1997 - 2002
*/

#include <Image.H>
#include <Usage.H>
#include <MatchFlow.H>
#include <MultiScale.H>

typedef Image<float> image;
typedef MatchFlow<float,true,true> Matcher;

int main(int argc, char **argv) {

    char *n1, *n2; float alpha; int Nzoom;
    float sigma=0.25, dt; int iter;
    char *metric;

```

10

```

Usage Call ( argc, argv, "I1 I2 dt iter alpha Nzoom metric",
              n1, n2, dt, iter, alpha, Nzoom, metric );
                                                    20

Call(); image I1[Nzoom+1], I2[Nzoom+1], u[Nzoom+1], v[Nzoom+1];
I1[0]=n1; I2[0]=n2;

I1[0]>>u[0]>>v[0]; u[0]=v[0]=0.0;

for( int zoom = 1; zoom < Nzoom+1; zoom ++ ) {
                                                    30
    MultiScale::Zoom ( I1[zoom-1], I1[zoom] );
    MultiScale::Zoom ( I2[zoom-1], I2[zoom] );
    MultiScale::Zoom ( u[zoom-1], u[zoom] );
    MultiScale::Zoom ( v[zoom-1], v[zoom] );
}

for( int zoom = Nzoom; zoom >= 0; zoom -- ) {

    Matcher M;
                                                    40
    M ( I1[zoom],I2[zoom],u[zoom],v[zoom],sigma,alpha,dt,iter,metric );

    if ( zoom ) {
        MultiScale::DeZoom ( u[zoom], u[zoom-1] );
        MultiScale::DeZoom ( v[zoom], v[zoom-1] );
        u[zoom-1] *= 2.0;
        v[zoom-1] *= 2.0;
    }
}
                                                    50

u[0].SaveINR("U.inr"); v[0].SaveINR("V.inr");
}

```

Bibliography

- [1] L. Alvarez, R. Deriche, T. Papadopoulo, and J. Sanchez. Symmetrical dense optical flow estimation with occlusions detection. To appear in ECCV 2002.
- [2] L. Alvarez, R. Deriche, and F. Santana. Recursivity and PDE's in image processing . In *Proceedings 15th International Conference on Pattern Recognition*, volume I, pages 242–248, September 2000.
- [3] L. Alvarez, R. Deriche, J. Weickert, and J. Sánchez. Dense disparity map estimation respecting image discontinuities: A PDE and scale-space based approach. *International Journal of Visual Communication and Image Representation, Special Issue on Partial Differential Equations in Image Processing, Computer Vision and Computer Graphics*, 2000.
- [4] L. Alvarez, F. Guichard, P.L. Lions, and J.M. Morel. Axioms and fundamental equations of image processing. *Archive for Rational Mechanics and Analysis*, 123(3):199–257, 1993.
- [5] L. Alvarez, J. Weickert, and J. Sánchez. A scale-space approach to nonlocal optical flow calculations. In Mads Nielsen, P. Johansen, O.F. Olsen, and J. Weickert, editors, *Scale-Space Theories in Computer Vision*, volume 1682 of *Lecture Notes in Computer Science*, pages 235–246. Springer–Verlag, 1999.
- [6] L. Alvarez, J. Weickert, and J. Sánchez. Reliable Estimation of Dense Optical Flow Fields with Large Displacements. Technical report, Cuadernos del Instituto Universitario de Ciencias y Tecnologías Cibernéticas, 2000. A revised version has appeared at IJCV 39(1):41-56,2000.
- [7] Y. Amit. A nonlinear variational problem for image matching. *SIAM Journal on Scientific Computing*, 15(1), January 1994.
- [8] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *The International Journal of Computer Vision*, 2(3):283–310, January 1989.
- [9] J. B. Antoine Maintz and M. A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36, 1998.

- [10] G. Aubert, R. Deriche, and P. Kornprobst. Computing optical flow via variational techniques. *SIAM Journal of Applied Mathematics*, 60(1):156–182, 1999.
- [11] G. Aubert and P. Kornprobst. A mathematical study of the relaxed optical flow problem in the space BV. *SIAM Journal on Mathematical Analysis*, 30(6):1282–1308, 1999.
- [12] G. Aubert and P. Kornprobst. *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*, volume 147 of *Applied Mathematical Sciences*. Springer-Verlag, December 2001. Utiliser aubert-kornprobst:02.
- [13] J.L. Barron, D.J. Fleet, and S.S. Beauchemin. Performance of optical flow techniques. *The International Journal of Computer Vision*, 12(1):43–77, 1994.
- [14] J. Bergen, P. Anandan, K. Hanna, and R. Hingorani. Hierarchical Model-Based Motion Estimation. In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision*, pages 237–252, Santa Margherita, Italy, May 1992. Springer-Verlag.
- [15] M. Bertero, T. A. Poggio, and V. Torre. Ill-posed problems in early vision. *Proc. IEEE*, 76(8):869–889, August 1988.
- [16] M. Black and P. Anandan. Robust incremental optical flow. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, Urbana Champaign, IL, June 1992. IEEE.
- [17] D. Bosq. *Nonparametric Statistics for Stochastic Processes*, volume 110 of *Lecture Notes in Statistics*. Springer-Verlag, 2nd edition, 1998.
- [18] H. Brezis. *Analyse fonctionnelle. Théorie et applications*. Masson, 1983.
- [19] P. Cachier and N. Ayache. Regularization in non-rigid registration: I. trade-off between smoothness and intensity similarity. Technical Report 4188, INRIA, May 2001.
- [20] P. Cachier and N. Ayache. Regularization in non-rigid registration: Ii. isotropic energies, filters, and splines. Technical Report 4243, INRIA, July 2001.
- [21] P. Cachier and X. Pennec. 3d non-rigid registration by gradient descent on a gaussian weighted similarity measure using convolutions. In *Proceedings of MMBIA*, pages 182–189, June 2000.
- [22] P. Cachier and D. Rey. Symmetrization of the non-rigid registration problem using inversion-invariant energies: Application to multiple sclerosis. In *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2000*, volume 1935 of *Lecture Notes in Computer Science*. Springer, 2000.

- [23] Christophe Chefd'hotel, Gerardo Hermosillo, and Olivier Faugeras. Flows of diffeomorphisms for multimodal image registration. In *International Symposium on Biomedical Imaging*. IEEE, 2002.
- [24] G. E. Christensen and J. He. Consistent nonlinear elastic image registration. In *MMBIA'2001*, December 2001.
- [25] Gary Christensen, MI Miller, and MW Vannier. A 3d deformable magnetic resonance textbook based on elasticity. In *Proceedings of the American Association for Artificial Intelligence, Symposium: Applications of Computer Vision in Medical Image Processing*, 1994.
- [26] G.E. Christensen, R.D. Rabbitt, and M.I. Miller. Deformable template using large deformation kinematics. *IEEE Transactions on Image Processing*, 5(10):1437–1447, 1996.
- [27] P.G. Ciarlet. *Mathematical Elasticity*, volume 1. North Holland, 1988.
- [28] I. Cohen. Nonlinear variational method for optical flow computation. In *Scandinavian Conference on Image Analysis*, volume 1, pages 523–530, 1993.
- [29] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Machine Intell.*, 2001.
- [30] D. Comaniciu, R. Visvanathan, and P. Meer. The variable bandwidth mean shift and data-driven scale selection. In *Proceedings of the 8th International Conference on Computer Vision*, volume 1, pages 438–445, Vancouver, BC, Canada, July 2001.
- [31] R. Deriche. Fast algorithms for low-level vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(12):78–88, January 1990.
- [32] R. Deriche, P. Kornprobst, and G. Aubert. Optical flow estimation while preserving its discontinuities: A variational approach. In *Proceedings of the 2nd Asian Conference on Computer Vision*, volume 2, pages 71–80, Singapore, December 1995.
- [33] L.C. Evans. *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics*. Proceedings of the American Mathematical Society, 1998.
- [34] O. Faugeras. *Three-Dimensional Computer Vision: a Geometric Viewpoint*. MIT Press, 1993.
- [35] Olivier Faugeras, Bernard Hotz, Hervé Mathieu, Thierry Viéville, Zhengyou Zhang, Pascal Fua, Eric Théron, Laurent Moll, Gérard Berry, Jean Vuillemin,

- Patrice Bertin, and Catherine Proy. Real time correlation based stereo: algorithm implementations and applications. Technical Report 2013, INRIA Sophia-Antipolis, France, 1993.
- [36] Olivier Faugeras and Renaud Keriven. Variational principles, surface evolution, PDE's, level set methods and the stereo problem. *IEEE Transactions on Image Processing*, 7(3):336–344, March 1998.
- [37] T. Gaens, F. Maes D. Vandermeulen, and P. Suetens. Non-rigid multimodal image registration using mutual information. In J.van Leeuwen G. Goos, J. Hartmanis, editor, *First International Conference on Medical Image Computing and Computer-Assisted Intervention*, volume 1496 of *Lecture Notes in Computer Science*. Springer, 1998.
- [38] F. Girosi, Alessandro Verri, and Vincente Torre. Constraints for the Computation of Optical Flow. In *Proceedings Workshop on Visual Motion*, pages 116–124. IEEE Computer Society, 1989.
- [39] F. Guichard and L. Rudin. Accurate estimation of discontinuous optical flow by minimizing divergence related functionals. In *Proceedings of the International Conference on Image Processing*, volume I, pages 497–500, 1996.
- [40] S.N. Gupta and J.L. Prince. On div-curl regularization for motion estimation in 3-d volumetric imaging. In *Proceedings of the International Conference on Image Processing*, pages 929–932, 1996.
- [41] N. Hata, T. Dohi, S. Warfield, W. Wells III, R. Kikinis, and F. A. Jolesz. Multimodality deformable registration of pre-and intra-operative images for mri-guided brain surgery. In J.van Leeuwen G. Goos, J. Hartmanis, editor, *First International Conference on Medical Image Computing and Computer-Assisted Intervention*, volume 1496 of *Lecture Notes in Computer Science*. Springer, 1998.
- [42] F. Heitz and P. Bouthemy. Multimodal estimation of discontinuous optical flow using markov random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(12):1217–1232, December 1993.
- [43] D. Hill. *Combination of 3D medical images from multiple modalities*. PhD thesis, University of London, December 1993.
- [44] B.K. Horn and B.G. Schunck. Determining Optical Flow. *Artificial Intelligence*, 17:185–203, 1981.
- [45] M. Irani and P. Anandan. About direct methods. In *Vision algorithms : theory and practice. Proceedings of the International workshop on vision algorithms, Corfu, Greece, September 21-22, 1999*. Springer, 2000.

- [46] Jan J. Koenderink. Optic Flow. *Vision Research*, 26(1):161–180, 1986.
- [47] Jan J. Koenderink and Andrea J. van Doorn. Blur and Disorder. In Mads Nielsen, Peter Johansen, Ole F. Olsen, and Joachim Weickert, editors, *Scale-Space Theories in Computer Vision, Second International Conference, Scale-Space'99*, volume 1682 of *Lecture Note in Computer Science*, pages 1–9. Springer, 1999.
- [48] J.J. Koenderink. The structure of images. *Biological Cybernetics*, 50:363–370, 1984.
- [49] J. Kybic, P. Thevenaz, A. Nirkko, and M. Unser. Unwarping of unidirectionally distorted EPI images. *IEEE Transactions on Medical Imaging*, 19:80–93, 2000.
- [50] M.E. Leventon and W.E.L. Grimson. Multi-Modal Volume Registration Using Joint Intensity Distributions. In W.M. Wells, A. Colchester, and S. Delp, editors, *Medical Image Computing and Computer-Assisted Intervention-MICCAI'98*, number 1496 in *Lecture Notes in Computer Science*, Cambridge, MA, USA, October 1998. Springer.
- [51] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *International Joint Conference on Artificial Intelligence*, pages 674–679, 1981.
- [52] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE transactions on Medical Imaging*, 16(2):187–198, April 1997.
- [53] J.B.A. Maintz, H.W. Meijering, and M.A. Viergever. General multimodal elastic registration based on mutual information. In *Medical Imaging 1998 - Image Processing*, volume 3338, pages 144–154. SPIE, 1998.
- [54] M. Mattavelli and A. Nicoulin. Motion estimation relaxing the constant brightness constraint. In *Proceedings of the International Conference on Image Processing*, volume II, pages 770–774, 1994.
- [55] E. Mémin and P. Pérez. A multigrid approach for hierarchical motion estimation. In *Proceedings of the 6th International Conference on Computer Vision*, pages 933–938. IEEE Computer Society Press, Bombay, India, January 1998.
- [56] E. Mémin and P. Pérez. Dense/parametric estimation of fluid flows. In *IEEE Int. Conf. on Image Processing, ICIP'99*, Kobe, Japan, October 1999.
- [57] Chuck Meyer, Jennifer Boes, Boklye Kim, and Peyton Bland. Evaluation of control point selection in automatic, mutual information driven, 3d warping. In J.van Leeuwen G. Goos, J. Hartmanis, editor, *First International Conference*

- on Medical Image Computing and Computer-Assisted Intervention, Proceedings*, volume 1496 of *Lecture Notes in Computer Science*, October 1998.
- [58] M. Miller and L. Younes. Group actions, homeomorphisms, and matching : A general framework. *International Journal of Computer Vision*, 41(1/2):61–84, 2001.
- [59] A. Mitiche and P. Bouthemy. Computation and analysis of image motion: a synopsis of current problems and methods. *The International Journal of Computer Vision*, 19(1):29–55, July 1996.
- [60] H-H. Nagel. On the estimation of optical flow: relations between different approaches and some new results. *Artificial Intelligence Journal*, 33:299–324, 1987.
- [61] H.H. Nagel. Direct estimation of optical flow and of its derivatives. In G.A. Orban and H.H. Nagel, editors, *Artificial and Biological Vision Systems*, Basic Research Series, pages 191–224. Springer–Verlag, 1992.
- [62] H.H. Nagel and W. Enkelmann. An investigation of smoothness constraint for the estimation of displacement vector fields from images sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:565–593, 1986.
- [63] P. Nési. Variational approach to optical flow estimation managing discontinuities. *Image and Vision Computing*, 11(7):419–439, September 1993.
- [64] T. Netsch, P. Rosch, A. van Muiswinkel, and J. Weese. Towards real-time multi-modality 3d medical image registration. In *Proceedings of the 8th International Conference on Computer Vision*, Vancouver, Canada, 2001. IEEE Computer Society, IEEE Computer Society Press.
- [65] M. Otte and H.H. Nagel. Optical flow estimation: Advances and comparisons. In Jan-Olof Eklundh, editor, *Proceedings of the 3rd European Conference on Computer Vision*, volume 800 of *Lecture Notes in Computer Science*, pages 51–70. Springer–Verlag, 1994.
- [66] S. Ourselin, A. Roche, S. Prima, and N. Ayache. Block matching: A general framework to improve robustness of rigid registration of medical images. In *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2000*, volume 1935 of *Lecture Notes in Computer Science*. Springer, 2000.
- [67] E. Parzen. On the estimation of probability density function. *Ann. Math. Statist.*, 33:1065–1076, 1962.
- [68] A. Pazy. *Semigroups of Linear Operators and Applications to Partial Differential Equations*. Springer–Verlag, 1983.

- [69] X. Pennec, P. Cachier, and N. Ayache. Understanding the “demon’s algorithm”: 3d non-rigid registration by gradient descent. In *Second International Conference on Medical Image Computing and Computer-Assisted Intervention*, volume 1679 of *Lecture Notes in Computer Science*, pages 597–605. Springer, September 1999.
- [70] G. Penney, J. Weese, J.A. Little, P.Desmedt, D. LG. Hill, and D.J. Hawkes. A comparison of similarity measures for use in 2d-3d medical image registration. In J.van Leeuwen G. Goos, J. Hartmanis, editor, *First International Conference on Medical Image Computing and Computer-Assisted Intervention*, volume 1496 of *Lecture Notes in Computer Science*. Springer, 1998.
- [71] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, July 1990.
- [72] M. Proesmans, L. Van Gool, E. Pauwels, and A. Oosterlinck. Determination of Optical Flow and its Discontinuities using Non-Linear Diffusion. In *Proceedings of the 3rd ECCV, II*, number 801 in *Lecture Notes in Computer Science*, pages 295–304. Springer-Verlag, 1994.
- [73] A. Roche. *Recalage d’images médicales par inférence statistique*. PhD thesis, Université de Nice Sophia-Antipolis, February 2001.
- [74] A. Roche, A. Guimond, J. Meunier, and N. Ayache. Multimodal Elastic Matching of Brain Images. In *Proceedings of the 6th European Conference on Computer Vision*, Dublin, Ireland, June 2000.
- [75] A. Roche, G. Malandain, and N. Ayache. Unifying maximum likelihood approaches in medical image registration. *International Journal of Imaging Systems and Technology: Special Issue on 3D Imaging*, 11(1):71–80, 2000.
- [76] A. Roche, G. Malandain, X. Pennec, and N. Ayache. The correlation ratio as new similarity metric for multimodal image registration. In W.M. Wells, A. Colchester, and S. Delp, editors, *Medical Image Computing and Computer-Assisted Intervention-MICCAI’98*, number 1496 in *Lecture Notes in Computer Science*, pages 1115–1124, Cambridge, MA, USA, October 1998. Springer.
- [77] Alexis Roche, Grégoire Malandain, Xavier Pennec, and Nicholas Ayache. Multimodal image registration by maximization of the correlation ratio. Technical Report 3378, INRIA, August 1998.
- [78] D. Rückert, C. Hayes, C. Studholme, P. Summers, M. Leach, and D.J. Hawkes. Non-rigid registration of breast mr images using mutual information. In W.M. Wells, A. Colchester, and S. Delp, editors, *Medical Image Computing and*

- Computer-Assisted Intervention-MICCAI'98*, number 1496 in Lecture Notes in Computer Science, Cambridge, MA, USA, October 1998. Springer.
- [79] Daniel Scharstein and Richard Szeliski. Stereo matching with nonlinear diffusion. *International Journal of Computer Vision*, 28(2):155–174, June 1998.
- [80] C. Schnörr. Determining optical flow for irregular domains by minimizing quadratic functionals of a certain class. *The International Journal of Computer Vision*, 6(1):25–38, 1991.
- [81] J.P. Thirion. Image matching as a diffusion process: An analogy with maxwell's demons. *Medical Image Analysis*, 2(3):243–260, 1998.
- [82] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 839–846, January 1998.
- [83] A. Trouvé and L. Younes. Mise en correspondance par difféomorphismes en une dimension: définition et maximisation de fonctionnelles. In *12^{ème} Congrès RFIA '00*, Paris, February 2000.
- [84] Alain Trouvé. Diffeomorphisms groups and pattern matching in image analysis. *International Journal of Computer Vision*, 28(3):213–21, 1998.
- [85] A. Verri, F. Girosi, and V. Torre. Differential techniques for optical flow. *Journal of the Optical Society of America A*, 7:912–922, 1990.
- [86] A. Verri and T. Poggio. Motion field and optical flow: qualitative properties. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(5):490–498, 1989.
- [87] Paul Viola. *Alignment by Maximisation of Mutual Information*. PhD thesis, MIT, 1995.
- [88] Paul Viola and William M. Wells III. Alignment by maximization of mutual information. *The International Journal of Computer Vision*, 24(2):137–154, 1997.
- [89] Youngmei Wang. Physical model based non-rigid registration incorporating statistical shape information. In J.van Leeuwen G. Goos, J. Hartmanis, editor, *First International Conference on Medical Image Computing and Computer-Assisted Intervention*, volume 1496 of *Lecture Notes in Computer Science*. Springer, 1998.
- [90] J. Weickert. On discontinuity-preserving optic flow. In *Proc. Computer Vision and Mobile Robotics Workshop*, pages 115–122, Santorini, September 1998.

-
- [91] J. Weickert and C. Schnörr. A theoretical framework for convex regularizers in pde-based computation of image motion. *The International Journal of Computer Vision*, 45(3):245–264, December 2001.
- [92] J. Weickert and C. Schnörr. Variational optic flow computation with a spatio-temporal smoothness constraint. *Journal of Mathematical Imaging and Vision*, 14(3):245–255, May 2001.
- [93] W.M. Wells III, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis. Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1(1):35–51, 1996.
- [94] R.P. Woods, J.C. Maziotto, and S.R. Cherry. Mri-pet registration with automated algorithm. *Journal of computer assisted tomography*, 17(4):536–546, 1993.
- [95] Z. Zhang, R. Deriche, O. Faugeras, and Q.T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence Journal*, 78:87–119, October 1995.

