Distance-based shape statistics for image segmentation with priors

Guillaume Charpiat

École doctorale de l'École Polytechnique

Thèse

Présentée pour le grade de docteur de l'école Polytechnique

Spécialité : Informatique

Distance-based shape statistics for image segmentation with priors

Par

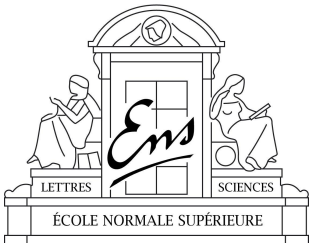
Guillaume Charpiat

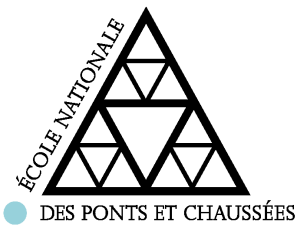
Soutenue le mercredi 13 décembre 2006

Devant le jury composé de :

Grégoire Allaire	
Olivier Faugeras	Directeur de thèse
Renaud Keriven	
David Mumford	Rapporteur
Guillermo Sapiro	Rapporteur
Alain Trouvé	Rapporteur

Équipe Odyssée
École Normale Supérieure / INRIA / ENPC





Remerciements

Je tiens tout d'abord à remercier Olivier Faugeras et Renaud Keriven pour la qualité de leur encadrement et de leur enseignement qui m'a particulièrement motivé à explorer avec eux le domaine de la vision par ordinateur et de la vision biologique. C'est leurs conseils et points de vue, en particulier leur conviction inaltérable en la possible et nécessaire alliance de la rigueur mathématique et de la performance effective des algorithmes, qui au long de ces années de DEA et de thèse m'ont progressivement initié au métier de la recherche. Je leur sais gré autant de m'avoir guidé et proposé des pistes théoriques et des applications intéressantes que de l'aide pratique technique et logicielle.

Je tiens également à remercier les membres du jury pour avoir accepté d'y participer et pour leur relecture en dépit du travail que cela représente. Je remercie aussi Grégoire Allaire pour m'avoir proposé comme monitorat d'assister son enseignement, ce qui m'a permis de découvrir le domaine de l'analyse numérique ; David Mumford pour la justesse mathématique de sa relecture ; Guillermo Sapiro pour m'avoir accueilli chaleureusement à Minnéapolis et pour avoir fait le déplacement ce jour ; et Alain Trouvé pour la qualité de son enseignement de cours de DEA et de ses travaux et réflexions dans le domaine, l'un des facteurs qui m'ont poussé vers l'étude des déformations de formes.

Je suis très reconnaissant à l'ensemble des enseignants à l'ENS dont j'ai suivi les cours de m'avoir fait ainsi découvrir autant de domaines passionnants.

Je voudrais remercier particulièrement Pierre Maurel, Jean-Philippe Pons et Jean-Yves Audibert pour les échanges et collaborations scientifiques qui apparaissent dans cet ouvrage.

Je remercie plus généralement l'équipe Odyssée dans son ensemble, qu'elle soit parisienne, campésienne ou sophipolitaine, ainsi que les autres certisois et assimilés (notamment Nikos Paragios), pour les discussions scientifiques intéressantes et constructives qui permettent de mieux appréhender le domaine, comme pour le bon esprit, l'ambiance amicale qui y règnent.

Merci en particulier aux Parisiens (Pierre, Irène, Thomas, Romain et les nouveaux) pour la bonne humeur, la vie quotidienne au labo...

Merci à Jill pour la relecture et la correction de l'anglais.

Profitons de l'occasion pour saluer ma famille et mes amis Marc, Guillaume, Arnaud, Liguori, Raphaël, Sylvain, Edouard, Antonin et Stéphanie, Yann, Sandra, Pierre, Eileen, Radu, Alina, Romain, Marie, Mickaël, Patrick, Luc... et les autres...

Enfin je remercie le lecteur de l'intérêt qu'il porte à cet ouvrage.

Introduction

Since the development of mechanics, one has wondered how far any task could be automated and performed by a machine. One has often tried to give machines a human facet, either by copying superficially and roughly the appearance of humans, or trying to model more deeply their internal mechanisms. The myths or dreams of robots or machines that behave like humans are much older than the technological breakthroughs that have led to the creation of computers. These ones however, due to their still increasing huge capacities in terms of quantity of information and programming, have aroused a great, productive enthusiasm for algorithmics. Nevertheless, the question of delimiting which tasks can be solved by a computer is still addressed, and people usually disagree on the possibility of the creation of an artificial "intelligence". Over time, the notion of "intelligence" has much evolved, loosing certainty, as the computer science has succeeded in resolving harder tasks. The best definition of "intelligence" seems to simply be "everything that can not be done by a machine (yet)". For instance the ability to defeat a human chess champion had been for a long time considered as requiring intelligence, but the intensive use of raw calculation power has changed the perspective. Beside these philosophical or scientific questions on the potentiality of machines, the industry has been asking for the automatization of more and more intricate tasks in order to decrease their costs or the required time to achieve them. An example of a domain where a lot of things remain to be done, from both practical and cognitive point of view, is the image processing. Indeed at the optimistic era of early computer programming the recognition of objects in images was thought to be an easy task, but still after decades the state of the art is not very satisfying, even if some methods have shown their efficiency in some limited applications.

The goal of image processing, or computer vision, consists in a first step in solving usual visual tasks that human beings accomplish everyday very naturally without even thinking of them. Examples of such tasks are : finding the contour of an object, following it, identifying it, seeing it in three dimensions whereas the eyes get only two-dimensional images, imagining its missing parts if it is partially occluded. And then, in a second step, some visual tasks require a little more reflexion, like analyzing and describing visual scenes with a few words. These tasks are respectively referred to as image segmentation, tracking, object classification, stereo reconstruction, inpainting and scene interpretation. They show themselves very useful in many applications, such as in medical imaging, video monitoring, and character recognition (known as OCR). For example the detection and identification of characters in an image allows to numerize scanned books and to sort letters at the post office by their ZIP code. Obviously the time gained for these repetitive tasks is huge.

The (static) image segmentation is often considered as being a low-level first step that needs to be done before going on towards more "semantic" tasks. This approach is known as "bottom-up". However, on one hand, the segmentation of a tracked object is really helped by the information coming from the motion (optical flow), which brings up the problem of what exactly should be considered as a low-level tool, and on the other hand, high-level information on objects or scene interpretation can also help the segmentation process. This last point of view is named "top-down".

How to find the contour of an object in an image ? Without any information about the

object, one standard method consists in searching for the regions of greatest homogeneity, with highest differences between them. Of course the quantity that is supposed to be homogeneous has to be precised. A possible homogeneity criterion is the spatial coherence of the color or the texture. Texture descriptors are local and based on the intensity of the image in the neighborhood of the point of interest and can involve multi-scale convolutions, like wavelets, or statistical tools, like Markov fields. The minimization of the chosen criterion should be made amongst all possible contours, but this set is far too huge to be entirely explored. If the criterion to minimize admits a simple enough formula that satisfies some conditions, then the global minimum can be found in a low-degree polynomial time thanks to the graph cuts technique. Depending on the criterion, other minimization methods consist in dealing with regions and successively merging or splitting them.

However in general the only method available is the gradient descent with respect to the boundaries, and this has naturally led to the variational approach where an arbitrary contour is initially drawn and then evolves in order to minimize the energy. This contour can be represented by a polygon or list of points (*snake*), or by splines, etc., or by the zero-level of a function defined on the whole image. In this last case the whole function evolves, even if the only center of interest is its zero-level, and this allows topological changes. A great advantage of the gradient descent method is that the minimized energy can explicitly deal with criteria dedicated to the contour itself. For instance the contour could be asked to be attracted by regions of high intensity gradient, which is a model of heterogeneity of an image at the boundary between two homogeneous regions. But, above all, criteria concerning the *shape* of the object can be used. For instance its smoothness is easily described by the curvature of the boundary, and this restriction that the boundary should be smooth (or should not) can be seen as a prior on the shape of the object.

Without more information about the object to be found, it is difficult to find better suited criteria. In fact the search for the best criterion without any knowledge about the object could be seen as an attempt to describe as best as possible the properties of usual objects in the real world and consequently it is already a *prior* on the usual world or more precisely on the objects of interest that it contains. Obviously in the case where more information about the object is available, its incorporation into the previous framework will improve the results, but this could be done in several manners. For example the programmer could build a model ex nihilo of the object, for example a mechanical model with rigid parts of fixed length and rotation centers, and then a small number of parameters naturally arise so that the minimization now depends only on these parameters. The color or the texture of the object could also be known. In all cases, the additional information removes some variables or sets more precisely their probability distribution.

The main point here is that these probability distributions could be learned rather than fixed by hand. Ideally the learning would be completely unsupervised, that is to say that one could expect the computer to elaborate progressively heuristics as the number of images that it has been given grows. However, to my knowledge, the art of successfully extracting heuristics from raw images without any prior on their structure and without any given example of segmentation does not exist yet and could be considered as a part of the "artificial intelligence" concept. Less ambitious but more realistic, the supervised or semi-supervised learning usually consists in bringing a sample set of examples already segmented and in estimating the probability distributions of some predefined variables (color, texture, area, diameter, etc.). Priors on the shape of the contour itself are however more intricate to deal with since a shape is not naturally equipped with a small number of parameters. A

simplistic approach consists in fixing a particular shape and searching for the best location of it in a new image by minimizing usual criteria with respect to the translation, rotation and scaling parameters. Of course such a prior do not take into account any kind of shape variability except rigid motion and is not well designed for learning. Most often, some shape parameters are introduced by hand by associating manually to some chosen points on a reference shape their correspondences on all other samples, and then some basic statistics like PCA (principal component analysis) are applied to the obtained coordinate vectors. The issue would then be to manage to automatically find the correspondences, that is to say, to establish meaningful correspondences between any two shapes. In a more intrinsic point of view, the whole shape rather than some points only could be considered, and this avoids the quest for particular points on shapes and their manual matching. However the notion of shape statistics supposes the ability to *compare* shapes in some manner, so that a possible way to start would be to establish correspondences between whole shapes, which is often done via shape warping.

Warping any shape onto any other one is a difficult problem also. More precisely, the difficulty is to warp shapes automatically so that the obtained correspondences seem quite intuitive to a human eye. Different approaches have been proposed, notably the mathematical morphology, based on dilatation and erosion, and the variational method, based on the gradient descent of the distance between the shapes. In the variational method, the choice of the distance to be minimized is crucial since it will act qualitatively on the obtained evolution paths. Amongst most used distances are the Hausdorff distance, the symmetric difference area, and the L^2 or $W^{1,2}$ norm between the associated signed distance functions. A variation where the distance depends on the whole path between the shapes allows to consider path-based metrics and consequently geodesics.

In any case, the problem of computing shape statistics from a sample set of shapes and their associated correspondences is addressed. Contrary to the usual case of vectors in $\mathbb{R}^n$ where PCA can be applied or histograms can be drawn, the set of all deformations that a shape may undergo is intrinsically infinite dimensioned, and therefore either a dimensionality reduction or a way to deal with infinite dimensions is needed. In practice, some tools like kernel methods (support vector machine) or graph Laplacian are designed to classify high-dimensioned data or to represent it in a low-dimensional space, and indeed in this purpose some authors have applied SVM techniques to images, which are then seen as pixel vectors. However such an application lacks the image structure and looses important notions such as the continuity or the curvature of a shape (not to say the notion of shape itself), and the related information that has been thus scattered is much more difficult to retrieve, especially if the training set is small. The point is that kernel methods show themselves helpful when there is no inherent structure or particular quantity of interest that can be previously described and modeled. As long as the dimensionality reduction tools are not efficient enough, or as long as there is no "artificial intelligence" that can guess the quantities of interest, the task addressed to the machine has to be deeply thought and precisely modeled.

The design of shape priors for the image segmentation consequently starts with an investigation of the notion of *shape*, a definition of the set of admissible shapes, coming along with a study of this set and a way to compare shapes, that is to say intrinsic distances between shapes. Then the way to *move* into this set can be also investigated and this naturally leads to the notions of deformation and shape gradient. Finally, the question of modeling shape distributions can be thought in terms of deformation distributions so that usual statistics can be performed on them.

This framework is followed here. The thesis is divided into three parts. In the first one the notion of *shape* is explored, with particular attention to sets of shapes and associated distances. The second part deals with shape warping, and more precisely with shape derivatives, which happens to be an inspiring topic. In the last part a definition of the mean shape of a sample set of shapes is given, as well as the one of characteristic deformations that convey the shape variability, and then this shape variability is turned into a shape prior for the image segmentation task. In order to make the reading easier, a small abstract has been added to the head of each chapter.

More precisely, the first part *Shapes and Distances* begins with the description of several distances between subsets of $\mathbb{R}^2$ that are often considered in the computer vision literature, namely the L^2 norm of the difference of characteristic functions, the L^∞ and $W^{1,2}$ norms of the difference of distance functions and the Hausdorff distance. The set $\mathcal{S}$ of all shapes is introduced as the set of curves with smooth boundaries, with bounded curvature and positive Federer reach. The topologies related to the three previous distances are then shown to be equivalent on this set. In chapter 2, a family of smooth approximations of the Hausdorff distance is built and studied. In particular it is shown that with suitable choices of parameters, these approximations can be as close as desired to the Hausdorff distance. These approximations are then shown to be continuous with respect to the previously studied topologies. The notion of *shape* is the subject of Chapter 3, which proposes different concepts of *shape* and associated distances. The relation between the Hausdorff distance of full objects and the Hausdorff distance of their boundaries is described and they are shown to be equal for close enough shapes in $\mathcal{S}$. Then the Hausdorff distance is extended in order to take into account more information, like color, local orientation or other local shape descriptors. A criterion between images, namely the local cross-correlation, is also introduced and shown to be closely related to a distance. The part ends with a brief study of "static" statistics that can be computed with the only knowledge of distances between shapes without any shape warping. The technique of the graph Laplacian is applied to sample sets of curves and leads to low-dimensional maps of them, whereas shapes are intrinsically infinite dimensioned. However in the case of sparse data this kind of technique seems limited.

A more dynamical way to explore the set of shapes and to compare them consists in warping any shape onto any other one. This is the subject of the second part, *Shape Warping*. Its first chapter introduces the notion of differentiation in $\mathcal{S}$ and after a brief theoretical study of this new structure on $\mathcal{S}$, the shape gradient of the approximation of the Hausdorff distance is computed and used to warp a shape onto another one by minimizing the distance with respect to one shape in a gradient descent framework. Chapter 6 is a small digression on the implementation of shape gradient in practice. Shapes are often represented by a finite number of parameters, such as the position of the vertices of a polyhedron, but the differentiation of a criterion with respect to these parameters introduces a prior which is specific to the representation. An intrinsic differentiation is proposed instead. Back to the shape matching task, chapter 7 presents a way to incorporate priors on minimization flows into the gradient descent framework. This is done by changing the inner product that defines the notion of *gradient*. The usual gradients in the computer vision literature are indeed related to the L^2 inner product but the consideration of other structures allows to set priors on the minimization flows. This is of great interest for usual variational methods since in practice the local minima to which they lead are then qualitatively different. This framework is applied to induce several degrees of spatial consistency in the warping flow, such as a Gaussian smoothing or a rigidification, while keeping the properties of a gradient

descent. Chapter 8 goes further and states an equivalent definition of the gradient which allows to consider more general priors. The gradient is seen itself as the minimizer of a criterion which can take into account any reasonable prior on the deformation field. An example which favors locally rigid motions is shown.

Finally, in part *Shape Statistics and Priors*, the tools developed in the two previous parts are used to build shape statistics which are themselves incorporated into shape priors. In chapter 9, a definition of the mean of a sample set of shapes is proposed, as well as the definition of characteristic modes of deformation that express the shape variability. These modes are computed from the knowledge of the gradients of the chosen distance between the mean shape and each of the sample shapes with respect to the mean shape. Examples are shown. Chapter 10 carries on the same spirit but with application to sample sets of images instead of contours. The Hausdorff distance is replaced by the local cross-correlation, and the instantaneous deformation fields by diffeomorphisms of the whole image. A mean image is defined, as well as eigenmodes of deformation which include both spatial and intensity variations. The example of a face database is shown. Then statistics on the spatial deformation fields are used to classify faces and recognize their expressions via SVM. The thesis ends with a chapter dedicated to the definition of shape priors that use shape statistics and to their application to an image segmentation task. Different probability distributions are proposed, such as a kernel method based on shape gradients and the Mahalanobis distance of the shape gradient with knowledge of eigenmodes and their associated standard deviations. The shape priors are then adapted in order to be invariant with respect to rigid motion. Since their minimization involves the second cross-derivative of the chosen distance, this one is computed in the case of the approximation of the Hausdorff distance. The importance of shape priors is then illustrated by examples.

Most of chapters 1, 2, 5 and 9 comes from the article [18] *Approximations of Shape Metrics and Application to Shape Warping and Empirical Shape Statistics* with Olivier Faugeras and Renaud Keriven published in the journal *Foundations of Computational Mathematics*, 2004, and turned into a chapter [20] with the same title in the book *Statistics and Analysis of Shapes*, H. Krim & A. Yezzi editors, 2006, with an additional extension to surfaces in $\mathbb{R}^3$ by Pierre Maurel. A short and preliminary version of this work can be found in the article [17] *Shape Metrics, Warping and Statistics* in the proceedings of the *International Conference on Image Processing* held in 2003. Most of chapters 7 and 8 comes from the article [23] *Generalized Gradients : Priors on Minimization Flows* with Pierre Maurel, Jean-Philippe Pons, Renaud Keriven and Olivier Faugeras, published in the *International Journal of Computer Vision*, 2007. It is the continuation of the work begun in the article [21] *Designing Spatially Coherent Minimizing Flows for Variational Problems Based on Active Contours* with Jean-Philippe Pons, Renaud Keriven and Olivier Faugeras, in the proceedings of *International Conference on Computer Vision* held in 2005. Chapter 10 comes from the article [19] *Image Statistics based on Diffeomorphic Matching* with Olivier Faugeras and Renaud Keriven, in the proceedings of the same conference, except the last part concerning expression recognition, which comes from the ENPC research report of the same title and to which Jean-Yves Audibert has also contributed, providing classification tools. The technique exposed in chapter 4 is briefly presented in the article [22] *Distance-Based Shape Statistics* with Pierre Maurel, Renaud Keriven and Olivier Faugeras, in the proceedings of the *International Conference on Acoustics, Speech, and Signal Processing*, Special Session *Statistical Inferences on Nonlinear Manifolds with Applications in Signal and Image Processing* held in 2006.

Most programs make intensive use of C++ libraries[1] developed by the Odyssee Team, namely the Level-Set library (LSLIB) and the WinLib. Computations in chapters 7 and 11 use some additions by Pierre Maurel to the LSLIB. Figures 7.2, 7.4, 7.5 and 7.6 by Pierre Maurel, who appears on figure 8.4. Computations in chapter 10 use at some level a source code by Jean-Philippe Pons based on yar++, another Odyssee Team image library, developed by Gerardo Hermosillo.

[1] `http ://cermics.enpc.fr/~keriven/software/`

Contents

Résumé en français

Motivation

Le but de cette thèse est de construire un terme d'a priori sur la forme pour aider la segmentation d'images lorsque l'on dispose d'un ensemble d'exemples de contours auxquels on voudrait que le résultat de la segmentation *ressemble*. Pour cela on cherche tout d'abord à exprimer des statistiques sur ces contours, puis à déduire de ces statistiques un critère exprimant la probabilité d'une nouvelle forme étant donné l'échantillon d'apprentissage.

Rappel sur l'approche variationnelle

L'approche variationnelle en segmentation d'images consiste à définir un critère E dépendant d'une courbe C ou surface, de choisir une certaine position initiale C_0 et de minimiser ce critère par descente de gradient :

$$\begin{aligned} C(0) &= C_0 \\ \partial_t C(t) &= -\nabla_C E\big(C(t)\big) \end{aligned}$$

afin de trouver le meilleur contour C pour ce critère. Malheureusement une descente de gradient garantit seulement de trouver un minimum local et en pratique les résultats dépendent beaucoup de l'initialisation. Par ailleurs les critères sont assez difficiles à déterminer, le choix de l'énergie à minimiser dépend fortement du type de données et de la tâche précise que l'on cherche à accomplir en segmentant l'image.

Les critères usuels sont basés soit sur des propriétés locales le long de la contour C, telle que l'estimation de la présence éventuelle de bords en un point de ce contour à partir du gradient de l'intensité de l'image en ce point, soit sur des propriétés propres aux deux régions délimitées par le contour, telles que l'homogénéité de la couleur ou de la texture dans chacune des deux régions en question.

Ces critères ne porte aucune information ou restriction sur la forme que doit avoir le résultat de la segmentation, sinon éventuellement sa régularité (grâce à des termes portant sur la longueur de la courbe) ou quelques paramètres tels que son diamètre.

On souhaiterait ici construire un nouveau critère qui ne dépende que de la forme de la courbe (et non pas de l'intensité de l'image) et qui exprime à quel point cette courbe est typique ou non des exemples observés auparavant (fournis éventuellement par une segmentation manuelle de quelques images analogues).

Exigences

Il serait hautement souhaitable que l'approche et le critère à définir vérifient un certain nombre de propriétés :

- que l'approche soit bien posée mathématiquement ;
- que les quantités exprimées soient intrinsèques aux formes et non dépendant de leur paramétrisation ;
- que l'approche soit facilement généralisable au cas d'hypersurfaces dans un $\mathbb{R}^n$ quelconque, notamment au cas de surfaces dans un environnement tri-dimensionnel ;
- que le critère rende compte au mieux de la variabilité de la forme ;
- que le critère soit dérivable afin que l'on puisse effectuer une descente de gradient ;
- que le critère puisse être invariant aux transformations rigides (translation, rotation, changement d'échelle).

Travaux déjà existants sur ce sujet

Méthode Parzen

Il faut mentionner ici les travaux de Cremers [28, 29, 27] qui consistent à appliquer la méthode Parzen aux courbes : une distance d étant choisie sur l'espace des courbes, on définit le critère

$$E(C) = \sum_{i=1}^{N} \exp^{-\frac{d(C,C_i)^2}{2\sigma^2}}$$

où C_i est l'ensemble des N courbes de l'ensemble d'apprentissage et σ un paramètre à fixer. L'un des problèmes de cette approche est qu'elle nécessite un ensemble d'apprentissage très dense, sans quoi pour les petites valeurs de σ la courbe C sera attirée vers un échantillon déjà observé et pour les grandes valeurs de σ la courbe C ne sera certes pas exactement sur un des exemples C_i mais la déformation qui relie ces deux courbes sera *quelconque* et non nécessairement représentative des déformations que l'on peut observer dans l'ensemble échantillon. C'est justement au fait que la déformation entre la courbe C et les échantillons C_i doit être *typique* que l'on s'attache dans cette thèse.

ACP sur fonction distance

Un certain nombre de travaux, tels ceux de Leventon [73, 74] ou de Rousson et Paragios [94, 98] s'appuient sur la représentation par fonction distance signée de la forme et expriment des statistiques via l'application d'une analyse en composantes principales de ces fonctions distance. Le problème ici est qu'une combinaison linéaire de fonctions distance n'est généralement pas elle-même une fonction distance ; or la décomposition en composantes principales suppose intrinsèquement que les combinaisons linéaires de fonctions distances font sens. Des phénomènes assez étranges peuvent survenir lorsque l'on moyenne ainsi des fonctions distance.

Plan de la thèse

La thèse se découpe en trois parties : en premier lieu, une étude mathématique de l'ensemble des formes et des métriques que l'on peut poser dessus, ainsi qu'une brève étude de ce que l'on peut faire lorsque l'on ne peut disposer que des distances entre les formes (sans aucun déplacement dans l'espace des formes, c'est-à-dire sans déformation). En second lieu, une étude de la notion de dérivation d'une énergie par rapport à une courbe et plus généralement des déformations dans l'espace des formes : le gradient de forme se révèle dépendre fortement du produit scalaire posé dans l'espace tangent ; ce produit scalaire est alors vu comme un a priori sur les déformations d'un intérêt primordial. La troisième partie combine les résultats des deux parties précédentes afin de définir la moyenne d'un ensemble de formes ainsi que les modes de déformation caractéristiques associés. Sur ces statistiques l'on construit un a priori sur la forme qui est illustré par un exemple de segmentation d'image. Des statistiques similaires sont également définies pour un ensemble d'images (et non de formes) et illustrées sur une base de données de visages ; l'application est cette fois la reconnaissance d'expressions de visages.

Première partie : Formes et distances

Cette première partie est consacrée à l'ensemble des formes et à ses métriques.

Nous définissons tout d'abord (chapitre 1) mathématiquement l'ensemble de "toutes les formes" $\mathcal{S}$ et nous étudions plusieurs métriques sur cet ensemble. En particulier nous montrons que trois des métriques les plus courantes sont topologiquement équivalentes.

Nous choisissons parmi elles la distance de Hausdorff qui a l'avantage d'exprimer plus d'information géométrique au sens où, si l'on essaye de déformer une courbe de façon à minimiser sa distance à une autre courbe, le champ de déformation à suivre se révèle dans certains cas plus naturel que la distance L^2 entre les fonctions distance par exemple. Nous étudions alors plus précisément cette distance (chapitre 3) et proposons quelques extensions pour prendre en considération d'autres notions telles que l'orientation locale de la courbe. Nous proposons également (chapitre 2) une approximation dérivable de cette distance afin de pouvoir effectuer une descente de gradient par la suite.

Que peut-on obtenir comme statistiques sur un ensemble de formes si l'on a accès seulement à la distance entre tout couple de formes ? On peut représenter un échantillon de N courbes comme un graphe dont chaque courbe est un noeud (chapitre 4). Les fonctions propres du laplacien de ce graphe peuvent alors être utilisées comme un système de coordonnées, ce qui permet d'avoir une représentation de faible dimension de l'ensemble des N courbes.

Chapitre 1 : Ensembles de formes et métriques usuelles

Pour définir l'ensemble des formes dans lequel on va travailler, on s'inspire des travaux de Delfour et Zolésio [33]. Étant donné un domaine compact D de $\mathbb{R}^n$ (l'image qui contient les formes), on considère l'ensemble $\mathcal{S}$ des ensembles de points de D qui sont à la fois :

- dans C^2 : au sens où leur frontière est non vide et peut-être localement représentée comme le graphe d'une fonction C^2 ;
- dans l'ensemble de Federer $\mathcal{F}(h_0)$: la fonction distance signée à la forme est à distance $\geqslant h_0$ de son squelette.

Les métriques les plus courantes sur les ensembles de formes sont :

- la distance de Hausdorff
- l'aire de la différence symétrique
- la norme L^2 ou $W^{1,2}$ de la différence des fonctions distance associées.

Le point important de ce chapitre est que l'on prouve que les topologies relatives au trois distances ci-dessus sont équivalentes sur $\mathcal{S}$. Il est à noter que la contrainte $\mathcal{F}(h_0)$ y joue un rôle primordial et que l'équivalence n'a plus lieu sans.

Un des résultats de compacité obtenus lors de l'étude de $\mathcal{S}$ permet également de prouver que l'enveloppe semi-continue inférieurement d'une fonction continue sur $\mathcal{S}$ atteint son minimum sur $\overline{\mathcal{S}}$ (quelle que soit la topologie choisie parmi les trois).

L'équivalence entre les trois topologies n'implique pas pour autant que les trois distances se comportent de façon similaire à plus grande échelle. On s'intéressera désormais en particulier à la distance de Hausdorff, car celle-ci semble exprimer plus d'information géométrique que les autres (à des échelles non epsilonesques).

Chapitre 2 : Approximations de distances de formes

La distance de Hausdorff entre deux courbes n'étant pas dérivable par rapport à ces courbes, on définit dans ce chapitre une famille d'approximation de la distance de Hausdorff

qui consiste à remplacer les $\inf_{x\in A} f(x)$ et $\sup_{x\in A} f(x)$ où x est un point de la courbe A et f une fonction régulière par

$$\Psi^{-1}\left(\frac{1}{|A|}\int_A \Psi\left(f(x)\right)\ dx\right)$$

avec Ψ fonction croissante pour approximer le sup et décroissante pour approximer l'inf. Cette astuce n'est pas sans rappeler l'approximation $\|\cdot\|_p \longrightarrow \|\cdot\|_\infty$ quand p tend vers l'infini.

L'on montre que ces approximations sont continues pour les topologies étudiées au chapitre précédent.

Une astuce similaire peut être utilisée pour approximer de façon dérivable et continue la norme L^2 ou $W^{1,2}$ des fonctions distance signée associées aux courbes.

Le cas de l'approximation de la distance de Hausdorff est étudié plus particulièrement, et l'on quantifie l'erreur entre la distance et son approximation en fonction des paramètres de celle-ci (les paramètres sont les analogues de p dans $\|\cdot\|_p \longrightarrow \|\cdot\|_\infty$). En particulier l'on montre la convergence de l'approximation vers la vraie distance quand les paramètres tendent vers l'infini. Nous n'avons cependant pas étudié la convergence du gradient de l'approximation vers le "gradient" de la distance de Hausdorff (en supposant que l'on puisse définir celui-ci) vu que les calculs nécessaires seraient bien fastidieux. Ce point cependant mériterait d'être étudié plus précisément vu l'usage intensif des gradients qui est fait par la suite. On peut néanmoins conjecturer, vu les résultats obtenus en pratique, que cette convergence semble bien avoir lieu. Il est à noter que dans les applications pratiques l'on considèrera de relativement faibles valeurs des paramètres, afin que les déformations soient suffisamment lisses, plutôt que des valeurs élevées, ce qui aurait pour conséquence de faire ressembler les gradients à une somme de pics pointus (et qui n'est guère souhaitable pour construire des chemins entre des formes).

Chapitre 3 : Différentes notion de *forme* et distances associées

La notion intuitive de *forme* peut prendre plusieurs aspects : contour d'un objet, l'objet lui-même, avec ou sans couleur... Dans ce chapitre on essaye de mathématiser ces différentes notions et de poser des distances sur les ensembles de formes correspondants.

On étudie tout d'abord le lien entre la distance de Hausdorff entre les "formes pleines", c'est-à-dire les objets, et la distance de Hausdorff entre les frontières de ces objets. On montre notamment que sur l'espace $\mathcal{S}$ introduit précédemment, si l'une de ces deux distances est suffisamment petite, alors l'autre également et elles sont alors égales.

On présente ensuite une façon d'étendre la distance de Hausdorff pour qu'elle puisse prendre en compte des descripteurs locaux de la forme, telles la courbure, l'orientation (c'est-à-dire la normale à la forme au point considéré), la couleur... Pour cela on troque la distance euclidienne $d(x, y)$ qui est la brique de construction de la distance de Hausdorff pour une distance qui prend également en compte ces descripteurs locaux. On montre en particulier que cette extension de la distance de Hausdorff est toujours une distance.

L'on compare enfin la notion d'objet "plein" à la notion d'image. On présente la corrélation croisée locale, critère régulièrement utilisé dans la littérature pour comparer deux images. On montre que ce critère découle d'une distance (au sens mathématique du terme). On présente également la technique dite de multi-résolution, qui permet d'éviter un certain nombre de minima locaux lors de la déformation d'une image vers une autre par minimisation de la corrélation croisée locale.

Chapitre 4 : Statistiques basées sur la seule connaissance des distances

L'étude d'espaces des formes et des distances que l'on peut définir dessus mène naturellement à la question de savoir ce que l'on peut extraire comme type de statistiques à partir de la seule connaissance des distances entre les formes (c'est-à-dire sans aucune déformation de forme, sans aucun mouvement dans cet espace des formes).

Étant donné un échantillon de formes et les distances entre elles, on peut construire un graphe où chaque forme est un noeud du graphe qui est relié à ses K plus proches voisins (pour une valeur prédéfinie de K). Sur ce graphe on peut définir une version discrète du laplacien et calculer ses fonctions propres, qui ont la particularité d'être les "meilleures" fonctions pouvant servir de système de coordonnées (pour un certain critère simple de régularité). L'on peut alors représenter l'ensemble de formes étudié dans un espace de faible dimension grâce aux premières fonctions propres.

Utiliser cette représentation pour construire un a priori de forme pour la segmentation d'images n'est pas chose aisée. Quelques pistes peut-être du côté des extensions des cartes harmoniques qui permettent d'affecter à un nouveau point des coordonnées, en supposant que l'on puisse dériver simplement ces coordonnées par rapport à la courbe afin de la "projeter" sur la "variété" des courbes déjà observées. Cependant il est difficile de concevoir comment se mouvoir dans l'espace des formes sans avoir étudié préalablement la déformation de formes, d'où la partie suivante.

Deuxième partie : Déformation de contours

Cette partie est consacrée à la définition d'un "meilleur" champ de déformation à appliquer à une forme pour minimiser une énergie.

Tout d'abord (chapitre 5) la dérivation d'une énergie par rapport à une forme est mathématiquement définie. On peut alors appliquer une descente de gradient à la distance entre deux formes pour construire un chemin les reliant. Des exemples dans le plan et en trois dimensions sont montrés.

La notion de dérivation intrinsèque est alors abordée (chapitre 6) : il existe en effet une certaine manière de calculer, discrétiser et implémenter le gradient d'une énergie par rapport à une forme de façon à ce que le résultat dépende le moins possible du choix particulier de discrétisation et de représentation de la forme.

Le gradient dépend du produit scalaire choisi dans l'espace tangent à la forme. Cette dépendance est explorée (chapitre 7) et d'autres produits scalaires sont proposés afin de poser des a priori sur les déformations.

Une extension de la définition de gradient est alors proposée (chapitre 8) afin de prendre en compte des critères plus variés.

Chapitre 5 : Gradient de forme et déformation de forme

Pour minimiser un critère E dépendant d'une forme Γ par descente de gradient, il faut tout d'abord savoir exprimer le gradient en question. L'espace tangent à une courbe (ou surface, etc.) est l'ensemble des déformations infinitésimales que l'on peut appliquer à cette courbe, ce sont donc des champs de déformation définis sur la courbe. On définit tout d'abord la dérivée directionnelle dans une direction v (qui est un de ces champs de déformation infinitésimaux) par la dérivée de Gâteaux :

$$\mathcal{G}_\Gamma(E(\Gamma, \Gamma_0), \boldsymbol{\beta}) = \lim_{\varepsilon \to 0} \frac{E(\Gamma + \varepsilon\boldsymbol{\beta}, \Gamma_0) - E(\Gamma, \Gamma_0)}{\varepsilon}.$$

L'addition d'une courbe et d'un champ de déformation peut être définie simplement si l'on voit la courbe comme une fonction du cercle (ou autre chose selon sa topologie) à valeurs dans $\mathbb{R}^2$ (ou $\mathbb{R}^n$ plus généralement) : la somme de la courbe et du champ est la fonction définie sur ce même cercle mais qui renvoie la somme de l'ancienne valeur dans $\mathbb{R}^2$ et du champ de déformation. L'ensemble obtenu par cette somme ne dépend pas de la paramétrisation choisie.

Dans le cas où cette dérivée directionnelle est linéaire et continue en v, le théorème de Riesz permet de représenter la dérivée par un champ de déformation, nommément le gradient :

$$\mathcal{G}_\Gamma(E(\Gamma, \Gamma_0), \boldsymbol{\beta}) = \langle \nabla E(\Gamma, \Gamma_0), \beta \rangle.$$

Ce gradient dépend naturellement du produit scalaire que l'on a choisi dans l'espace tangent. Le plus couramment, c'est le produit $L^2(\Gamma)$ qui est choisi.

Pour effectuer une descente de gradient il faut d'abord avoir défini la notion de dérivée d'une famille (continue) de courbes indexée par un paramètre par rapport à ce paramètre (le temps t dans le cas de la descente de gradient). Il existe deux manières naturelles de définir cette dérivée, en utilisant la distance de Hausdorff ou en utilisant la métrique reliée aux produits scalaires des espaces tangents. Heureusement il se trouve que les deux définitions coïncident ici. Par analogie avec les variétés, on étudie également une notion proche de celle du transport ainsi que les géodésiques dans l'espace des formes. En particulier, une descente de gradient effectuée pour minimiser la distance entre deux courbes ne conduit généralement pas à une géodésique.

La descente de gradient sur la distance $d(\Gamma, \Gamma_2)$ entre deux courbes par rapport à Γ permet de construire progressivement un chemin reliant Γ à Γ_2, et donc de *transformer* l'une en l'autre. En pratique la restriction que les courbes devrait rester dans l'espace $\mathcal{S}$ introduit au premier chapitre, en particulier la contrainte de Federer $\mathcal{F}(h_0)$, n'a pas été implémentée. Il est alors possible qu'une descente de gradient fasse sortir de ce cadre théorique, on perd alors l'équivalence des topologies liées aux métriques étudiées aux moments précis où la courbe qui évolue change de topologie (i.e. de nombre de composantes connexes).

On calcule le gradient de l'approximation dérivable de la distance de Hausdorff et on l'utilise pour construire des chemins reliant deux courbes par descente de gradient sur cette énergie. Le tout est implémenté avec la technique des level-sets et illustré par de exemples. Le gradient d'une approximation de la norme $W^{1,2}$ entre les fonctions distance signée est également calculé.

Chapitre 6 : Dérivation intrinsèque

Ce court chapitre est une réflexion sur la façon d'implémenter les descentes de gradient. En pratique, les courbes sont souvent représentées par un nombre fini (petit) de paramètres, comme les positions des sommets d'un polygone ou les paramètres des splines. Calculer le gradient de l'énergie par rapport aux paramètres induit un a priori spécifique à la représentation de la courbe dans les déformations obtenues.

Il est possible de définir un gradient plus intrinsèque au sens où le choix de discrétisation de la courbe jouera un rôle plus faible. Pour cela on exprime l'ensemble des déformations possibles (par exemple, si la courbe est représentée par un polygone, l'ensemble généré par les déplacements de ses sommets) et l'on projette sur cet ensemble le gradient usuel, qui est, lui, intrinsèque (il ne dépend pas de la représentation choisie). On obtient ainsi le meilleur

champ de déformation disponible vu la représentation choisie, qui soit le plus proche possible du vrai gradient par rapport à la courbe (en tant que courbe quelconque, sans la restriction qu'elle doive rester un polygone). On montre que le champ de déformation obtenu résulte en fait d'un choix particulier de produit scalaire et peut être vu comme un gradient lui-même.

Chapitre 7 : Gradients généralisés comme a priori sur les flots de minimisation

On a vu précédemment lors de la définition du gradient que celui-ci dépend du produit scalaire. Dans ce chapitre l'on étudie différents produits scalaires possibles ainsi que les conséquences sur le gradient.

Bien que l'énergie minimisée soit la même, une descente de gradient effectuée pour deux produits scalaires différents n'empruntera pas nécessairement les mêmes chemins et aboutira éventuellement dans des minima locaux différents. Le choix du produit scalaire peut être alors vu comme un a priori sur le type de chemin à prendre, ou plus exactement sur le type de déformation instantanée que subit la courbe.

Quelques produits scalaires particuliers sont étudiés. Par exemple le produit scalaire H^1 permet d'obtenir un champ de déformation plus lisse. Etant donné un sous-espace vectoriel de transformations "préférées", il est possible de définir un produit scalaire qui va naturellement privilégier les transformations de ce sous-espace. Ainsi l'on peut privilégier les similitudes (translation, rotation, homothétie) afin que la *forme* de la courbe évoluant soit au mieux préservée lors de l'évolution. L'intérêt de passer par un changement de produit scalaire pour cela est que le cadre théorique de la descente de gradient reste identique et que l'on a les mêmes garanties que pour le flot minimisant habituel L^2 d'aboutir dans un minimum local de l'énergie minimisée. L'espoir est que les minima locaux trouvés soient "meilleurs" grâce aux a priori ainsi introduits sur le chemin à suivre. Le lissage d'un champ par une gaussienne peut être vu comme l'application d'une transformation linéaire symétrique définie positive à ce champ, ce qui permet de conserver également les mêmes propriétés de convergence vers un minimum local.

On ne peut cependant appliquer successivement deux transformations linéaires symétriques définies positives à un même champ vu que la transformation composée qui en résulte n'est elle pas symétrique. On propose une solution générique pour ce faire, ainsi qu'une autre solution particulière adaptée au cas de la rigidification du mouvement combinée au lissage du "bruit" restant.

D'un point de vue pratique, ces nouveaux flots minimisants sont spatiallement plus cohérents et induisent des évolutions beaucoup plus intuitives.

Chapitre 8 : Gradients étendus : des a priori plus généraux

Les changements de produit scalaire étudiés au chapitre précédent peuvent être vus comme des transformations linéaires que l'on applique au gradient usuel. On peut cependant étendre la définition de *gradient* de façon à poser des a priori plus généraux sur les champs de déformation.

Etant donné un produit scalaire F, le gradient $\nabla_F E$ qui lui est associé vérifie la propriété suivante :

$$-\nabla_F E(\Gamma) = \arg\min_v \left[\delta E(\Gamma, v) \; + \; \frac{1}{2}\|v\|_F^2 \right]$$

Cette propriété est intéressante puisqu'elle ne fait intervenir que la norme associée à F et non pas le produit scalaire lui-même, et que l'on peut voir le gradient comme le meilleur

champ de déformation qui minimise au mieux l'énergie pour un coût (sa norme F) le plus faible possible. La norme F est ainsi un critère sur le coût des déformations ; on peut choisir un critère R plus général, non nécessairement découlant d'un produit scalaire, et définir :

$$-\nabla_R E(\Gamma) = \arg\min_v [\ \delta E(\Gamma, v)\ +\ R(v)\ ]$$

Pour calculer ce champ $\nabla_R E$, il peut être nécessaire dans certains cas de résoudre le problème de minimisation ci-dessus par la méthode variationnelle. Il se trouve que si R atteint son minimum global en 0, alors si l'on stoppe la minimisation en cours de route au lieu d'attendre la convergence, le champ obtenu fera quand même décroître (strictement) l'énergie, ce qui est utile en pratique.

Comme exemple on définit un critère R qui favorise les déformations "localement rigides", c'est-à-dire les déformations composées de rotations, translations, homothéties qui ne concernent qu'une partie de la courbe. Le calcul du gradient de l'approximation de la distance de Hausdorff associé à ce critère R particulier s'est révélé instable à cause d'une minimisation difficile (du critère ci-dessus), donc il serait intéressant soit de trouver une meilleure méthode de minimisation pour calculer ce gradient, soit de changer le critère de façon adéquate pour que la minimisation soit plus stable. Néanmoins même avec ce problème de stabilité les résultats obtenus sont encourageants.

Troisième partie : Statistiques et a priori de formes

Les outils développés dans les parties précédentes permettent de définir la moyenne et les modes de variations d'un ensemble de formes, ou, de façon relativement similaire, d'images. Ces statistiques sont utilisées dans des applications de classification et de segmentation.

La moyenne empirique M est définie (chapitre 9) comme le centre de masse des N formes de l'ensemble d'apprentissage. Le gradient de la distance entre la moyenne M et une des formes C_i par rapport à la moyenne est un champ de déformation défini sur M, noté α_i. Ces champs constituent la brique de base de toute la suite. Une analyse en composantes principales permet d'obtenir des modes propres de déformation.

De façon similaire (chapitre 10) l'on peut définir la moyenne d'un ensemble d'images via l'introduction de difféomorphismes. Des statistiques peuvent être obtenues sur ces difféomorphismes et sont appliquées à la reconnaissance d'expression de visages.

Des statistiques de forme calculées précédemment grâce aux champs α_i l'on peut construire un critère exprimant la "probabilité" empirique d'un champ de déformation défini sur la moyenne (chapitre 11). Ainsi pour estimer la probabilité d'une nouvelle forme l'on estimera la probabilité de son champ de déformation α associé. Ce critère permet d'avoir un a priori sur la forme du contour qui évolue lors d'une segmentation d'images, et l'on montre l'intérêt de cet a priori en pratique par un exemple de segmentation d'images.

Chapitre 9 : Statistiques de formes : moyenne empirique et modes de variation

Considérons un ensemble de N contours C_i. On définit la moyenne M de ces contours de façon classique comme le nouveau contour qui minimise $\sum_i d(M, C_i)^2$ (formule du centre de masse). Pour la distance d on choisit comme précédemment l'approximation de la distance de Hausdorff. On calcule la moyenne par descente de gradient (sur M) grâce aux outils exposés dans la partie précédente.

L'inconvénient de cette formulation (somme toute classique) est qu'il est a priori possible qu'il existe plusieurs minima locaux ou globaux. Ceci dit, d'expérience, dans les cas testés, la moyenne semble être unique.

Chaque contour C_i induit sur la courbe M un champ de déformation :

$$\alpha_i = -\nabla_M d(M, C_i)^2.$$

Comme tous ces champs α_i sont définis dans le même espace (l'espace tangent à la courbe M), il est possible de les comparer les uns aux autres et d'exprimer des statistiques. Une analyse en composantes principales effectuée sur ces champs permet en particulier d'obtenir des modes propres de déformation, notés β_k, qui ont la particuliarité d'être orthogonaux entre eux et d'avoir chacun une valeur propre associée σ_k. L'analyse en composantes principales repose sur l'extraction des vecteurs propres de la matrice de corrélation des champs α_i ; cette corrélation entre deux champs α_i et α_j n'est autre que le produit scalaire $\langle \alpha_i \, | \alpha_j \rangle$ (même produit scalaire que celui utilisé pour calculer les gradients α_i, de préférence).

Des exemples de moyennes et de modes sont montrés. D'un certain point de vue, le point fondamental de cette approche statistique est l'espoir que les combinaisons linéaires de champs de déformations infinitésimaux dans l'espace tangent à la courbe M font sens. L'espace tangent étant un espace vectoriel, l'approche est validée.

Chapitre 10 : Statistiques d'images et classification d'objets

De manière similaire au chapitre précédent, on peut s'intéresser au problème de la définition (et du calcul) de la moyenne et des modes de déformation caractéristiques d'un ensemble donné de N images I_i (et non pas de contours, à la différence du reste de la thèse). A chaque image I_i on associe un difféomorphisme h_i, et l'on considère la corrélation croisée locale comme critère de ressemblance de deux images ainsi déformées $LCC(I_i \circ h_i, I_j \circ h_j)$. On cherche à maximiser la ressemblance entre toutes les paires d'images déformées par descente de gradient sur les déformations h_i. Pour cela on prend également en compte un critère de régularité $\sum_i R(h_i)$ afin que les déformations restent régulières.

A la convergence, les images $I_i \circ h_i$ observées se correspondent pixel à pixel, et l'on définit alors la moyenne de l'ensemble des images I_i comme la moyenne (pixel par pixel) des images déformées $I_i \circ h_i$.

L'approche est testée sur la base de visage de Yale et les résultats montrent que le gain en netteté (par rapport à une moyenne directe sans déformations préalables) est énorme.

On calcule alors des statistiques sur les déformations spatiales h_i (via une ACP comme précédemment) ainsi que sur les variations d'intensité $I_i \circ h_i - M$, voire sur les deux en même temps. Sur la base de visages, les modes obtenus paraissent sensés.

On utilise la moyenne M pour effectuer une tâche de classification de visages selon leur expression (joie, surprise, tristesse, etc.). Pour ce faire on applique SVM (Support vector machine) aux déformations spatiales obtenues lors de la mise en correspondance de M et des images. La classification qui en résulte est bonne (24 erreurs sur 65 sachant qu'un tirage aléatoire ferait 52 erreurs). Pour mieux faire, on suppose que l'on dispose d'une information supplémentaire, à savoir que l'on connait non seulement l'image à classer mais également le même visage avec une expression neutre. On calcule la déformation qui envoie le visage neutre sur le visage expressif et l'on ramène cette déformation sur la moyenne grâce au calcul de la déformation envoyant la moyenne sur le visage neutre. Appliquer SVM à ces nouvelles déformations fait descendre le nombre d'erreurs à 12.

Un point méritant d'être étudié davantage ici est l'incorporation des statistiques obtenues par ACP dans la tâche de classification.

Chapitre 11 : Segmentation d'images avec a priori de forme

On reprend maintenant les statistiques de formes exprimées au chapitre 9 et l'on en déduit un a priori de forme pour la segmentation d'images.

L'a priori étudié ici consiste à attribuer à un contour C une probabilité en fonction de la plausibilité du champ de déformation instantané α de la moyenne M vers C : $\alpha = -\nabla_M d^2(M, C)$.

Etant donné les modes de déformations β_k et leur écart-types associés σ_k, un critère simple est le suivant :

$$P(C) = P(\alpha) = \prod_k e^{-\frac{\langle \beta_k | \alpha \rangle_{L^2}^2}{2\sigma_k^2}} \times e^{-\frac{\|\text{Rem.}(\alpha)\|_2^2}{2\sigma_{\text{noise}}^2}}$$

où $\text{Rem.}(\alpha) = \alpha - \sum_k \langle \alpha | \beta_k \rangle_{L^2}$ est la partie de α qui ne peut s'expliquer par les modes propres, et où σ_{noise} est un paramètre décrivant l'amplitude typique de ce bruit.

La minimisation de ce critère par rapport à la courbe C implique notamment le calcul de la dérivée de α par rapport à C, qui n'est autre que $\nabla_C \nabla_M d^2(M, C)$, la dérivée seconde croisée de la distance. Dans le cas de l'approximation de la distance de Hausdorff, cette dérivée seconde est particulièrement fastidieuse à obtenir et le résultat fait plusieurs pages.

Enfin, il serait souhaitable que ce critère soit invariant par rapport aux similitudes (translations, rotations et homothéties) afin que deux mêmes formes à des emplacements différents d'une image puissent être considérées comme une unique même forme. Pour cela, on remplace dans tous les critères définis précédemment toute occurrence d'un contour C par l'infimum sur toutes les similitudes possibles $\inf_R ... R(C)$ de ce critère. En pratique on associe à toute courbe une similitude, et aux dérivées par rapport aux courbes (lors des minimisations) on ajoute une dérivée par rapport à la similitude associée. Ainsi la définition de la moyenne M d'un ensemble de courbes C_i devient la courbe M qui minimise

$$\sum_i \inf_{R_i} d\big(M, R_i(C_i)\big)^2$$

et de même le critère de forme $P(C)$ devient $\inf_R P\big(R(C)\big)$.

L'intérêt de ce critère est montré par la segmentation d'une image comportant une étoile de mer, connaissant un échantillon de 12 contours d'autres étoiles de mer. L'a priori sur la forme permet en effet de ne pas suivre certaines déformations demandées par le critère de segmentation (basé sur l'homogénéité des histogrammes d'intensité des deux régions délimitées par le contour) mais non caractéristiques d'une étoile de mer.

Conclusion

La définition et l'étude d'un ensemble de "toutes les formes" a permis d'établir l'équivalence topologique des métriques habituelles sur les formes. En s'appuyant sur une approximation dérivable de l'une d'entre elles, la distance de Hausdorff, on a pu construire des chemins (par descente de gradient) reliant une forme initiale à une forme cible. Pour cela il a fallu définir le gradient d'une énergie dépendant d'une forme, et il est apparu que le produit

scalaire auquel est lié ce gradient joue un rôle primordial dans le type d'évolution obtenue. Le gradient de l'approximation de la distance est ensuite considéré comme une brique de base pour définir les notions de moyenne, modes caractéristiques d'un ensemble de contours, et pour construire un critère d'a priori sur la forme pour la segmentation d'images. Une parenthèse a été également ouverte pour exprimer des statistiques sur des images. Une piste intéressante serait de concilier les statistiques sur les images et sur les courbes pour à la fois segmenter et classer une image en disposant ainsi d'un maximum d'information a priori.

Part I

Shapes and Distances

Chapter 1

Shape Sets and Basic Distances

Abstract

This chapter is dedicated to the definition of relevant sets of shapes and that of defining a metric on them. Following a recent research monograph by Delfour and Zolesio [33], we consider the characteristic functions of the subsets of $\mathbb{R}^2$ and their distance functions. The L^2 norm of the difference of characteristic functions, the L^∞ and the $W^{1,2}$ norms of the difference of distance functions define interesting topologies, in particular the well-known Hausdorff distance. Because of practical considerations arising from the fact that we deal with image shapes defined on finite grids of pixels we restrict our attention to subsets of $\mathbb{R}^2$ of positive reach in the sense of Federer [45], with smooth boundaries of bounded curvature. For this particular set of shapes we show that the three previous topologies are equivalent.

1.1 Introduction

Learning shape models from examples, using them to recognize new instances of the same class of shapes are fascinating problems that have attracted the attention of many scientists for many years. Central to this problem is the notion of a random shape which in itself has occupied people for decades. Frechet [48] is probably one of the first mathematicians to develop some interest for the analysis of random shapes, i.e. curves. He was followed by Matheron [76] who founded with Serra the french school of mathematical morphology and by David Kendall [57, 65, 66] and his colleagues, e.g. Small [104]. In addition, and independently, a rich body of theory and practice for the statistical analysis of shapes has been developed by Bookstein [10], Dryden and Mardia [39], Carne [13], Cootes, Taylor and colleagues [26]. Except for the mostly theoretical work of Frechet and Matheron, the tools developed by these authors are very much tied to the point-wise representation of the shapes they study: objects are represented by a finite number of salient points or landmarks. This is an important difference with our work which deals explicitly with curves as such, independently of their sampling or even parametrization.

In effect, our work bears more resemblance with that of several other authors. Like in Grenander's theory of patterns [53, 54], we consider shapes as points of an infinite dimensional manifold but we do not model the variations of the shapes by the action of Lie groups on this manifold, except in the case of such finite-dimensional Lie groups as rigid displacements (translation and rotation) or affine transformations (including scaling). For infinite dimensional groups such as diffeomorphisms [41, 112] which smoothly change the objects'

shapes previous authors have been dependent upon the choice of parameterizations and origins of coordinates [118, 119, 114, 113, 81, 55]. For them, warping a shape onto another requires the construction of families of diffeomorphisms that use these parameterizations. Our approach, based upon the use of the distance functions, does not require the arbitrary choice of parameterizations and origins. From our viewpoint this is already very nice in two dimensions but becomes even nicer in three dimensions and higher where finding parameterizations and tracking origins of coordinates can be a real problem: this is not required in our case. Another piece of related work is that of Soatto and Yezzi [105] who tackle the problem of jointly extracting and characterizing the motion of a shape and its deformation. In order to do this they find inspiration in the above work on the use of diffeomorphisms and propose the use of a distance between shapes (based on the set-symmetric difference described in section 1.2.2). This distance poses a number of problems that we address in the same section where we propose two other distances which we believe to be more suitable. They also use a signed distance score but it is non-symmetric with respect to the two regions and is not an approximation to a distance.

Some of these authors have also tried to build a Riemannian structure on the set of shapes, i.e. to go from an infinitesimal metric structure to a global one. The infinitesimal structure is defined by an inner product in the tangent space (the set of normal deformation fields) and has to vary continuously from point to point, i.e. from shape to shape. The Riemannian metric is then used to compute geodesic curves between two shapes: these geodesics define a way of warping either shape onto the other. The distance between the shapes is then given by the length of the geodesic path. This is dealt with in the work of Trouvé and Younes [118, 119, 112, 114, 113, 120] and, more recently, in the work of Klassen and Srivastava [68], again at the cost of working with parameterizations. The problem with these approaches, beside that of having to deal with parameterizations of the shapes, is that there exist global metric structures on the set of shapes (see section 1.2.2) which are useful and relevant to the problem of the comparison of shapes but that do not derive from an infinitesimal structure. Our approach can be seen as taking the problem from exactly the opposite viewpoint from the previous one: we start with a global metric on the set of shapes and build smooth functions (in effect smooth approximations of these metrics) that are dissimilarity measures, or energy functions; we then minimize these functions using techniques of the calculus of variation by computing their gradient and performing infinitesimal gradient descent (see part II): this minimization defines another way of warping either shape onto the other. In this endeavor we build on the seminal work of Delfour and Zolesio who have introduced new families of sets, complete metric topologies, and compactness theorems. This work is now available in book form [33]. The book provides a fairly broad coverage and a synthetic treatment of the field along with many new important results, examples, and constructions which have not been published elsewhere. Its full impact on image processing and robotics has yet to be fully assessed.

Section 1.2 sets the stage and introduces some notations and tools. In particular in section 1.2.2 we discuss three of the main topologies that can be defined on sets of shapes and motivate the choice of two of them. In section 1.3 we introduce the particular set of shapes we work with in this chapter, show that it has nice compactness properties and that the three topologies defined in the previous section are in fact equivalent on this set of shapes.

1.2 Shapes and shape topologies

To define fully the notion of a shape is beyond the scope of this chapter in which we use a limited, i.e purely *geometric*, definition. It could be argued that the perceptual shape of an object also depends upon the distribution of illumination, the reflectance and texture of its surface; these aspects are not discussed here. In our context we define a shape to be a measurable subset of $\mathbb{R}^2$. Since we are driven by image applications we also assume that all our shapes are contained in a hold-all open bounded subset of $\mathbb{R}^2$ which we denote by D. The reader can think of D as the "image".

In the next section we will restrict our interest to a more limited set of shapes but presently this is sufficient to allow us to introduce some methods for representing shapes.

1.2.1 Definitions

Since, as mentioned in the introduction, we want to be independent of any particular parametrization of the shape, we use two main ingredients, the *characteristic function* of a shape Ω

$$\chi_\Omega(x) = 1 \quad \text{if} \quad x \in \Omega \quad \text{and} \quad 0 \quad \text{if} \quad x \notin \Omega,$$

and the *distance function* to a shape Ω

$$d_\Omega(x) = \inf_{y \in \Omega} |y - x| = \inf_{y \in \Omega} d(x, y) \quad \text{if} \quad \Omega \neq \emptyset \quad \text{and} \quad +\infty \quad \text{if} \quad \Omega = \emptyset.$$

Note the important property [33, chapter 4, theorem 2.1]:

$$d_{\Omega_1} = d_{\Omega_2} \Longleftrightarrow \overline{\Omega}_1 = \overline{\Omega}_2 \tag{1.1}$$

Also of interest is the distance function to the complement of the shape, $d_{\complement\Omega}$ and the distance function to its boundary, $d_{\partial\Omega}$. In the case where $\Omega = \partial\Omega$ and Ω is closed, we have

$$d_\Omega = d_{\partial\Omega} \quad d_{\complement\Omega} = 0$$

We note $C_d(D)$ the set of distance functions of nonempty sets of D. Similarly, we note $C_d^c(D)$ the set of distance functions to the complements of open subsets of D (for technical reasons which are irrelevant here, it is sufficient to consider open sets).

Another function of great interest is the *oriented distance function* b_Ω defined as

$$b_\Omega = d_\Omega - d_{\complement\Omega}$$

Note that for closed sets such that $\Omega = \partial\Omega$, one has $b_\Omega = d_\Omega$.

We briefly recall some well known results about these two functions. The integral of the characteristic function is equal to the measure (area) $m(\Omega)$ of Ω:

$$\int_\Omega \chi_\Omega(x)\ dx = m(\Omega)$$

Note that this integral does not change if we add to or subtract from Ω a measurable set of Lebesgue measure 0 (also called a negligible set).

Concerning the distance functions, they are continuous, in effect Lipschitz continuous with a Lipschitz constant equal to 1 [31, 33]:

$$|d_\Omega(x) - d_\Omega(y)| \leq |x - y| \quad \forall x, y, \in D.$$

Thanks to the Rademacher theorem [42], this implies that d_Ω is differentiable almost everywhere in D, i.e. outside of a negligible set, and that the magnitude of its gradient, when it exists, is less than or equal to 1

$$|\nabla d_\Omega(x)| \leq 1 \quad \text{a.e..}$$

The same is true of $d_{\complement\Omega}$ and b_Ω (if $\partial\Omega \neq \emptyset$ for the second), [33, Chapter 5, theorem 2.1].

Closely related to the various distance functions (more precisely to their gradients) are the projections associated with $\overline{\Omega}$ and $\overline{\complement\Omega}$. These are also related to the notion of skeleton. We recall some definitions. The first one is adapted from [33, Chapter 4 definition 3.1]:

Definition 1 (Projections and skeletons).

- *Given $\Omega \subset D$, $\Omega \neq \emptyset$ (resp. $\complement\Omega \neq \emptyset$), the set of projections of $x \in D$ on Ω (resp. on $\complement\Omega$) is given by*

$$\Pi_\Omega(x) \stackrel{def}{=} \{p \in \overline{\Omega} : |p - x| = d_\Omega(x)\}$$
$$(resp.\, \Pi_{\complement\Omega}(x) \stackrel{def}{=} \{p \in \overline{\complement\Omega} : |p - x| = d_{\complement\Omega}(x)\})$$

  The elements of $\Pi_\Omega(x)$ (resp. $\Pi_{\complement\Omega}(x)$) are called projections onto $\overline{\Omega}$ (resp. $\overline{\complement\Omega}$).

- *Given $\Omega \subset D$, $\Omega \neq \emptyset$ (resp. $\complement\Omega \neq \emptyset$), the set of points where the projection on Ω (resp. $\complement\Omega$) is not unique is called the exterior (resp. interior) skeleton $Sk_{ext}(\Omega)$ (resp. $Sk_{int}(\Omega)$). We define $Sk(\Omega) = Sk_{ext}(\Omega) \cup Sk_{int}(\Omega)$.*

The following properties of the skeletons can be found e.g. in [33, Chapter 4, theorems 3.1 and 3.2]

Proposition 2. *The exterior (resp. interior) skeleton is exactly the subset of $\complement\overline{\Omega}$ (resp. of $int(\Omega)$) where the function d_Ω (resp. $d_{\complement\Omega}$) is not differentiable. Moreover the exterior and interior skeletons and the boundary $\partial\Omega$ is exactly the subset of D where $d_{\partial\Omega}$ is not differentiable.*

At each x of $\complement\Omega \backslash \mathrm{Sk}_{\mathrm{ext}}(\Omega)$, the gradient of the distance function $d_{\partial\Omega}$ is well-defined, of unit norm, and points away from the projection $y = \Pi_\Omega(x)$ of x onto Ω, see figure 1.1. Similar considerations apply to the case where $x \in \Omega$.

We introduce an additional definition that will be useful in the sequel.

Definition 3. *Given $\Omega \subset D$, $\Omega \neq \emptyset$, and a real number $h > 0$, the h-tubular neighborhood of Ω is defined as*

$$U_h(\Omega) \stackrel{def}{=} \{y \in D : d_\Omega(y) < h\}$$

1.2.2 Some shape topologies

The next question we want to address is that of the definition of the similarity between two shapes. This question of similarity is closely connected to that of metrics on sets of shapes which in turn touches that of what is known as shape topologies. We now briefly review three main similarity measures between shapes which turn out to define three distances.

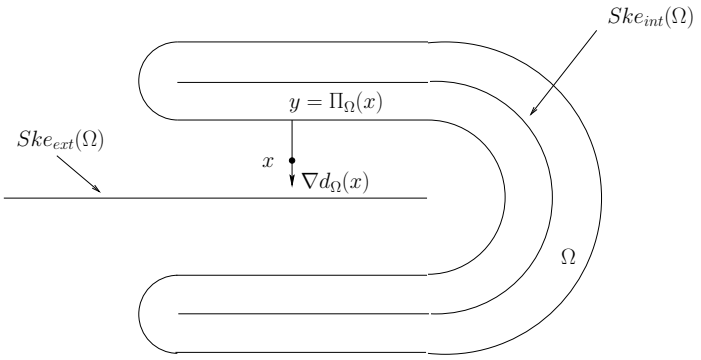


Figure 1.1: An example of skeletons.

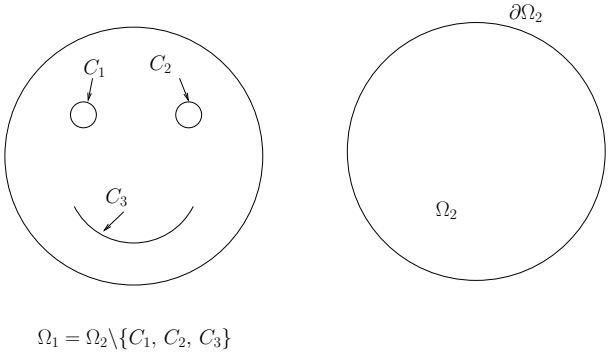


Figure 1.2: Two shapes whose distance ρ_2 is equal to 0; Ω_1 is obtained by removing from the disk Ω_2 the three curves C_1, C_2, C_3: $\rho_2(\Omega_1, \Omega_2) = 0$

Characteristic functions

The similarity measure we are about to define is based upon the characteristic functions of the two shapes we want to compare. We denote by $X(D)$ the set of characteristic functions of measurable subsets of D.

Given two such sets Ω_1 and Ω_2, we define their distance

$$\rho_2(\Omega_1, \Omega_2) \;=\; \|\chi_{\Omega_1} - \chi_{\Omega_2}\|_{L^2} \;=\; \left(\int_D \left(\chi_{\Omega_1}(x) - \chi_{\Omega_2}(x)\right)^2 \; dx\right)^{1/2}$$

This definition also shows that this measure does not "see" differences between two shapes that are of measure 0 (see figure 1.2 adapted from [33, Chapter 3, Figure 3.1]) since the integral does not change if we modify the values of χ_{Ω_1} or χ_{Ω_2} over negligible sets. In other words, this is not a distance between the two shapes Ω_1 and Ω_2 but between their equivalence classes $[\Omega_1]_m$ and $[\Omega_2]_m$ of measurable sets. Given a measurable subset Ω of D, we define its equivalence class $[\Omega]_m$ as $[\Omega]_m = \{\Omega' | \Omega'$ is measurable and $\Omega \,\Delta\, \Omega'$ is negligible$\}$, where $\Omega \,\Delta\, \Omega'$ is the symmetric difference

$$\Omega \Delta \Omega' \;=\; \complement_\Omega \Omega' \cup \complement_{\Omega'} \Omega.$$

The proof that this defines a distance follows from the fact that the L^2 norm defines a distance over the set of equivalence classes of square integrable functions (see e.g. [100, 42]).

This is nice and one has even more ([33, Chapter 3, Theorem 2.1]): the set $X(D)$ is closed and bounded in $L^2(D)$ and $\rho_2(\cdot, \cdot)$ defines a complete metric structure on the set of equivalence classes of measurable subsets of D. Note that ρ_2 is closely related to the symmetric difference:

$$\rho_2(\Omega_1, \Omega_2) = m(\Omega_1 \Delta \Omega_2)^{\frac{1}{2}} \tag{1.2}$$

The completeness is important in applications: any Cauchy sequence of characteristic functions $\{\chi_{\Omega_n}\}$ converges for this distance to a characteristic function χ_Ω of a limit set Ω. Unfortunately in applications not all sequences are Cauchy sequences, for example the minimizing sequences of the energy functions defined in section 2.1, and one often requires more, i.e. that any sequence of characteristic functions contains a subsequence that converges to a characteristic function. This stronger property, called *compactness*, *is not* satisfied by $X(D)$ (see [33, Chapter 3]).

Distance functions

We therefore turn ourselves toward a different similarity measure which is based upon the distance function to a shape. As in the case of characteristic functions, we define equivalent sets and say that two subsets Ω_1 and Ω_2 of D are equivalent iff $\overline{\Omega}_1 = \overline{\Omega}_2$. We note $[\Omega]_d$ the corresponding equivalence class of Ω. Let $\mathcal{T}(D)$ be the set of these equivalence classes. The application

$$[\Omega]_d \to d_\Omega \quad \mathcal{T}(D) \to C_d(D) \subset C(\overline{D})$$

is injective according to (1.1). We can therefore identify the set $C_d(D)$ of distance functions to sets of D with the just defined set of equivalence classes of sets. Since $C_d(D)$ is a subset of the set $C(\overline{D})$ of continuous functions on $\overline{D}$, a Banach space[1] when endowed with the norm

$$\|f\|_{C(D)} = \sup_{x \in D} |f(x)|,$$

it can be shown (e.g. [33]), that the similarity measure

$$\rho([\Omega_1]_d, [\Omega_2]_d) = \|d_{\Omega_1} - d_{\Omega_2}\|_{C(D)} = \sup_{x \in D} |d_{\Omega_1}(x) - d_{\Omega_2}(x)|, \tag{1.3}$$

is a distance on the set of equivalence classes of sets which induces on this set a complete metric. Moreover, because we have assumed D bounded, the corresponding topology is identical to the one induced by the well-known Hausdorff metric (see [76, 102, 33])

$$\rho_H([\Omega_1]_d, [\Omega_2]_d) = \max\left\{ \sup_{x \in \Omega_2} d_{\Omega_1}(x), \sup_{x \in \Omega_1} d_{\Omega_2}(x) \right\}. \tag{1.4}$$

In fact we have even more than the identity of the two topologies, see [33, Chapter 4, Theorem 2.2]:

Proposition 4. *If the hold-all set D is bounded $\rho = \rho_H$.*

An important improvement with respect to the situation in the previous section is the (see [33, Chapter 4, Theorem 2.2])

[1] A Banach space is a complete normed vector space.

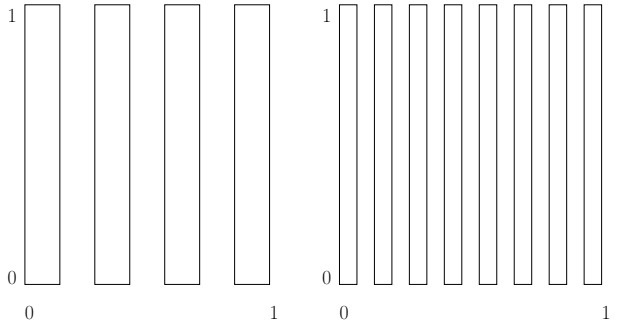


Figure 1.3: Two shapes in the sequence $\{\Omega_n\}$, see text: (left) Ω_4 and (right), Ω_8.

Theorem 5. *The set $C_d(D)$ is compact in the set $C(\overline{D})$ for the topology defined by the Hausdorff distance.*

In particular, from any sequence $\{d_{\Omega_n}\}$ of distance functions to sets Ω_n one can extract a sequence converging toward the distance function d_Ω to a subset Ω of D.

It would appear that we have reached an interesting stage and that the Hausdorff distance is what we want to measure shape similarities. Unfortunately this is not so because the convergence of areas and perimeters is lost in the Hausdorff metric, as shown in the following example taken from [33, Chapter 4, Example 4.1 and Figure 4.3].

Consider the sequence $\{\Omega_n\}$ of sets in the open square $]-1,2[^2$:

$$\Omega_n = \{(x,y) \in D : \frac{2k}{2n} \leq x \leq \frac{2k+1}{2n},\ 0 \leq k < n\}$$

Figure 1.3 shows the sets Ω_4 and Ω_8. This defines n vertical stripes of equal width $1/2n$ each distant of $1/2n$. It is easy to verify that, for all $n \geq 1$, $m(\Omega_n) = 1/2$ and $|\partial\Omega_n| = 2n+1$. Moreover, if S is the unit square $[0,1]^2$, for all $x \in S$, $d_{\Omega_n}(x) \leq 1/4n$, hence $d_{\Omega_n} \to d_S$ in $C(\overline{D})$. The sequence $\{\Omega_n\}$ converges to S for the Hausdorff distance but since $m(\overline{\Omega}_n) = m(\Omega_n) = 1/2 \nrightarrow 1 = m(S)$, $\chi_{\Omega_n} \nrightarrow \chi_S$ in $L^2(D)$ and hence we do not have convergence for the ρ_2 topology. Note also that $|\partial\Omega_n| = 2n+1 \nrightarrow |\partial S| = 4$.

Distance functions and their gradients

In order to recover continuity of the area one can proceed as follows. If we recall that the gradient of a distance function is of magnitude equal to 1 except on a subset of measure 0 of D, one concludes that it is square integrable on D. Hence the distance functions and their gradients are square-integrable, they belong to the Sobolev space $W^{1,2}(D)$, a Banach space for the vector norm

$$\|f-g\|_{W^{1,2}(D)} = \|f-g\|_{L^2(D)} + \|\nabla f - \nabla g\|_{\mathbf{L}^2(D)},$$

where $\mathbf{L}^2(D) = L^2(D) \times L^2(D)$. This defines a similarity measure for two shapes

$$\rho_D([\Omega_1]_d, [\Omega_2]_d) = \|d_{\Omega_1} - d_{\Omega_2}\|_{W^{1,2}(D)}$$

which turns out to define a complete metric structure on $\mathcal{T}(D)$. The corresponding topology is called the $W^{1,2}$-topology. For this metric, the set $C_d(D)$ of distance functions is closed in $W^{1,2}(D)$, and the mapping

$$d_\Omega \to \chi_{\overline{\Omega}} = 1 - |\nabla d_\Omega| : C_d(D) \subset W^{1,2}(D) \to L^2(D)$$

is "Lipschitz continuous":

$$\|\chi_{\overline{\Omega}_1} - \chi_{\overline{\Omega}_2}\|_{L^2(D)} \leq \|\nabla d_{\Omega_1} - \nabla d_{\Omega_2}\|_{\mathbf{L}^2(D)} \leq \|d_{\Omega_1} - d_{\Omega_2}\|_{W^{1,2}(D)}, \tag{1.5}$$

which indeed shows that areas are continuous for the $W^{1,2}$-topology, see [33, Chapter 4, Theorem 4.1].

$C_d(D)$ is not compact for this topology but a subset of it of great practical interest is, see section 1.3.

1.3 The set $\mathcal{S}$ of all shapes and its properties

We now have all the necessary ingredients to be more precise in the definition of shapes.

1.3.1 The set of all shapes

We restrict ourselves to sets of D with compact boundary and consider three different sets of shapes. The first one is adapted from [33, Chapter 4, definition 5.1]:

Definition 6 (Set $\mathcal{DZ}$ of sets of bounded curvature). *The set $\mathcal{DZ}$ of sets of bounded curvature contains those subsets Ω of $\overline{D}$, Ω, $\complement\Omega \neq \emptyset$ such that ∇d_Ω and $\nabla d_{\complement\Omega}$ are in $BV(D)^2$, where $BV(D)$ is the set of functions of bounded variations.*

This is a large set (too large for our applications) which we use as a "frame of reference". $\mathcal{DZ}$ was introduced by Delfour and Zolésio [31, 32] and contains the sets $\mathcal{F}$ and $\mathcal{C}_2$ introduced below. For technical reasons related to compactness properties (see section 1.3.2) we consider the following subset of $\mathcal{DZ}$.

Definition 7 (Set $\mathcal{DZ}_0$). *The set $\mathcal{DZ}_0$ is the subset of $\mathcal{DZ}$ such that there exists $c_0 > 0$ such that for all $\Omega \in \mathcal{DZ}_0$,*

$$\|D^2 d_\Omega\|_{M^1(D)} \leq c_0 \text{ and } \|D^2 d_{\complement\Omega}\|_{M^1(D)} \leq c_0$$

$M^1(D)$ is the set of bounded measures on D and $\|D^2 d_\Omega\|_{M^1(D)}$ is defined as follows. Let $\mathbf{\Phi}$ be a 2×2 matrix of functions in $C^1(D)$, we have

$$\|D^2 d_\Omega\|_{M^1(D)} = \sup_{\mathbf{\Phi} \in C^1(D)^{2\times 2},\, \|\mathbf{\Phi}\|_C \leq 1} \left| \int_D \nabla d_\Omega \cdot \mathbf{div}\mathbf{\Phi}\, dx \right|,$$

where

$$\|\mathbf{\Phi}\|_C = \sup_{x \in D} |\mathbf{\Phi}(x)|_{\mathbb{R}^{2\times 2}},$$

and

$$\mathbf{div}\mathbf{\Phi} = [div\Phi_1,\, div\Phi_2],$$

where Φ_i, $i = 1, 2$ are the row vectors of the matrix $\mathbf{\Phi}$.

The set $\mathcal{DZ}_0$ has the following property (see [33, Chapter 4, Theorem 5.2])

Proposition 8. *Any $\Omega \in \mathcal{DZ}_0$ has a finite perimeter upper-bounded by $2c_0$.*

We next introduce three related sets of shapes.

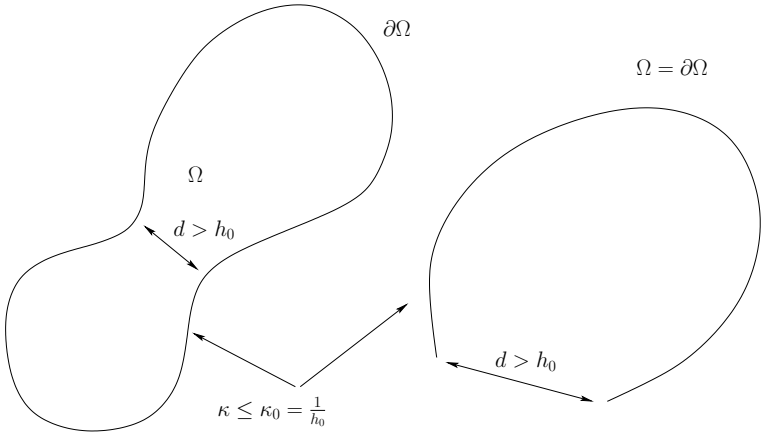


Figure 1.4: Examples of admissible shapes: a simple, closed, regular curve (left); a simple, open regular curve (right). In both cases the curvature is upperbounded by κ_0 and the pinch distance is larger than h_0.

Definition 9 (Sets of smooth shapes). *The set $\mathcal{C}_0$ (resp. $\mathcal{C}_1$, $\mathcal{C}_2$) of smooth shapes is the set of subsets of D whose boundary is non-empty and can be locally represented as the graph of a C^0 (resp. C^1, C^2) function. One further distinguishes the sets $\mathcal{C}_i^c$ and $\mathcal{C}_i^o$, $i = 0, 1, 2$ of subsets whose boundary is closed and open[2], respectively.*

Note that this implies that the boundary is a simple regular curve (hence compact) since otherwise it cannot be represented as the graph of a C^0 (resp. C^1, C^2) function in the vicinity of a multiple point. When $\Omega \in \mathcal{C}_i^o$, the set is identical to its boundary: $\Omega = \partial\Omega = \Gamma$. Also note that $\mathcal{C}_2 \subset \mathcal{C}_1 \subset \mathcal{DZ}$ ([31, 32]).

The third set has been introduced by Federer [45].

Definition 10 (Set $\mathcal{F}$ of shapes of positive reach). *A nonempty subset Ω of D is said to have positive reach if there exists $h > 0$ such that $\Pi_\Omega(x)$ is a singleton for every $x \in U_h(\Omega)$. The maximum h for which the property holds is called the reach of Ω and is noted reach(Ω).*

We will also be interested in the subsets, called h_0-Federer's sets and noted $\mathcal{F}_{h_0}$, $h_0 > 0$, of $\mathcal{F}$ which contain all Federer's sets Ω such that reach$(\Omega) \geq h_0$. Note that $\mathcal{C}_i$, $i = 1, 2 \subset \mathcal{F}$ but $\mathcal{C}_i \not\subset \mathcal{F}_{h_0}$.

We are now ready to define the set of shapes of interest.

Definition 11 (Set of all shapes). *The set, noted $\mathcal{S}$, of all shapes (of interest) is the subset of $\mathcal{C}_2$ whose elements are also h_0-Federer's sets for a given and fixed $h_0 > 0$.*

$$\mathcal{S} \stackrel{def}{=} \mathcal{C}_2 \cap \mathcal{F}_{h_0}$$

This set contains the two subsets $\mathcal{S}^c$ and $\mathcal{S}^o$ obtained by considering $\mathcal{C}_2^c$ and $\mathcal{C}_2^o$, respectively.

Note that $\mathcal{S} \subset \mathcal{DZ}$. Note also that the curvature of $\partial\Omega$ is well defined and upperbounded by $1/h_0$, noted κ_0. Hence, c_0 in definition 7 can be chosen in such a way that $\mathcal{S} \subset \mathcal{DZ}_0$.

At this point, we can represent regular (i.e. C^2) simple curves with and without boundaries that do not curve or pinch too much (in the sense of κ_0 and h_0, see figure 1.4). Polygons

[2]Meaning here without and with endpoints, respectively.

and other non-smooth structures are not explicitly included in our theory which assumes smooth shapes. In practice they are, thanks to the fact that we intersect $\mathcal{C}_2$ with $\mathcal{F}_{h_0}$. If h_0 is chosen to be smaller than the smallest distance between pixels, we will not see the difference between a polygon or a non-smooth shape and its approximation by an element of $\mathcal{S}$.

There are two reasons why we choose $\mathcal{S}$ as our frame of reference. The first one is because our implementations work with discrete objects defined on an underlying discrete square grid of pixels. As a result we are not able to describe details smaller than the distance between two pixels. This is our unit, our absolute yardstick, and h_0 is chosen to be smaller than or equal to it. The second reason is that $\mathcal{S}$ is included in $\mathcal{DZ}_0$ which, as shown in section 1.3.2, is compact. This will turn out to be important when minimizing shape functionals.

In the remaining of this chapter, the distances ρ (ρ_H) and ρ_D use the distance functions of the boundaries of the sets we consider.

The question of the deformation of a shape by an element of a group of transformations could be raised at this point. What we have in mind here is the question of deciding whether a square and the same square rotated by 45 degrees are the same shape. There is no real answer to this question, more precisely the answer depends on the application. Note that the group in question can be finite dimensional, as in the case of the Euclidean and affine groups which are the most common in applications, or infinite dimensional. In this work we will, for the most part, not consider the action of groups of transformations on shapes.

1.3.2 Compactness properties

Interestingly enough, the definition of the set $\mathcal{DZ}_0$ (definition 7) implies that it is compact for all three topologies. This is the result of the following theorem whose proof can be found in [33, Chapter 4, Theorems 8.2, 8.3].

Theorem 12. *Let D be a nonempty bounded regular*[3] *open subset of $\mathbb{R}^2$ and $\mathcal{DZ}$ the set defined in definition 6. The embedding*

$$BC(D) = \{d_\Omega \in C_d(D) \cap C_d^c(D) : \nabla d_\Omega, \nabla d_{\complement\Omega} \in BV(D)^2\} \to W^{1,2}(D),$$

is compact.

This means that for any bounded sequence $\{\Omega_n\}$, $\emptyset \neq \Omega_n$ of elements of $\mathcal{DZ}$, i.e. for any sequence of $\mathcal{DZ}_0$, there exists a set $\Omega \neq \emptyset$ of $\mathcal{DZ}$ such that there exists a subsequence Ω_{n_k} such that

$$d_{\Omega_{n_k}} \to d_\Omega \quad \text{and} \quad d_{\complement\Omega_{n_k}} \to d_{\complement\Omega} \quad \text{in} \quad W^{1,2}(D).$$

Since $b_\Omega = d_\Omega - d_{\complement\Omega}$, we also have the convergence of $b_{\Omega_{n_k}}$ to b_Ω, and since the mapping $b_\Omega \to |b_\Omega| = d_{\partial\Omega}$ is continuous in $W^{1,2}(D)$ (see [33, Chapter 5, Theorem 5.1 (iv)]), we also have the convergence of $d_{\partial\Omega_{n_k}}$ to $d_{\partial\Omega}$. The convergence for the ρ_2 distance follows from equation (1.5):

$$\chi_{\Omega_{n_k}} \to \chi_\Omega \quad \text{in} \quad L^2(D),$$

and the convergence for the Hausdorff distance follows from theorem 5, taking subsequences if necessary.

In other words, the set $\mathcal{DZ}_0$ is compact for the topologies defined by the ρ_2, Hausdorff and $W^{1,2}$ distances.

[3] Regular means uniformly Lipschitzian in the sense of [33, Chapter 2, Definition 5.1].

Note that, even though $\mathcal{S} \subset \mathcal{DZ}_0$, this does not imply that it is compact for either one of these three topologies. But it does imply that its closure $\overline{\mathcal{S}}$ for each of these topologies is compact in the compact set $\mathcal{DZ}_0$.

1.3.3 Comparison between the three topologies on $\mathcal{S}$

The three topologies we have considered turn out to be closely related on $\mathcal{S}$. This is summarized in the following

Theorem 13. *The three topologies defined by the three distances ρ_2, ρ_D and ρ_H are equivalent on $\mathcal{S}^c$. The two topologies defined by ρ_D and ρ_H are equivalent on $\mathcal{S}^o$.*

This means that, for example, given a set Ω of $\mathcal{S}^c$, a sequence $\{\Omega_n\}$ of elements of $\mathcal{S}^c$ converging toward $\Omega \in \mathcal{S}^c$ for any of the three distances ρ_2, ρ (ρ_H) and ρ_D also converges toward the *same* Ω for the other two distances.

We now proceed with the proof of theorem 13. Being a bit lengthy, we have split it in a series of lemmas and propositions.

We start with a lemma.

Lemma 14. *Let $\{f_n\}$ be a sequence of uniformly Lipschitz functions $K \to \mathbb{R}^m$, K a compact of $\mathbb{R}^2$, converging for the L^2 norm toward a Lipschitz continuous function f. Then, the convergence is uniform.*

Proof. The L^2 convergence of continuous functions implies the convergence a.e.. Let us show that this implies the convergence everywhere. We note L the Lipschitz constant. Let x_0 be a point of K such that $f_n(x_0)$ does not converge toward $f(x_0)$. There exists $\varepsilon_0 > 0$ such that for all $n_0 \geq 0$, $\exists n > n_0$, $|f_n(x_0) - f(x_0)| > \varepsilon_0$.

f being continuous at x_0, there exists $\eta > 0$ such that for all y in K such that $d(x_0, y) < \eta$, $|f(y) - f(x_0)| < \varepsilon_0/3$.

Consider now the ys of K such that $d(x_0, y) < \inf(\varepsilon_0/3L, \eta)$. There exists at least one of them, noted y_0, such that $f_n(y_0)$ converges to $f(y_0)$ because the convergence is a.e..

We write

$$|f_n(x_0) - f(x_0)| \leq |f_n(x_0) - f_n(y_0)| + |f_n(y_0) - f(y_0)| + |f(y_0) - f(x_0)|$$

The first term in the right handside of this inequality is less than or equal to $\varepsilon_0/3$ because of the uniform Lipschitz hypothesis. Because $f_n(y_0)$ converges to $f(y_0)$ there exists $N_0(\varepsilon_0, y)$ such that for all $n \geq N_0$, $|f_n(y_0) - f(y_0)| \leq \varepsilon_0/3$. The third term is also less than or equal to $\varepsilon_0/3$ because of the hypothesis on f. Hence

$$|f_n(x_0) - f(x_0)| \leq \varepsilon_0 \, \forall n \geq N_0,$$

a contradiction. The sequence $\{f_n\}$ converges toward f everywhere in K and since the f_ns are uniformly Lipschitz, the convergence is uniform (see e.g. [35]). □

This lemma is useful for proving the following

Proposition 15. *In $\mathcal{S}$, the $W^{1,2}$ convergence of sequences of distance functions implies their Hausdorff convergence.*

Proof. The $W^{1,2}$ convergence implies the L^2 convergence of the distance functions. According to lemma 14 this implies the uniform convergence of the distance functions and hence the Hausdorff convergence. □

We also have the converse

Proposition 16. *In $\mathcal{S}$, the Hausdorff convergence of sequences of distance functions implies their $W^{1,2}$ convergence.*

Proof. We consider the boundary Γ of a shape Ω of $\mathcal{S}$. The inequality

$$\|d_{\Gamma_1} - d_{\Gamma_2}\|_{L^2} \leq \rho(\Gamma_1, \Gamma_2) m(D)^{1/2}$$

shows that the Hausdorff convergence implies the L^2 convergence of the distance functions. For the $W^{1,2}$ topology we also need the convergence of the L^2 norm of the gradient.

Consider a sequence $\{\Omega_n\}$ of elements of $\mathcal{S}$ whose boundaries Γ_n converge for the Hausdorff distance toward $\Gamma \in \mathcal{S}$. If we prove the convergence a.e. of $\nabla(d_{\Gamma_n} - d_\Gamma)$ to 0, the Lebesgue dominated convergence theorem will give us the L^2 convergence toward 0 since

$$|\nabla(d_{\Gamma_n} - d_\Gamma)| \leq 2 \quad \text{a.e.}$$

Because we are in $\mathcal{S}$, all skeletons are negligible (zero Lebesgue measure), [24]. Consider the union $Sk = \Gamma \cup Sk(\Gamma) \cup_n Sk(\Gamma_n) \cup_n \Gamma_n$; as a countable union of negligible sets it is negligible. Let x be a point of $D \backslash Sk$, y_n its projection on Γ_n, y its projection on Γ. According to definition 1 and proposition 2, all distance functions of interest are differentiable at x. We prove that the angle between the vectors $\overrightarrow{xy_n}$ and $\overrightarrow{xy}$ goes to 0 by proving that $y_n \to y$. By compactness of $\overline{D}$ there exists a subsequence $\{y_{n_k}\}$ of $\{y_n\}$ converging toward $z \in \Gamma$. If $z = y$ we are done. If $z \neq y$ we prove a contradiction. Indeed, since the distance is continuous

$$\lim_{k \to \infty} d(x, y_{n_k}) = d(x, y).$$

But we also have, by definition, $d(x, y_{n_k}) = d_{\Gamma_{n_k}}(x)$; since $\Gamma_{n_k} \to \Gamma$ for the Hausdorff distance, $d_{\Gamma_{n_k}} \to d_\Gamma$ everywhere in D and therefore $\lim_{k \to \infty} d_{\Gamma_{n_k}}(x) = d_\Gamma(x)$. Hence $d(x, y) = d(x, z)$ and $x \in Sk(\Gamma)$, a contradiction.

We have shown that all converging subsequences of $\{y_n\}$ converged to $z = \Pi_\Gamma(x)$. In order to conclude, we must show that the sequence $\{y_n\}$ converges to z. Indeed, let us assume that there exists a subsequence $\{y_{n_k}\}$ not converging. There exists an $\varepsilon_0 > 0$ such that there is an infinity of values of k for which y_{n_k} is outside the open disc $B(z, \varepsilon_0)$. Let us note $\{y_{n_l}\}$ the corresponding subsequence. Because of compactness again there exists a converging subsequence of $\{y_{n_l}\}$ which has to converge toward z but this is impossible since all y_{n_l} are outside of $B(z, \varepsilon_0)$. Hence the sequence $\{y_n\}$ converges toward z and we have proved that $\nabla(d_{\Gamma_n} - d_\Gamma) \to 0$ a.e.. □

We now compare the topologies induced by the ρ_2 and the Hausdorff distances. This makes only sense in $\mathcal{S}^c$. The first result is in the following

Proposition 17. *In $\mathcal{S}^c$, the Hausdorff convergence of sequences of distance functions to the boundaries implies the L^2 convergence of the corresponding characteristic functions of the sets.*

Proof. The proof is based on the proof of proposition 23 below where we show that if $\rho_H(\Gamma_1, \Gamma_2) < \varepsilon < h_0$, given a C^2 parametrization $p \in [0, 1] \to \Gamma_1(p)$ of Γ_1, we can build a C^2 parametrization $p \in [0, 1] \to \Gamma_2(p)$ of Γ_2 such that the vector $\overrightarrow{\Gamma_1(p)\Gamma_2(p)}$ is normal to Γ_2 for all p's. Let s_2 be the arc-length on Γ_2, L_2 its length. The integral

$\int_0^{L_2} \|\overrightarrow{\Gamma_1(s_2)\Gamma_2(s_2)}\|\, ds_2$ is equal to $m(\Omega_1 \Delta \Omega_2)$, hence to $(\rho_2(\Omega_1, \Omega_2))^2$ (equation (1.2)). Since $\|\overrightarrow{\Gamma_1(s_2)\Gamma_2(s_2)}\| \leq \max_p d_{\Gamma_2}(\Gamma_1(p)) \leq \rho_H(\Gamma_1, \Gamma_2)$, we have $(\rho_2(\Omega_1, \Omega_2))^2 \leq \varepsilon L_2 \leq 2\varepsilon c_0$, according to proposition 8. □

We also prove the converse in the

Proposition 18. *In $\mathcal{S}^c$, the ρ_2 convergence of sequences of characteristic functions implies the Hausdorff convergence of the distance functions of the boundaries of the corresponding sets.*

In the proof we will need the following two lemmas and proposition.

Lemma 19. *Let Γ be a C^2 curve whose curvature is upperbounded by κ_0. Let C_1 and C_2 be two points of Γ, δ the length of the curve between C_1 and C_2:*

$$0 \leq \delta - d(C_1, C_2) \leq \frac{\delta^2 \kappa_0}{2} \tag{1.6}$$

Proof. The first inequality in (1.6) follows from the fact that the straight line is the shortest path between two points in the plane.

We parameterize Γ with its arc length s. We recall the Frenet formulae

$$\frac{d\Gamma}{ds} = \mathbf{t} \quad \frac{d\mathbf{t}}{ds} = \kappa \mathbf{n} \quad \frac{d\mathbf{n}}{ds} = -\kappa \mathbf{t},$$

where $\mathbf{t}$ and $\mathbf{n}$ are the unit tangent and normal vectors to Γ, respectively. We then write the second order Taylor expansion without remainder of $\Gamma(s_2) = C_2$ at $\Gamma(s_1) = C_1$

(1.7)

$$C_2 = C_1 + (s_2 - s_1)\mathbf{t}(s_1) + (s_2 - s_1)^2 \int_0^1 (1 - \zeta)\, \kappa(s_1 + \zeta(s_2 - s_1))\, \mathbf{n}(s_1 + \zeta(s_2 - s_1))\, d\zeta.$$

The second inequality in (1.6) follows from the fact that $|\kappa| \leq \kappa_0$ and $\delta = |s_2 - s_1|$. □

An easy consequence of this lemma is the

Proposition 20. *The length of a closed curve in $\mathcal{S}$ is greater than or equal to $2h_0$.*

Proof. We use the second inequality in (1.6) with $d(C_1, C_2) = 0$ from which the conclusion follows. □

The second lemma tells us that in a disc of small enough radius we cannot have too large a piece of a boundary of an element of $\mathcal{S}$.

Lemma 21. *Let $\varepsilon > 0$ be such that $2\varepsilon\kappa_0 << 1$. Then any disc of radius ε does not contain a connected piece of boundary of an element of $\mathcal{S}$ of length greater than h_0.*

Proof. The proof follows from the previous lemma. Let us first assume that the piece in question has a boundary, hence two different endpoints C_1 and C_2. By definition, $d(C_1, C_2) \leq 2\varepsilon$. Using (1.6) we conclude that the length δ of the curve between C_1 and C_2 must satisfy

$$\frac{\delta^2 \kappa_0}{2} - \delta + 2\varepsilon \geq 0.$$

The lefthand side is a second degree polynomial in the variable δ, noted $P(\delta)$, which has two positive roots $\delta_1 \leq \delta_2$:

$$\delta_1 = h_0(1 - \sqrt{1 - 2\varepsilon\kappa_0})$$

Since $P(0) > 0$, δ can continuously vary from 0 to its maximal value, and $\delta_1 < \delta_2$, we must have $\delta \leq \delta_1$. Moreover, since $2\varepsilon\kappa_0 << 1$, $\delta_1 < h_0/2$.

Let us now assume that the connected piece does not have a boundary, hence is a closed simple curve. We choose two distinct arbitrary points C_1 and C_2 on the curve, apply the previous analysis to the each of the two connected components, and conclude that the length of the curve is less than h_0. Because of proposition 20, this is impossible. □

We now prove proposition 18.

Proof. Let Ω_1 and Ω_2 two shapes of $\mathcal{S}^c$ with boundaries Γ_1 and Γ_2, $\varepsilon > 0$ such that $\rho_2(\Omega_1, \Omega_2) \leq \varepsilon$ and $2\kappa_0\varepsilon << 1$. We assume that there exists a point A of Γ_1 such that $d_{\Gamma_2}(A) > \varepsilon$ and prove a contradiction.

Consider the open disc $B(A, \varepsilon)$ of center A and radius ε. This disc does not contain any point of Γ_2 by hypothesis, since otherwise we would have $d_{\Gamma_2}(A) \leq \varepsilon$. Moreover, the curve Γ_1 is not included in $B(A, \varepsilon)$ because of the hypothesis $2\kappa_0\varepsilon << 1$ and lemma 21, therefore there must be a strictly positive even number of points of intersection between Γ_1 and the border of $B(A, \varepsilon)$. If there are more than two, the same reasoning as in the proof of proposition 23 below shows that there is a piece of skeleton of Γ_1 within $B(A, \varepsilon)$ and hence $\Omega_1 \notin \mathcal{F}_{h_0}$.

Let A_1 and A_2 be the endpoints of the arc of Γ_1 going through A. This arc divides $B(A, \varepsilon)$ in two parts, one of them belongs to $\Omega_1 \Delta \Omega_2$. The idea is that since $2\varepsilon\kappa_0 << 1$, the arc A_1AA_2 is equivalent to a line segment and each area is approximately equal to $\pi\varepsilon^2/2$, hence $\|\chi_{\Omega_1} - \chi_{\Omega_2}\|_{L^2} \geq \varepsilon\sqrt{\pi/2} > \varepsilon$, a contradiction.

In order to prove this, we parameterize Γ_1 between A_1 and A_2 by its arc-length s and compute an upperbound on the distance of $A(s)$ to the tangent line to $\partial\Omega_1$ at A. We choose A as the origin of arclength on Γ_1 and use equation (1.7):

$$A(s) = A + s\mathbf{t}(s) + s^2 \int_0^1 (1 - \zeta)\, \kappa(\zeta s)\, \mathbf{n}(\zeta s)\, d\zeta.$$

The distance of $A(s)$ to the line $(A, \mathbf{t}(0))$ is given by

$$(A(s) - A) \cdot \mathbf{n}(0) = s^2 \int_0^1 (1 - \zeta)\, \kappa(\zeta s)\, \mathbf{n}(\zeta s) \cdot \mathbf{n}(0)\, d\zeta.$$

We obtain an upper bound on its magnitude by

$$|(A(s) - A) \cdot \mathbf{n}(0)| \leq \frac{s^2}{2}\kappa_0 \tag{1.8}$$

The upper bound is maximal for $s = s_1$ ($A_1 = A(s_1)$ and $|s_1| \stackrel{\text{def}}{=} \delta_1$) or $s = s_2$ ($A_2 = A(s_2)$ and $s_2 \stackrel{\text{def}}{=} \delta_2$). We obtain upper bounds from (1.6); δ_1 and δ_2 must satisfy

$$\delta^2 \frac{\kappa_0}{2} - \delta + \varepsilon \geq 0.$$

In order for this to be true we must have

$$0 \leq \delta \leq \frac{1}{\kappa_0}(1 - \sqrt{1 - 2\varepsilon\kappa_0}) \stackrel{\text{def}}{=} \delta_m \quad \text{or} \quad \delta \geq \frac{1}{\kappa_0}(1 + \sqrt{1 - 2\varepsilon\kappa_0})$$

The second alternative is impossible since $B(A, \varepsilon)$ cannot contain an arc whose length is larger than $1/\kappa_0$ (lemma 21). There remains only the first alternative. Returning to (1.8), we find that $\delta_m^2\kappa_0/2$ is an upper bound on the distance of $A(s)$ to the tangent. Referring to figure 1.5 we conclude that the area of interest is bounded below by

$$\frac{\pi\varepsilon^2}{2} - \varepsilon\kappa_0\delta_m^2$$

Since $2\varepsilon\kappa_0 << 1$, we have $\delta_m = \frac{1}{\kappa_0}(\varepsilon\kappa_0 + o(\varepsilon\kappa_0)$ and therefore

$$\varepsilon\kappa_0\delta_m^2 = \frac{\varepsilon}{\kappa_0}((\varepsilon\kappa_0)^2 + o((\varepsilon\kappa_0)^2) = \varepsilon^2(\varepsilon\kappa_0 + o(\varepsilon\kappa_0)).$$

The area of interest is lower bounded by

$$\varepsilon^2(\frac{\pi}{2} - \varepsilon\kappa_0 + o(\varepsilon\kappa_0)),$$

and therefore, for $\varepsilon\kappa_0$ sufficiently small, its square root is strictly larger than ε. □

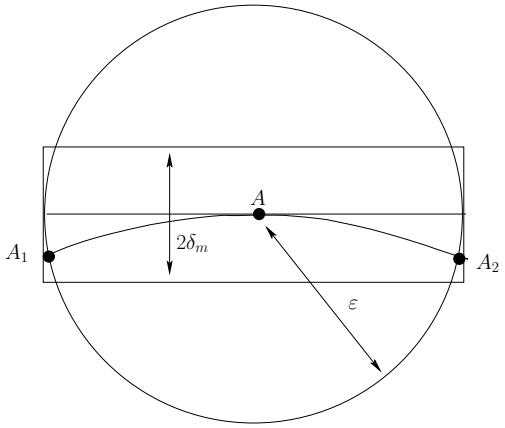


Figure 1.5: A lower bound on the area of $\Omega_1\Delta\Omega_2$ (see text).

This completes the proof of theorem 13.

1.3.4 Minima of a continuous function defined on $\mathcal{S}$

An interesting and practically important consequence of this analysis of the space $\mathcal{S}$ is the following. We know that $\mathcal{S}$ is included in $\mathcal{DZ}_0$, consider its closure $\overline{\mathcal{S}}$ for any one of the three topologies of interest. $\overline{\mathcal{S}}$ is a closed subset of the compact metric space $\mathcal{DZ}_0$ and is therefore compact as well. Given a continuous function $f : \mathcal{S} \to \mathbb{R}$ we consider its lower semi-continuous (l.s.c.) envelope $\underline{f}$ defined on $\overline{\mathcal{S}}$ as follows

$$\underline{f}(x) = \begin{cases} f(x) & \text{if} \quad x \in \mathcal{S} \\ \liminf_{y\to x,\, y\in\mathcal{S}} f(y) \end{cases}$$

The useful result for us is summarized in the

Proposition 22. *$\underline{f}$ is l.s.c. in $\overline{\mathcal{S}}$ and therefore has at least a minimum in $\overline{\mathcal{S}}$.*

Proof. In a metric space E, a real function f is said to be l.s.c. if and only if

$$f(x) \leq \liminf_{y \to x} f(y) \quad \forall x \in E.$$

Therefore $\underline{f}$ is l.s.c. by construction. The existence of minimum of an l.s.c. function defined on a compact metric space is well-known (see e.g. [25, 42]) and will be needed later to prove that some of our minimization problems are well-posed. □

Chapter 2

Shape Distance Approximations

Abstract

As in the next part we will need to take the derivative of the distances with respect to the shapes, and as the distances are not differentiable because they involve an infimum, we propose a family of smooth approximations of them which are continuous with respect to the Hausdorff topology, and hence with respect to the other two topologies.

2.1 How to approximate shape distances

The goal of this section is to provide smooth approximations of some of these distances. We start with some notations.

2.1.1 Averages

Let Γ be a given curve in $\mathcal{C}^1$ and consider an integrable function $f : \Gamma \to \mathbb{R}^n$. We denote by $\langle f \rangle_\Gamma$ the average of f along the curve Γ:

$$\langle f \rangle_\Gamma = \frac{1}{|\Gamma|} \int_\Gamma f = \frac{1}{|\Gamma|} \int_\Gamma f(x)\, d\Gamma(x) \tag{2.1}$$

For real positive integrable functions f, and for any continuous strictly monotonous (hence one to one) function φ from $\mathbb{R}^+$ or $\mathbb{R}^{+*}$ to $\mathbb{R}^+$ we will also need the φ-average of f along Γ which we define as

$$\langle f \rangle_\Gamma^\varphi = \varphi^{-1}\left(\frac{1}{|\Gamma|} \int_\Gamma \varphi \circ f\right) = \varphi^{-1}\left(\frac{1}{|\Gamma|} \int_\Gamma \varphi(f(x))\, d\Gamma(x)\right) \tag{2.2}$$

Note that φ^{-1} is also strictly monotonous and continuous from $\mathbb{R}^+$ to $\mathbb{R}^+$ or $\mathbb{R}^{+*}$. Also note that the unit of the φ-average of f is the same as that of f, thanks to the normalization by $|\Gamma|$.

The discrete version of the φ-average is also useful: let $a_i, i = 1, \ldots, n$ be n positive numbers, we note

$$\langle a_1, \cdots, a_n \rangle^\varphi = \varphi^{-1}\left(\frac{1}{n} \sum_{i=1}^{n} \varphi(a_i)\right), \tag{2.3}$$

their φ-average.

2.1.2 Approximations of the Hausdorff distance

We now build a series of smooth approximations of the Hausdorff distance $\rho_H(\Gamma, \Gamma')$ of two shapes Γ and Γ'. According to (1.4) we have to consider the functions $d_{\Gamma'} : \Gamma \to \mathbb{R}^+$ and $d_\Gamma : \Gamma' \to \mathbb{R}^+$. Let us focus on the second one. Since d_Γ is Lipschitz continuous on the bounded hold-all set D it is certainly integrable on the compact set Γ' and we have [100, Chapter 3, problem 4]

$$\lim_{\beta \to +\infty} \left(\frac{1}{|\Gamma'|} \int_{\Gamma'} d_\Gamma^\beta(x')\, d\Gamma'(x') \right)^{\frac{1}{\beta}} = \sup_{x' \in \Gamma'} d_\Gamma(x'). \tag{2.4}$$

Moreover, the function $\mathbb{R}^+ \to \mathbb{R}^+$ defined by $\beta \to \left(\frac{1}{|\Gamma'|} \int_{\Gamma'} d_\Gamma^\beta(x')\, d\Gamma'(x') \right)^{\frac{1}{\beta}}$ is monotonously increasing [100, Chapter 3, problem 5].

Similar properties hold for $d_{\Gamma'}$.

If we note p_β the function $\mathbb{R}^+ \to \mathbb{R}^+$ defined by $p_\beta(x) = x^\beta$ we can rewrite (2.4)

$$\lim_{\beta \to +\infty} \langle d_\Gamma \rangle_{\Gamma'}^{p_\beta} = \sup_{x' \in \Gamma'} d_\Gamma(x').$$

$\langle d_\Gamma \rangle_{\Gamma'}^{p_\beta}$ is therefore a monotonically increasing approximation of $\sup_{x' \in \Gamma'} d_\Gamma(x')$. We go one step further and approximate $d_{\Gamma'}(x)$.

Consider a continuous strictly monotonously decreasing function $\varphi : \mathbb{R}^+ \to \mathbb{R}^{+*}$. Because φ is strictly monotonously decreasing

$$\sup_{x' \in \Gamma'} \varphi(d(x, x')) = \varphi(\inf_{x' \in \Gamma'} d(x, x')) = \varphi(d_{\Gamma'}(x)),$$

and moreover

$$\lim_{\alpha \to +\infty} \left(\frac{1}{|\Gamma'|} \int_{\Gamma'} \varphi^\alpha(d(x, x'))\, d\Gamma'(x') \right)^{\frac{1}{\alpha}} = \sup_{x' \in \Gamma'} \varphi(d(x, x')).$$

Because φ is continuous and strictly monotonously decreasing, it is one to one and φ^{-1} is strictly monotonously decreasing and continuous. Therefore

$$d_{\Gamma'}(x) = \lim_{\alpha \to +\infty} \varphi^{-1} \left(\left(\frac{1}{|\Gamma'|} \int_{\Gamma'} \varphi^\alpha(d(x, x'))\, d\Gamma'(x') \right)^{\frac{1}{\alpha}} \right)$$

We can simplify this equation by introducing the function $\varphi_\alpha = p_\alpha \circ \varphi$:

$$d_{\Gamma'}(x) = \lim_{\alpha \to +\infty} \langle d(x, \cdot) \rangle_{\Gamma'}^{\varphi_\alpha} \tag{2.5}$$

Any $\alpha > 0$ provides us with an approximation, noted $\tilde{d}_{\Gamma'}$, of $d_{\Gamma'}$:

$$\tilde{d}_{\Gamma'}(x) = \langle d(x, \cdot) \rangle_{\Gamma'}^{\varphi_\alpha} \tag{2.6}$$

We have a similar expression for $\tilde{d}_\Gamma$.

Note that because $\left(\frac{1}{|\Gamma'|} \int_{\Gamma'} \varphi^\alpha(d(x, x'))\, d\Gamma'(x') \right)^{\frac{1}{\alpha}}$ increases with α toward its limit $\sup_{x'} \varphi(d(x, x')) = \varphi(d_{\Gamma'}(x))$, $\varphi^{-1} \left(\left(\frac{1}{|\Gamma'|} \int_{\Gamma'} \varphi^\alpha(d(x, x'))\, d\Gamma'(x') \right)^{\frac{1}{\alpha}} \right)$ decreases with α toward its limit $d_{\Gamma'}(x)$.

Examples of functions φ are

$$\begin{aligned}\varphi_1(z) &= \frac{1}{z+\varepsilon} \quad \varepsilon > 0,\, z \geq 0 \\ \varphi_2(z) &= \mu \exp(-\lambda z) \quad \mu,\, \lambda > 0,\, z \geq 0 \\ \varphi_3(z) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{z^2}{2\sigma^2}) \quad \sigma > 0,\, z \geq 0\end{aligned} \tag{2.7}$$

Putting all this together we have the following result

$$\begin{aligned}\sup_{x\in\Gamma} d_{\Gamma'}(x) &= \lim_{\alpha,\beta\to+\infty} \left\langle \langle d(\cdot,\,\cdot)\rangle_{\Gamma'}^{\varphi_\alpha} \right\rangle_{\Gamma}^{p_\beta} \\ \sup_{x\in\Gamma'} d_{\Gamma}(x) &= \lim_{\alpha,\beta\to+\infty} \left\langle \langle d(\cdot,\,\cdot)\rangle_{\Gamma}^{\varphi_\alpha} \right\rangle_{\Gamma'}^{p_\beta}\end{aligned}$$

Any positive values of α and β yield approximations of $\sup_{x\in\Gamma} d_{\Gamma'}(x)$ and $\sup_{x\in\Gamma'} d_{\Gamma}(x)$.

The last point to address is the max that appears in the definition of the Hausdorff distance. We use (2.3), choose $\varphi = p_\gamma$ and note that, for a_1 and a_2 positive,

$$\lim_{\gamma\to+\infty} \langle a_1,\, a_2\rangle^{p_\gamma} = \max(a_1,\, a_2).$$

This yields the following expression for the Hausdorff distance between two shapes Γ and Γ'

$$\rho_H(\Gamma,\,\Gamma') = \lim_{\alpha,\beta,\gamma\to+\infty} \left\langle \left\langle \langle d(\cdot,\,\cdot)\rangle_{\Gamma'}^{\varphi_\alpha} \right\rangle_{\Gamma}^{p_\beta},\, \left\langle \langle d(\cdot,\,\cdot)\rangle_{\Gamma}^{\varphi_\alpha} \right\rangle_{\Gamma'}^{p_\beta} \right\rangle^{p_\gamma}$$

This equation is symmetric and yields approximations $\tilde{\rho}_H$ of the Hausdorff distance for all positive values of α, β and γ:

$$\tilde{\rho}_H(\Gamma,\,\Gamma') = \left\langle \left\langle \langle d(\cdot,\,\cdot)\rangle_{\Gamma'}^{\varphi_\alpha} \right\rangle_{\Gamma}^{p_\beta},\, \left\langle \langle d(\cdot,\,\cdot)\rangle_{\Gamma}^{\varphi_\alpha} \right\rangle_{\Gamma'}^{p_\beta} \right\rangle^{p_\gamma} . \tag{2.8}$$

2.1.3 Continuity

This approximation is "nice" in several ways, the first one being the obvious one, stated in the following

Proposition 23. *For each triplet $(\alpha,\,\beta,\,\gamma)$ in $(\mathbb{R}^{+*})^3$ the function $\tilde{\rho}_H : \mathcal{S}\times\mathcal{S} \to \mathbb{R}^+$ defined by equation (2.8) is continuous for the Hausdorff topology.*

We first recall the following properties of the squared distance function $\eta_{\partial\Omega}$ of the boundary of an element Ω of $\mathcal{S}$ (see [3]):

Proposition 24. *$\eta_{\partial\Omega}$ is smooth, i.e. C^2, in $U_{h_0}(\partial\Omega)$ and for all $x \in \partial\Omega$, the Hessian matrix $\nabla^2\eta_{\partial\Omega}(x)$ is the (matrix of) orthogonal projection onto the normal to $\partial\Omega$ at x.*

We now prove proposition 23.

Proof. For each shape Ω of $\mathcal{S}$, we consider the square of the distance function of $\partial\Omega$, noted $\eta_{\partial\Omega}$. We next prove that the length is continuous for the Hausdorff topology on $\mathcal{S}$. Consider two shapes Ω_1 and Ω_2 of $\mathcal{S}$, their boundaries Γ_1 and Γ_2 and assume that $\rho_H(\Gamma_1,\,\Gamma_2) < \varepsilon$. Let $p \in [0,1] \to \Gamma_1(p)$ be a C^2 parametrization of Γ_1, we prove that the mapping

$$p \in [0,1] \to \Gamma_2(p) = \Gamma_1(p) - \frac{1}{2}\nabla\eta_{\Gamma_2}(\Gamma_1(p)), \tag{2.9}$$

is a one to one parametrization of Γ_2. If we choose $\varepsilon < h_0$, $\nabla\eta_{\Gamma_2}(\Gamma_1(p))$ is well-defined and continuous for all ps (proposition 24), hence $p \to \Gamma_1(p) - \frac{1}{2}\nabla\eta_{\Gamma_2}(\Gamma_1(p))$ is continuous.

It is injective: assume that there exist p_1 and p_2 in $[0, 1]$, $p_1 \neq p_2$ such that $\Gamma_2(p_1) = \Gamma_2(p_2)$, see figure 2.1 (if Γ_1 and Γ_2 are closed, $p_1, p_2 \notin \{0, 1\}$). Since the curvature of Γ_1 and Γ_2 is bounded by $1/h_0$, we choose $\varepsilon << h_0$. The two points $\Gamma_1(p_1)$ and $\Gamma_1(p_2)$ are in the disc of center $\Gamma_2(p_1)$ and radius ε since their distances to Γ_2 are by construction equal to $d(\Gamma_1(p_1), \Gamma_2(p_1))$ and $d(\Gamma_1(p_2), \Gamma_2(p_2))$, respectively, and are less than ε. Because of our choice of ε, the curvatures of the two curves within the disc are negligible and we can consider they are straight lines, as shown in figure 2.1. Therefore there must be a piece of the skeleton of Ω_1 within the disc and this contradicts the hypothesis that $\Omega_1 \in \mathcal{F}_{h_0}$.

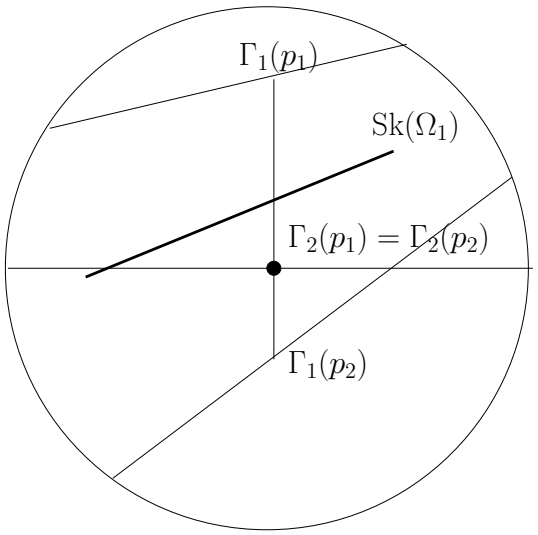


Figure 2.1: The mapping is injective.

It is surjective: we proceed by contradiction. Let us assume it is not surjective. Since the mapping (2.9) is continuous its image is connected and compact. Its complement $\hat{\Gamma}_2$ (assumed here to be non empty) is thus an open interval of Γ_2 (possibly two, if Γ_2 is an open curve). Let Γ_2^0 be one of the endpoints of this interval. There exists a value p_0 of p such that

$$\Gamma_2^0 = \Gamma_1(p_0) - \frac{1}{2}\nabla\eta_{\Gamma_2}(\Gamma_1(p_0))$$

Two cases can occur. Either $\Gamma_2^0 = \Gamma_1(p_0)$ and this implies that Γ_2 is not simple (see figure 2.2, left), or $\Gamma_2^0 \neq \Gamma_1(p_0)$ and this implies that $\Gamma_1(p_0)$ is on the skeleton of Γ_2, a contradiction if ε is small with respect to h_0 (see figure 2.2, right).

Using this parameterization, we now prove that the length is continuous for the Hausdorff metric. Given a shape Ω and a sequence $\{\Omega_n\}$ of shapes of $\mathcal{S}$ such that the boundaries Γ_n are converging to the boundary Γ of Ω for the Hausdorff topology, we show that $\lim_{n\to\infty} ||\Gamma_n| - |\Gamma|| = 0$. If n is large enough, we use the first part of the proof to parametrize Γ_n:

$$\Gamma_n(p) = \Gamma(p) - \frac{1}{2}\nabla\eta_{\Gamma_n}(\Gamma(p)), \tag{2.10}$$

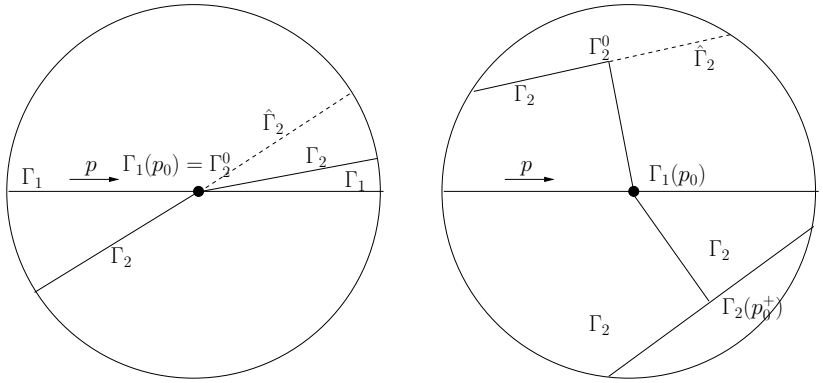


Figure 2.2: The mapping is surjective: The dotted line represents a piece of $\hat{\Gamma}_2$, see text.

and proceed from there:

$$|\,|\partial\Omega_n| - |\partial\Omega|\,| = \left| \int_0^1 |\Gamma_n'(p)|\,dp - \int_0^1 |\Gamma'(p)|\,dp \right| \leq \int_0^1 |\,|\Gamma_n'(p)| - |\Gamma'(p)|\,|\,dp \leq \int_0^1 |\Gamma_n'(p) - \Gamma'(p)|\,dp.$$

We take the derivative of (2.10) with respect to p:

$$\Gamma_n'(p) = \Gamma'(p) - \frac{1}{2}\nabla^2\eta_{\Gamma_n}(\Gamma(p))\Gamma'(p),$$

where $\nabla^2\eta_{\Gamma_n}$ is the second order derivative of η_{Γ_n}.

We are only interested in comparing the lengths of Γ and Γ_n where they differ. We can therefore exclude from the integral $\int |\Gamma_n'(p) - \Gamma'(p)|\,dp$ the values of p for which $\Gamma_n(p) = \Gamma(p)$ and assume that $\Gamma_n(p) \neq \Gamma(p)$. At these points, the first and second order derivatives of the distance function d_{Γ_n} are well-defined and (because $\Gamma_n \in \mathcal{S}$ and $\varepsilon << h_0$) there exists $M > 0$, independent of n, such that

$$|\,\nabla^2 d_{\Gamma_n}(x)\,| \leq M \,\forall x \notin \Gamma_n$$

Using the chain rule we obtain

$$\frac{1}{2}\nabla\eta_{\Gamma_n} = d_{\Gamma_n}\nabla d_{\Gamma_n} \quad \frac{1}{2}\nabla^2\eta_{\Gamma_n} = d_{\Gamma_n}\nabla^2 d_{\Gamma_n} + \nabla d_{\Gamma_n}(\nabla d_{\Gamma_n})^T,$$

and therefore

$$\begin{aligned}(2.11)\quad |\Gamma_n'(p) - \Gamma'(p)| &\leq |\nabla d_{\Gamma_n}(\Gamma(p)) \cdot \Gamma'(p)|\,|\nabla d_{\Gamma_n}(\Gamma(p))| + \\ & d_{\Gamma_n}(\Gamma(p))\,\|\nabla^2 d_{\Gamma_n}(\Gamma(p))\Gamma'(p)\| \leq |\nabla d_{\Gamma_n}(\Gamma(p)) \cdot \Gamma'(p)| + M d_{\Gamma_n}(\Gamma(p))\,|\Gamma'(p)|\end{aligned}$$

Consider the term $\nabla d_{\Gamma_n}(\Gamma(p)) \cdot \Gamma'(p)$. We write the following first order Taylor expansion without remainder

$$0 = d_{\Gamma_n}(\Gamma_n(p)) = d_{\Gamma_n}(\Gamma(p)) + \left(\int_0^1 (1-\zeta)\nabla d_{\Gamma_n}(\Gamma(p) + \zeta(\Gamma_n(p) - \Gamma(p)))\,d\zeta \right) \cdot (\Gamma_n(p) - \Gamma(p))$$

We take the derivative with respect to p:

$$\nabla d_{\Gamma_n}(\Gamma(p)) \cdot \Gamma'(p) + \left(\int_0^1 (1-\zeta) \nabla d_{\Gamma_n}(\Gamma(p) + \zeta(\Gamma_n(p) - \Gamma(p))) \, d\zeta \right) \cdot (\Gamma_n'(p) - \Gamma'(p)) +$$
$$\left(\int_0^1 (1-\zeta) \nabla^2 d_{\Gamma_n}(\Gamma(p) + \zeta(\Gamma_n(p) - \Gamma(p)))(\Gamma'(p) + \zeta(\Gamma_n'(p) - \Gamma'(p))) \, d\zeta \right) \cdot (\Gamma_n(p) - \Gamma(p)),$$

and obtain the upper bound

$$|\nabla d_{\Gamma_n}(\Gamma(p)) \cdot \Gamma'(p)| \leq \frac{1}{2} |\Gamma_n'(p) - \Gamma'(p)| + A d(\Gamma_n(p), \Gamma(p)).$$

We use it in (2.11) to obtain

$$\frac{1}{2} |\Gamma_n'(p) - \Gamma'(p)| \leq A d(\Gamma_n(p), \Gamma(p)) + M d_{\Gamma_n}(\Gamma(p)) \, |\Gamma'(p)|,$$

from which follows

$$| \, |\Gamma_n| - |\Gamma| \, | \leq 2(A + M|\Gamma|)\varepsilon. \tag{2.12}$$

We next prove that for all Lipschitz continuous functions f on D, the integral $\int_\Gamma f(x) \, d\Gamma(x)$ is continuous for the Hausdorff topology. Consider a shape Ω and a sequence $\{\Omega_n\}$ of shapes of $\mathcal{S}$ whose boundaries Γ_n are converging to the boundary Γ of Ω for the Hausdorff topology; we show that $\lim_{n \to \infty} \left| \int_{\Gamma_n} f(x) \, d\Gamma_n(x) - \int_\Gamma f(y) \, d\Gamma(y) \right| = 0$. We use once more the parametrization (2.10) and write

$$\left| \int_{\Gamma_n} f(x) \, d\Gamma_n(x) - \int_\Gamma f(y) \, d\Gamma(y) \right| =$$
$$\left| \int_0^1 \left(f(\Gamma_n(p)) |\Gamma_n'(p)| - f(\Gamma(p)) |\Gamma'(p)| \right) dp \right| \leq$$
$$\int_0^1 \left| f(\Gamma_n(p)) |\Gamma_n'(p)| - f(\Gamma(p)) |\Gamma'(p)| \right| dp \leq$$
$$\int_0^1 |f(\Gamma_n(p))| \, ||\Gamma_n'(p)| - |\Gamma'(p)|| \, dp + \int_0^1 |f(\Gamma_n(p)) - f(\Gamma(p))| \, |\Gamma'(p)| \, dp$$

f is continuous on the compact set $\overline{D}$ and is therefore upperbounded, $|f(x)| \leq K$, $\forall x \in \overline{D}$. It is also Lipschitz continuous, hence $|f(\Gamma_n(p)) - f(\Gamma(p))| \leq L d(\Gamma_n(p), \Gamma(p)) \leq L\varepsilon$. We combine this with (2.12) and obtain

$$\left| \int_{\Gamma_n} f(x) \, d\Gamma_n(x) - \int_\Gamma f(y) \, d\Gamma(y) \right| \leq \varepsilon((L + 2KM)|\Gamma| + 2KA).$$

We have used the Lipschitz hypothesis in the proof. It easy to verify that this hypothesis is satisfied since we are integrating along curves functions of the type $\varphi \circ d(\cdot, x)$. The functions φ are defined and at least C^1, hence Lipschitz continuous on $[0, diam(D)]$, where $diam(D)$ is the diameter of D. Hence $|\varphi \circ d(x_1, x) - \varphi \circ d(x_2, x)| \leq L_\varphi |d(x_1, x) - d(x_2, x)| \leq L_\varphi d(x_1, x_2)$. □

2.1.4 A possible extension of the approximation by changing the measure

The basic trick of the approximation of the Hausdorff distance relies on an integration of an application over the shape. This integration implies the choice of a particular measure on each shape, and so far was considered the canonical, intrinsic measure $d\Gamma(\mathbf{x})$ (sometimes improperly denoted by $d\mathbf{x}$ in this report for the sake of simplicity) expressing the local area of the shape Γ. However, there exist many other intrinsic measures (in the sense that they do not depend on the choice of the parameterization), such as $\left(1+\kappa(\mathbf{x})^2\right)\, d\Gamma(\mathbf{x})$ where κ is the curvature.

The new family of approximations would write, for any measure μ:

$$\langle f \rangle_{\Gamma,\Psi,\,\mu} = \Psi^{-1}\left(\frac{1}{\int_\Gamma d\mu}\int_\Gamma \Psi\circ f\; d\mu\right)$$

which rewrites, if μ is absolutely continuous with respect to the usual measure ($d\mu = q\, d\, Gamma$):

$$\langle f \rangle_{\Gamma,\Psi,\,q} = \Psi^{-1}\left(\frac{1}{\int_\Gamma q(\mathbf{x})\, d\Gamma(\mathbf{x})}\int_\Gamma q(\mathbf{x})\;\Psi\circ f(\mathbf{x})\; d\Gamma(\mathbf{x})\right)$$

However, giving to some points $\mathbf{x}$ a higher value of q means that these points are more important and have a greater impact on the integral. Since these integrals are used here to approximate the infimum of $d(\mathbf{x},\mathbf{y})$ for a fixed point $\mathbf{y}$ and a variable point $\mathbf{x}$ on the shape Γ, this is not necessarily a very good idea to give more importance to some points $\mathbf{x}$ which are not geometrically linked to $\mathbf{y}$. That is the reason why in the rest of the report the usual intrinsic measure $d\Gamma$ is considered. Note that, nevertheless, the choice of another intrinsic measure could be justified when the integral approximates the supremum of an application of $\mathbf{y}$ over the shape B. One could indeed arbitrarily choose to give more importance to the distance to Γ of the high-curvature points of the shape if the desired aim is to bring more attention to the corners of the shapes.

2.1.5 Other alternatives related to the Hausdorff distance

There exist several alternatives to the method presented in the previous sections if we use ρ (equation (1.3)) rather than ρ_H (equation (1.4)) to define the Hausdorff distance. A first alternative is to use the following approximation

$$\tilde{\rho}(\Gamma,\Gamma') = \langle |d_\Gamma - d_{\Gamma'}| \rangle_D^{p_\alpha},$$

where the bracket $\langle\, f(.)\,\rangle_D^{\varphi}$ is defined the obvious way for any integrable function $f : D \to \mathbb{R}^+$

$$\langle\, f\,\rangle_D^{\varphi} = \varphi^{-1}\left(\frac{1}{m(D)}\int_D \varphi(f(x))\, dx\right),$$

and which can be minimized, as in section 5.1.8, with respect to d_Γ. A second alternative is to approximate ρ using:

$$\tilde{\rho}(\Gamma,\,\Gamma') = \langle |\langle d(\cdot,\,\cdot)\rangle_{\Gamma'}^{\varphi_\beta} - \langle d(\cdot,\,\cdot)\rangle_{\Gamma}^{\varphi_\beta}| \rangle_D^{p_\alpha}, \tag{2.13}$$

and to compute its derivative with respect to Γ as we did in the previous section for $\tilde{\rho}_H$.

2.1.6 Approximations to the $W^{1,2}$ norm

The previous results can be used to construct approximations $\tilde{\rho}_D$ to the distance ρ_D defined in section 1.2.2:

$$\tilde{\rho}_D(\Gamma_1, \Gamma_2) = \|\tilde{d}_{\Gamma_1} - \tilde{d}_{\Gamma_2}\|_{W^{1,2}(D)}, \tag{2.14}$$

where $\tilde{d}_{\Gamma_i}$, $i = 1, 2$ is obtained from (2.6).

This approximation is also "nice" in the usual way and we have the

Proposition 25. *For each α in $\mathbb{R}^{+*}$ the function $\tilde{\rho}_D : \mathcal{S} \times \mathcal{S} \to \mathbb{R}^+$ is continuous for the $W^{1,2}$ topology.*

Its proof is left to the reader.

2.2 Quality of the approximation

Our hope is that, starting from Γ_1, we will follow the gradient (5.6) and smoothly converge to the curve Γ_2 where the minimum of $\tilde{\rho}_H$ is attained. Let us examine more closely these assumptions. First, it is clear from the expression (2.8) of $\tilde{\rho}_H$ that in general $\tilde{\rho}_H(\Gamma, \Gamma) \neq 0$, which implies in particular that $\tilde{\rho}_H$, unlike ρ_H, is not a distance. But worse things can happen: there may exist a shape Γ' such that $\tilde{\rho}_H(\Gamma, \Gamma')$ is strictly less than $\tilde{\rho}_H(\Gamma, \Gamma)$ or there may not exist any minima for the function $\Gamma \to \tilde{\rho}_H(\Gamma, \Gamma')$! This sounds like the end of our attempt to warp a shape onto another using an approximation of the Hausdorff distance. But things turn out not to be so bad. First, the existence of a minimum is guaranteed by proposition 23 which says that $\tilde{\rho}_H$ is continuous on $\mathcal{S}$ for the Hausdorff topology, theorem 12 which says that $\mathcal{DZ}_0$ is compact for this topology, and proposition 22 which tells us that the l.s.c. extension of $\tilde{\rho}_H(\cdot, \Gamma)$ has a minimum in the closure $\overline{\mathcal{S}}$ of $\mathcal{S}$ in $\mathcal{DZ}_0$.

We show in the next section that phenomena like the one described above are for all practical matters "invisible" since confined to an arbitrarily small Hausdorff ball centered at Γ.

2.2.1 Quality of the approximation $\tilde{\rho}_H$ of ρ_H

In this section we make more precise the idea that $\tilde{\rho}_H$ can be made arbitrarily close to ρ_H. Because of the form of (5.5) we seek upper and lower bounds of such quantities as $\langle f \rangle_\Gamma^\psi$, where f is a continuous real function defined on Γ. We note $f_{\max}$ and $f_{\min}$ the maximum and minimum values of f on Γ.

The expression

$$\langle f \rangle_\Gamma^\psi = \psi^{-1}\left(\frac{1}{|\Gamma|}\int_\Gamma \psi \circ f\right),$$

yields, if ψ is strictly increasing:

$$\langle f \rangle_\Gamma^\psi \leqslant \psi^{-1}\left(\frac{1}{|\Gamma|}\int_\Gamma \psi \circ f_{\max}\right) = f_{\max}$$

and, similarly:

$$\langle f \rangle_\Gamma^\psi \geqslant f_{\min}$$

If $f \geqslant f_{\text{moy}}$ on a set F of the curve Γ, of length $|F|$ $(\leqslant |\Gamma|)$:

$$\begin{aligned}
\langle f \rangle_{\Gamma}^{\psi} &= \psi^{-1}\left(\frac{1}{|\Gamma|}\int_F \psi \circ f + \frac{1}{|\Gamma|}\int_{\Gamma \setminus F} \psi \circ f\right) \\
&\geqslant \psi^{-1}\left(\frac{|F|}{|\Gamma|}\psi \circ f_{\text{moy}} + \frac{|\Gamma| - |F|}{|\Gamma|}\psi \circ f_{\min}\right) \\
&\geqslant \psi^{-1}\left(\frac{|F|}{|\Gamma|}\psi \circ f_{\text{moy}}\right)
\end{aligned}$$

To analyze this lower bound, we introduce the following notation. Given Δ, $\alpha \geqslant 0$, we note $\mathcal{P}(\Delta, \alpha)$ the following property:

$$\mathcal{P}(\Delta, \alpha): \quad \forall x \in \mathbb{R}^+,\ \Delta\psi(x) \geqslant \psi(\alpha x)$$

This property is satisfied for $\psi(x) = x^{\beta}$, $\beta \geq 0$. The best pairs (Δ, α) verifying $\mathcal{P}$ are such that $\Delta = \alpha^{\beta}$. In the sequel, we say that a function ψ is admissible for $\mathcal{P}$ if

$$\forall \Delta \in]0;1[,\ \exists \alpha \in]0;1[, \mathcal{P}(\Delta, \alpha),$$

and, conversely,

$$\forall \alpha \in]0;1[,\ \exists \Delta \in]0;1[, \mathcal{P}(\Delta, \alpha)$$

Let us assume that ψ is admissible and note that we can rewrite $\mathcal{P}(\Delta, \alpha)$

$$\forall x \in \mathbb{R}^+,\ \psi^{-1}\left(\Delta\psi(x)\right) \geqslant \alpha x.$$

For $\Delta = \frac{|F|}{|\Gamma|}$ and $x = f_{\text{moy}}$ we obtain for the largest $\alpha(\Delta)$ the following lowerbound

$$\langle f \rangle_{\Gamma}^{\psi} \geqslant \psi^{-1}\left(\Delta\psi(f_{\text{moy}})\right) \geqslant \alpha f_{\text{moy}}.$$

In words, for each arbitrary percentage Δ there exists an α such that if $|\{f \geqslant f_{\text{moy}}\}| \geqslant \Delta|\Gamma|$, then $\langle f \rangle_{\Gamma}^{\psi} \geqslant \alpha f_{\text{moy}}$. Conversely, for a given value of α, there exists a Δ such that it is sufficient that $|\{f \geqslant f_{\text{moy}}\}| \geqslant \Delta|\Gamma|$ to have $\langle f \rangle_{\Gamma}^{\psi} \geqslant \alpha f_{\text{moy}}$.

For each choice of (Δ, α), the bracket $\langle f \rangle_{\Gamma}^{\psi}$ acts as a filter which only "looks" at the values of f along Γ such that the subset F of Γ where they are reached is of relative length $\frac{|F|}{|\Gamma|} \geqslant \Delta$, meaning that one neglects the "details of relative importance $\leqslant \Delta$", and that the accuracy of the filter is relative, since it depends upon the product of α $(\leqslant 1)$ with f_{moy}.

One has even more: the above admissible family of functions ψ allows one to select an arbitrary accuracy, i.e. to choose both Δ as close as possible to 0, and α as close as possible to 1, the best pairs $(\alpha,\ \Delta)$ for $\psi(x) = x^{\beta}$ satisfying $\Delta = \alpha^{\beta}$, it is sufficient to choose β large enough.

Similar properties hold for such brackets as $\langle f \rangle_{\Gamma}^{\varphi}$ where φ is strictly decreasing. We have, as in the previous case:

$$f_{\min} \leqslant \langle f \rangle_{\Gamma}^{\varphi} \leqslant f_{\max}$$

Proceeding as before,if $|\{f \leqslant f_{\text{moy}}\}| \geqslant \Delta|\Gamma|$ and the pair (Δ, α) satisfies $\mathcal{P}$ for the function φ, we obtain:

$$\begin{aligned}
\frac{1}{|\Gamma|}\int_{\Gamma} \varphi \circ f &\geqslant \Delta\varphi(f_{\text{moy}}) \\
\langle f \rangle_{\Gamma}^{\varphi} &\leqslant \varphi^{-1}\left(\Delta\varphi(f_{\text{moy}})\right) \\
\langle f \rangle_{\Gamma}^{\varphi} &\leqslant \alpha f_{\text{moy}}
\end{aligned}$$

Admissible functions are $\varphi(x) = x^{-\beta}$, $\beta > 0$; the accuracy increases when α tends to 1^- and Δ to 0^+; this is always possible by choosing large values of β, and $\Delta = \alpha^{-\beta}$.

We now have all the ingredients for comparing $\tilde{\rho}_H$ and ρ_H. We start with two definitions.

Definition 26. *Let Γ be a shape. For each point P of D we note (see figure 2.3):*

$$d_\Delta(P,\Gamma) = \inf\left\{x \in \mathbb{R}^+; |\{Q \in \Gamma; d(P,Q) \leqslant x\}| \geqslant \Delta|\Gamma|\right\}$$

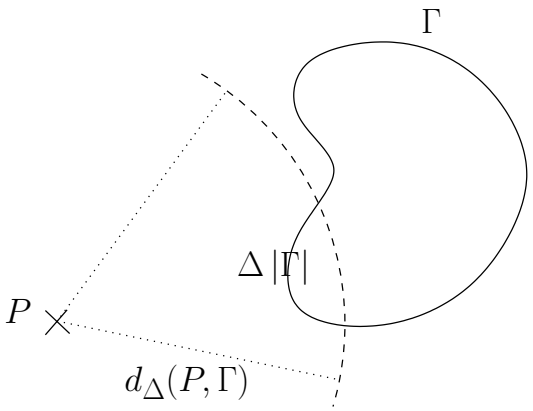


Figure 2.3: Geometric interpretation of $d_\Delta(P,\Gamma)$: Δ is the "percentage" of points of Γ whose distance to P is less than $d_\Delta(P,\Gamma)$.

Definition 27. *Let Γ and Γ' be two shapes, we define (see figure 2.4)*

$$d^\Delta(\Gamma',\Gamma) = \sup\left\{x \in \mathbb{R}^+; |\{Q \in \Gamma; d(Q,\Gamma') \geqslant x\}| \geqslant \Delta|\Gamma|\right\}$$

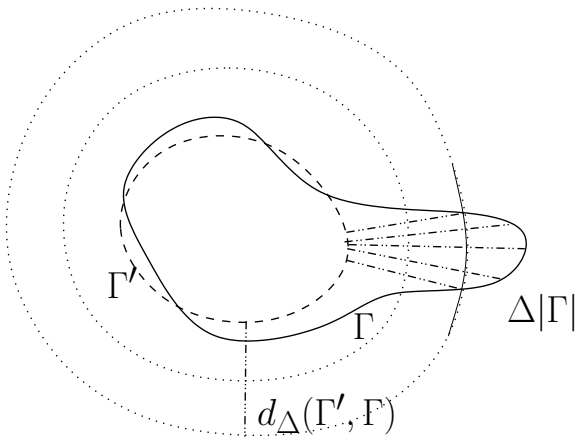


Figure 2.4: Geometric interpretation of $d^\Delta(\Gamma',\Gamma)$: Δ is the "percentage" of points of Γ whose distance to Γ' is greater than $d^\Delta(\Gamma',\Gamma)$.

If φ (respectively ψ) is an admissible function, we note (Δ_φ, α_φ) (respectively (Δ_ψ, α_ψ)) a pair (Δ,α) for the bracket $\langle\cdot\rangle_\Gamma^\varphi$ (respectively, $\langle\cdot\rangle_\Gamma^\psi$).

The following proposition relates $\tilde{\rho}_H$ to d_Δ and d^Δ.

Proposition 28. *The following relation is satisfied by $\tilde{\rho}_H$, d_Δ and d^Δ:*

$$\alpha_\psi\alpha_\psi \max\left(d^{\Delta_\psi}(\Gamma,\Gamma'), d^{\Delta_\psi}(\Gamma',\Gamma)\right) \leqslant \tilde{\rho}_H(\Gamma,\Gamma') \leqslant \alpha_\varphi \max\left(\sup_{P\in\Gamma'} d_{\Delta_\varphi}(P,\Gamma), \sup_{P\in\Gamma} d_{\Delta_\varphi}(P,\Gamma')\right)$$

Proof. We notice that

$$\forall P \in \mathbb{R}^2, \quad d(P,\Gamma) \leqslant \langle d(P,\cdot)\rangle_\Gamma^\varphi \leqslant \alpha_\varphi d_{\Delta_\varphi}(P,\Gamma),$$

and therefore

$$\alpha_\psi \;\; d^{\Delta_\psi}(\Gamma,\Gamma') \;\; \leqslant \;\; \left\langle\langle d(\cdot,\cdot)\rangle_\Gamma^\varphi\right\rangle_{\Gamma'}^\psi \;\; \leqslant \;\; \alpha_\varphi \;\; \sup_{P\in\Gamma'} d_{\Delta_\varphi}(P,\Gamma).$$

$\tilde{\rho}_H$ is a discrete bracket $\langle\cdot,\cdot\rangle_\theta$ of two such terms, θ an increasing function. We note α_θ an α associated to $\Delta = \frac{1}{2}$ through $\mathcal{P}$ for θ. For all positive a and b we have

$$\theta^{-1}\left(\frac{1}{2}\theta(\max(a,b))\right) \leqslant \langle a,b\rangle_\theta \leqslant \max(a,b)$$

$$\alpha_\theta \max(a,b) \leqslant \langle a,b\rangle_\theta \leqslant \max(a,b),$$

from where the conclusion follows. □

We now relate d_Δ and d^Δ to the Hausdorff distance ρ_H.

Proposition 29. *For all $P \in D$ and for all shapes Γ and Γ' we have*

$$d(P,\Gamma) \leqslant d_\Delta(P,\Gamma) \leqslant d(P,\Gamma) + \frac{\Delta}{2}|\Gamma|,$$

and

$$d_H(\Gamma,\Gamma') - \Delta\frac{|\Gamma|+|\Gamma'|}{2} \leqslant d^\Delta(\Gamma',\Gamma) \leqslant d_H(\Gamma,\Gamma') + \Delta\frac{|\Gamma|+|\Gamma'|}{2}$$

Proof. The lowerbound on $d_\Delta(P,\Gamma)$ is easy to obtain, the upperbound can be obtained by contradiction as follows: let us assume that there exists a point P and a curve Γ such that the upperbound is not satisfied. Hence

$$d_\Delta(P,\Gamma) > d(P,\Gamma) + \frac{\Delta}{2}|\Gamma|$$

Γ being compact, there exists a point Q of Γ such that $d(P,Q) = d(P,\Gamma)$. Let us now consider Γ as a C^2 function from $[0,1]$ to $\mathbb{R}^2$ such that $|\Gamma'(p)| = \mathrm{c}^{\mathrm{ste}} = |\Gamma|$ for all ps in $[0,1]$. Let $q \in [0,1]$ such that $\Gamma(q) = Q$, and consider the image by Γ of $I = \{p| \; |p-q| \leqslant \Delta/2\}$ (assuming $q \in]\Delta/2, 1-\Delta/2[$, otherwise the proof can be easily modified). By construction

$$|\Gamma(I)| = |I||\Gamma| = \Delta|\Gamma|,$$

and for all point R of $\Gamma(I)$ of parameter r

$$PR \leqslant PQ + QR \leqslant d(P,A) + |r-q||\Gamma| \leqslant d(P,A) + \frac{1}{2}\Delta|A|.$$

We have found a measurable subset of the curve Γ of length larger than or equal to $\Delta|\Gamma|$ such that all its points are at a distance of P less than $d(P,\Gamma) + \frac{1}{2}\Delta|\Gamma|$, a contradiction.

The proof of the second set of inequalities proceeds in a similar fashion by considering subsets of the curves Γ and Γ' centered at points P of Γ and Q of Γ' such that $\rho_H(\Gamma,\Gamma') = d(P,Q)$; this is always possible since Γ and Γ' are compact. □

By combining propositions 28 and 29 we obtain the

Proposition 30. *$\tilde{\rho}_H(\Gamma, \Gamma')$ has the following upper and lower bounds*

$$\alpha_\theta\alpha_\psi(\rho_H(\Gamma,\Gamma') - \Delta_\psi\frac{|\Gamma|+|\Gamma'|}{2}) \leq \tilde{\rho}_H(\Gamma,\Gamma') \leq \alpha_\varphi(\rho_H(\Gamma,\Gamma') + \Delta_\varphi\frac{|\Gamma|+|\Gamma'|}{2}). \tag{2.15}$$

We can now characterize the shapes Γ and Γ' such that

$$\tilde{\rho}_H(\Gamma,\, \Gamma') < \tilde{\rho}_H(\Gamma,\, \Gamma). \tag{2.16}$$

Theorem 31. *The condition (2.16) is equivalent to*

$$\rho_H(\Gamma,\Gamma') < 4c_0\Delta,$$

where the constant c_0 is defined in definition 7 and theorem 8, and Δ in the proof.

Proof. We use the upper and lower bounds (2.15) derived in proposition 30 and write

$$\alpha_\theta\alpha_\psi(\rho_H(\Gamma,\Gamma') - \Delta_\psi\frac{|\Gamma|+|\Gamma'|}{2}) < \alpha_\varphi\Delta_\varphi|\Gamma|$$

To simplify the analysis, let us assume that $\alpha_\theta\alpha_\psi = \alpha_\varphi$ and $\Delta_\psi = \Delta_\varphi = \Delta$, we obtain

$$\rho_H(\Gamma,\Gamma') < (\frac{3}{2}|\Gamma| + \frac{1}{2}|\Gamma'|)\Delta,$$

and hence (proposition 8)

$$\rho_H(\Gamma,\Gamma') < 4c_0\Delta,$$

Conversely, if Γ' is not in the Hausdorff ball with center Γ and radius $4c_0\Delta$, we necessarily have $\tilde{\rho}_H(\Gamma,\, \Gamma') > \tilde{\rho}_H(\Gamma,\, \Gamma)$. □

From this we conclude that, since Δ can be made arbitrarily close to 0, and the length of shapes is bounded, strange phenomena such as a shape Γ' closer to a shape Γ than Γ itself (in the sense of $\tilde{\rho}_H$) cannot occur or rather will be "invisible" to our algorithms.

For completeness we now present an example of such a phenomenon. In detail, we construct a pair (Γ_1, Γ_2) of curves of $\mathcal{S}$ such that $\tilde{\rho}_H(\Gamma_1,\, \Gamma_2) < \tilde{\rho}_H(\Gamma_1,\, \Gamma_1)$. We assume for simplicity that the function θ in (5.5) is the identity. Let O be a point in the plane and consider the family (C_r), $r > 0$ of circles of center O and radius r. We note $T(r)$ the distance $\tilde{\rho}_H(C_r, C_r)$ and, for all point P, $D(P,r) = \langle d(P,\cdot)\rangle^{\varphi}_{C_r}$. Notice that $T(r) = \langle D(\cdot,r)\rangle^{\psi}_{C_r}$. For symmetry reasons (rotation invariance) $D(P,r)$ is constant on C_r, we note $D(r)$ this value. Hence we have $T(r) = D(r)$. Let us compute $D(r)$ (see figure 2.5):

$$D(r) = \varphi^{-1}\left(\frac{1}{2\pi r}\int_0^{2\pi}\varphi\left(2r\sin\frac{\theta}{2}\right)rd\theta\right) = \varphi^{-1}\left(\frac{1}{2\pi}\int_0^{2\pi}\varphi\left(2r\sin\frac{\theta}{2}\right)d\theta\right)$$

The function $r \to r\sin\frac{\theta}{2}$ is strictly increasing for each $0 < \theta < 2\pi$, the functions φ and φ^{-1} are strictly decreasing, hence $r \to T(r)$ is strictly increasing. In particular, C_r is not a local minimum of $\Gamma \to \tilde{\rho}_H(\Gamma,\Gamma)$ and hence not a local minimum of $\Gamma \to \tilde{\rho}_H(C_r,\Gamma)$. Therefore there exists $\varepsilon > 0$ such that $\tilde{\rho}_H(C_r, C_{r-\varepsilon}) < \tilde{\rho}_H(C_r, C_r)$, see figure 2.5.

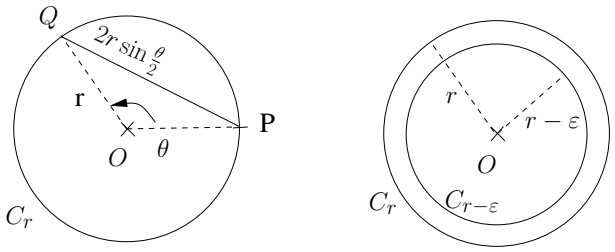


Figure 2.5: The curves C_r and $C_{r-\varepsilon}$ satisfy $\tilde{\rho}_H(C_r, C_{r-\varepsilon}) < \tilde{\rho}_H(C_r, C_r)$.

2.2.2 Some basic invariance properties of the approximation

The Hausdorff distance satisfies several natural basic properties, such as the invariance to a global rigid motion:

Proposition 32. *Let A and B be two shapes, and R a rigid motion. Then*

$$\rho_H(R(A), R(B)) = \rho_H(A, B).$$

and a simple relation for a scale transformation:

Proposition 33. *Let A and B be two shapes, and S a scale transformation with scale factor s. Then*

$$\rho_H(S(A), S(B)) = s\,\rho_H(A, B).$$

These properties are easily proved since the Hausdorff distance is based on the euclidean distance between points and that the effect of the studied deformations on the euclidean distance between points is simple.

The good news is that first property is also satisfied by the proposed family of approximations (for any applications φ, ψ and θ) and the second one is satisfied for the particular family we work with in practice (applications $x \mapsto x^\alpha$).

Proposition 34. *Let A and B be two shapes, and R a rigid motion. Then*

$$\tilde{\rho}_H(R(A), R(B)) = \tilde{\rho}_H(A, B).$$

This comes from the invariance of the basic constituent of the approximation. We recall that $\tilde{\rho}_H(A\,, B) = \left\langle \left\langle \langle d(\cdot,\,\cdot)\rangle_B^\varphi \right\rangle_A^\psi, \left\langle \langle d(\cdot,\,\cdot)\rangle_A^\varphi \right\rangle_B^\psi \right\rangle^\theta$:

Lemma 35. *Let A be a shape, and R a rigid motion. Then for any point $\mathbf{x}$,*

$$\langle d(R(\mathbf{x}),\,\cdot)\rangle_{R(A)}^\varphi \;=\; \langle d(\mathbf{x},\,\cdot)\rangle_A^\varphi\,.$$

Proof.

$$\begin{aligned}
\langle d(R(\mathbf{x}),\,\cdot)\rangle_{R(A)}^\varphi &= \varphi^{-1}\left(\frac{1}{|R(A)|}\int_{R(A)} \varphi(d(R(\mathbf{x}),\mathbf{z}))\,d\mathbf{z}\right)\\
&= \varphi^{-1}\left(\frac{1}{|A|}\int_{A} \varphi(d(R(\mathbf{x}),R(\mathbf{y})))\,d\mathbf{y}\right)\\
&\qquad\qquad \text{(since the Jacobian of } R \text{ is 1)}\\
&= \varphi^{-1}\left(\frac{1}{|A|}\int_{A} \varphi(d(\mathbf{x},\mathbf{y}))\,d\mathbf{y}\right) \;=\; \langle d(\mathbf{x},\,\cdot)\rangle_A^\varphi
\end{aligned}$$

□

We can now consider the second property:

Proposition 36. *Let A and B be two shapes, and S a scale transformation with scale factor s. We consider the particular case of approximation where there exist three real numbers α, β, γ such that $\varphi : x \mapsto x^\alpha, \psi : x \mapsto x^\beta, \theta : x \mapsto x^\gamma$. Then*

$$\rho_H(S(A), S(B)) = s\, \rho_H(A, B).$$

Once again, this comes from the basis constituent of the approximation:

Lemma 37. *Let A be a shape, f any application from A to $\mathbb{R}^+$, s a real positive number, and $\varphi : x \mapsto x^\alpha$. Then*

$$\langle s\, f(\cdot)\rangle_A^\varphi = s\ \langle f(\cdot)\rangle_A^\varphi .$$

Proof. Indeed:

$$\begin{aligned} \langle s\, f(\cdot)\rangle_A^\varphi &= \left(\frac{1}{|A|}\int_A (sf(\mathbf{x}))^\alpha d\mathbf{x}\right)^{1/\alpha} \\ &= \left(s^\alpha \frac{1}{|A|}\int_A f(\mathbf{x})^\alpha d\mathbf{x}\right)^{1/\alpha} \\ &= s\ \langle f(\cdot)\rangle_A^\varphi . \end{aligned}$$

The initial consideration of $f(\mathbf{x}) = d(\mathbf{x}, \mathbf{y})$ with $d(s\mathbf{x}, s\mathbf{y}) = s\, d(\mathbf{x}, \mathbf{y})$ ends the proof. □

Chapter 3

Different Notions of *Shape* and Associated Distances

Abstract

The intuitive notion of a *shape* has many facets. This chapter is an attempt to model these different aspects that have been forgotten so far and to make more explicit the links between the models. The Hausdorff distance between regions with non-empty interior is compared to the Hausdorff distance between their boundaries. Then local shape descriptors are incorporated into the distance between shapes.

3.1 Full shapes vs boundaries

3.1.1 Definitions

The set $\mathcal{S}$ of all shapes (of interest) introduced in chapter 1, definition 11, contains shapes that are linked to two different intuitive notions of "shape". The first notion of a "shape" is the boundary of an object, for instance a curve in the plane $\mathbb{R}^2$ or a two-dimensioned manifold in the space $\mathbb{R}^3$, like an empty sphere. The other notion of a "shape" is the object itself, that is to say not only its boundary but also its interior.

Definition 38. *The set $\mathcal{F}$ of full shapes is the subset of all shapes in $\mathcal{S}$ which satisfy that the closure of their interior is themselves.*

Definition 39. *The set $\mathcal{B}$ of boundaries is the subset of all shapes in $\mathcal{S}$ which can be written as the boundary of an element of $\mathcal{F}$.*

Before going on, one elementary property of these sets has to be checked:

Proposition 40. *The boundary of any full shape is in $\mathcal{S}$ and consequently in $\mathcal{B}$.*

Proof. This is quite straightforward since $\mathcal{S}$ is defined by properties of the distance function and the boundary of the shapes, and since a full shape and its boundary share the same distance function (except for a sign) and the same boundary. □

This duality between the full objects and their boundaries naturally leads to the question: if Ω_1 and Ω_2 are two full objects, then is their Hausdorff distance $d_H(\Omega_1, \Omega_2)$ linked to the Hausdorff distance $d_H(\partial\Omega_1, \partial\Omega_2)$ of their boundaries ?

3.1.2 Hausdorff distance between boundaries of shapes

The first remark to do is that they are not the same in the most general case, even if the two shapes are topologically homotope. On figure 3.1 are represented a disk (with boundary in dotted line) and a (thick) snake lying on the boundary of the disk. The Hausdorff distance between the two "full" shapes is high, since the snake is far from the middle of the disk. However the Hausdorff distance between the boundaries is low since the boundary of the snake is close to the circle.

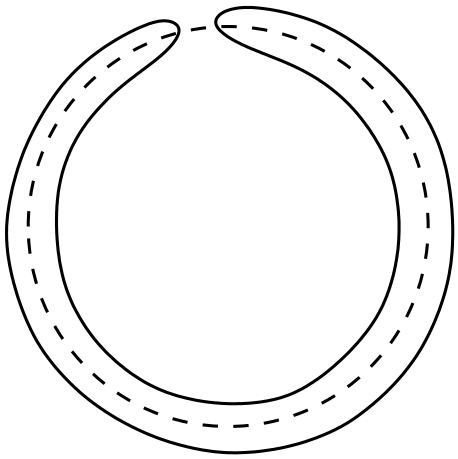


Figure 3.1: Example of two shapes whose the "full" Hausdorff distance is greater than the Hausdorff distance of the boundaries.

By contrast, the Hausdorff distance between the boundaries can be high while the Hausdorff distance between the "full" shapes is small (see figure 3.2).

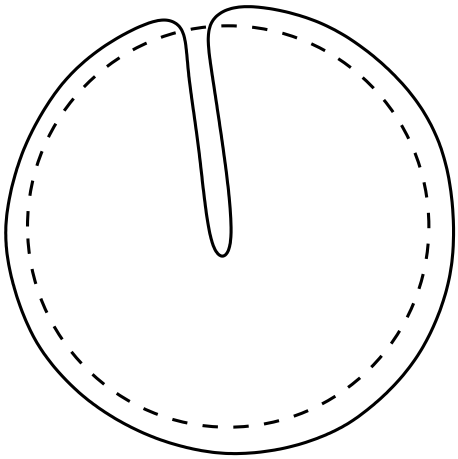


Figure 3.2: Example of two shapes whose the "full" Hausdorff distance is smaller than the Hausdorff distance of the boundaries.

Nevertheless these two distances appear to be often the same, for example for convex objects. The relation between them can be made more explicit. The reader not interested in details can skip to the corollary 43.

Given two full shapes Ω_1 and Ω_2 in $\mathcal{F}$ and a pair of points $\mathbf{p}_1 \in \Omega_1$ and $\mathbf{p}_2 \in \Omega_2$ which

realizes the Hausdorff distance:

$$d_H(\Omega_1,\Omega_2) = d(\mathbf{p}_1,\mathbf{p}_2) = \sup\left(\sup_{\mathbf{x}\in\Omega_1} d(\mathbf{x},\Omega_2),\ \sup_{\mathbf{y}\in\Omega_2} d(\mathbf{y},\Omega_1)\right)$$

with either $\mathbf{p}_1$ being the projection of $\mathbf{p}_2$ onto Ω_2, or $\mathbf{p}_2$ being the projection of $\mathbf{p}_1$ onto Ω_1, the projections being not necessarily unique. In the first case, $\mathbf{p}_1$ is the projection of $\mathbf{p}_2$, so it is on the boundary of Ω_1. Indeed, if it was not, then there would be a small open set Σ including $\mathbf{p}_1$ and included in Ω_1; but the function which associates to any point of Σ its distance to the point $\mathbf{p}_2$ is differentiable at point $\mathbf{p}_1$ – if $d(\mathbf{p}_1,\mathbf{p}_2)$ is not zero, i.e. if $d_H(\Omega_1,\Omega_2)$ is not null, which is assumed here – and its gradient is not zero. Therefore there would exist other points in $\Sigma \subset \Omega_1$ nearer from $\mathbf{p}_2$ than $\mathbf{p}_1$, which would contradict the hypothesis. The point $\mathbf{p}_1$ is consequently on the boundary of Ω_1. What about the point $\mathbf{p}_2$? $\mathbf{p}_2$ realizes the maximum in Ω_2 of the distance function to Ω_1. Therefore $\mathbf{p}_2$ either is on the boundary of Ω_2 or realizes a local maximum (in $\mathbb{R}^n$) of the distance function to Ω_1. In this last case, $\mathbf{p}_2$ appears to belong to the skeleton (see definition 1), since anywhere else the gradient of this distance function is well-defined and non-zero. The point $\mathbf{p}_2$ consequently belongs either to $\partial\Omega_2$ or to $\mathrm{Sk}(\Omega_1)\cap\Omega_2$, with the previous constraint that it also belongs to the complement $\complement\Omega_1$ of Ω_1; on the other side, the point $\mathbf{p}_1$ belongs to $\partial\Omega_1$. Note that the skeleton of Ω_1 is the part of the skeleton of $\partial\Omega_1$ which is not in Ω_1. Let us rewrite this in order, with the convention that the distance to the empty set and the supremum over the empty set are zero:

$$d_H(\Omega_1,\Omega_2) = \sup\left(\sup_{\mathbf{x}\in\Omega_1} d(\mathbf{x},\Omega_2),\ \sup_{\mathbf{y}\in\Omega_2} d(\mathbf{y},\Omega_1)\right)$$

$$\begin{aligned}
\sup_{\mathbf{x}\in\Omega_1} d(\mathbf{x},\Omega_2) &= \sup_{\mathbf{x}\in\partial\Omega_1\cup\Omega_1\cap\mathrm{Sk}(\Omega_2)} d(\mathbf{x},\Omega_2)\\
&= \sup_{\mathbf{x}\in\complement\Omega_2\cap(\partial\Omega_1\cup\Omega_1\cap\mathrm{Sk}(\Omega_2))} d(\mathbf{x},\Omega_2)\\
&= \sup_{\mathbf{x}\in\complement\Omega_2\cap(\partial\Omega_1\cup\Omega_1\cap\mathrm{Sk}(\Omega_2))} d(\mathbf{x},\partial\Omega_2)\\
&= \sup\left(\sup_{\mathbf{x}\in\partial\Omega_1\cap\complement\Omega_2} d(\mathbf{x},\partial\Omega_2),\ \sup_{\mathbf{x}\in\mathrm{Sk}(\Omega_2)\cap\Omega_1} d(\mathbf{x},\partial\Omega_2)\right)
\end{aligned}$$

Consequently:

Proposition 41. *Let Ω_1 and Ω_2 be two shapes in $\mathcal{F}$. Then $d_H(\Omega_1,\Omega_2)$ is equal to:*

$$\sup\Big(\sup_{\mathbf{x}\in\partial\Omega_1\cap\complement\Omega_2} d(\mathbf{x},\partial\Omega_2),\ \sup_{\mathbf{x}\in\mathrm{Sk}(\Omega_2)\cap\Omega_1} d(\mathbf{x},\partial\Omega_2),\ \sup_{\mathbf{y}\in\partial\Omega_2\cap\complement\Omega_1} d(\mathbf{y},\partial\Omega_1),\ \sup_{\mathbf{y}\in\mathrm{Sk}(\Omega_1)\cap\Omega_2} d(\mathbf{y},\partial\Omega_1)\Big).$$

Remark 42. *This splitting into four terms is to be compared to the expression of the Hausdorff distance between the boundaries $d_H(\partial\Omega_1,\partial\Omega_2) = \sup\left(\sup_{\mathbf{x}\in\partial\Omega_1} d(\mathbf{x},\partial\Omega_2),\ \sup_{\mathbf{y}\in\partial\Omega_2} d(\mathbf{y},\partial\Omega_1)\right)$ which is equal to:*

$$\sup\Big(\sup_{\mathbf{x}\in\partial\Omega_1\cap\complement\Omega_2} d(\mathbf{x},\partial\Omega_2),\ \sup_{\mathbf{x}\in\partial\Omega_1\cap\Omega_2} d(\mathbf{x},\partial\Omega_2),\ \sup_{\mathbf{y}\in\partial\Omega_2\cap\complement\Omega_1} d(\mathbf{y},\partial\Omega_1),\ \sup_{\mathbf{y}\in\partial\Omega_2\cap\Omega_1} d(\mathbf{y},\partial\Omega_1)\Big).$$

Two terms among the four ones are identical in the expressions of $d_H(\Omega_1,\Omega_2)$ *and* $d_H(\partial\Omega_1,\partial\Omega_2)$*, and the other ones are not correlated in the sense that we can find shapes so that* $\sup\left(\sup_{\mathbf{x}\in \mathrm{Sk}(\Omega_2)\cap\Omega_1} d(\mathbf{x},\partial\Omega_2), \sup_{\mathbf{y}\in \mathrm{Sk}(\Omega_1)\cap\Omega_2} d(\mathbf{y},\partial\Omega_1)\right)$ *is high whereas* $\sup\left(\sup_{\mathbf{x}\in \partial\Omega_1\cap\Omega_2} d(\mathbf{x},\partial\Omega_2), \sup_{\mathbf{y}\in \partial\Omega_2\cap\Omega_1} d(\mathbf{y},\partial\Omega_1)\right)$ *is small, and reciprocally.*

The conclusion is that $d_H(\Omega_1,\Omega_2)$ *is sometimes strictly greater than* $d_H(\partial\Omega_1,\partial\Omega_2)$*, sometimes strictly smaller, but they are equal one to the other when the term* $\sup\Big(\sup_{\mathbf{x}\in \partial\Omega_1\cap\complement\Omega_2} d(\mathbf{x},\partial\Omega_2), \sup_{\mathbf{y}\in \partial\Omega_2\cap\complement\Omega_1} d(\mathbf{y},\partial\Omega_1)\Big)$ *dominates.*

Corollary 43. *The Hausdorff distances* $d_H(\Omega_1,\Omega_2)$ *and* $d_H(\partial\Omega_1,\partial\Omega_2)$ *between two close enough shapes in* $\mathcal{F}$ *or* $\mathcal{B}$ *are equal one to the other, assuming in the case of the only knowledge of the boundaries that the associated full shapes are such that any point far enough (in comparison with the distance between the boundaries) from the two boundaries either belongs to both full shapes or belongs to none. This property is needed because the only knowledge of the boundaries does not carry information about where the interior and exterior of the associated full shapes are. This property is probably always true for bounded full shapes in* $\mathbb{R}^n$*. It will be denoted by:*

$$\mathcal{P}(\Omega_1,\Omega_2): \quad \complement\big([\partial\Omega_1+B(\eta)]\cup[\partial\Omega_2+B(\eta)]\big) \quad\subset\quad [\Omega_1\cap\Omega_2]\ \cup\ \big[\complement\Omega_1\cap\complement\Omega_2\big]$$

where $B(\eta)$ *stands for the ball of radius* $\eta = 2d_H(\partial\Omega_1,\partial\Omega_2)$ *and* $\partial\Omega_1+B(\eta)$ *for the set of points whose distance to* $\partial\Omega_1$ *is less than* η*. Then:*

$\forall\Omega_1\in\mathcal{F},\ \exists\varepsilon>0,\ \forall\Omega_2\in\mathcal{F}$,

$$\begin{cases} d_H(\Omega_1,\Omega_2)<\varepsilon \\ \ \ or \\ d_H(\partial\Omega_1,\partial\Omega_2)<\varepsilon \ \ and\ \ \mathcal{P}(\Omega_1,\Omega_2) \end{cases} \quad\Longrightarrow\quad d_H(\Omega_1,\Omega_2)=d_H(\partial\Omega_1,\partial\Omega_2).$$

Moreover, in that case the Hausdorff distance is symmetric in the sense that:

$$d_H(\Omega_1,\Omega_2)=d_H(\partial\Omega_1,\partial\Omega_2)=\sup_{\mathbf{y}\in\partial\Omega_2} d(\mathbf{y},\partial\Omega_1)=\sup_{\mathbf{x}\in\partial\Omega_1} d(\mathbf{x},\partial\Omega_2).$$

Proof. This result is not true for any subsets of $\mathbb{R}^n$. The condition of being shapes in $\mathcal{F}$ is very important here.

The definition of the set of shapes $\mathcal{S}$, which $\mathcal{F}$ and $\mathcal{B}$ are part of, implies that any shape $\Omega_1\in\mathcal{S}$ is far enough from its (exterior) skeleton: the euclidean distance between any point of $\mathrm{Sk}(\Omega_1)$ and any point of Ω_1 is greater than h_0. Consequently, if the Hausdorff distance between Ω_1 and Ω_2 is less than h_0, then Ω_2 and $\mathrm{Sk}(\Omega_1)$ do not intersect, and $\sup_{\mathbf{x}\in\mathrm{Sk}(\Omega_2)\cap\Omega_1} d(\mathbf{x},\partial\Omega_2)=\sup_{\mathbf{y}\in\mathrm{Sk}(\Omega_1)\cap\Omega_2} d(\mathbf{y},\partial\Omega_1)=0$. Therefore:

$$d_H(\Omega_1,\Omega_2)<h_0 \quad\Longrightarrow\quad d_H(\Omega_1,\Omega_2)=\sup\Big(\sup_{\mathbf{x}\in\partial\Omega_1\cap\complement\Omega_2} d(\mathbf{x},\partial\Omega_2), \sup_{\mathbf{y}\in\partial\Omega_2\cap\complement\Omega_1} d(\mathbf{y},\partial\Omega_1)\Big).$$

The same property stands if $d_H(\partial\Omega_1,\partial\Omega_2)<h_0/2$. Indeed if $\mathbf{x}$ belongs to $\mathrm{Sk}(\Omega_1)\subset\mathrm{Sk}(\partial\Omega_1)\cap\complement\Omega_1$ and to Ω_2, then the property $\mathcal{P}$ implies that $d(\mathbf{x},\partial\Omega_1)<2d_H(\partial\Omega_1,\partial\Omega_2)$ since $\mathbf{x}$ is in Ω_2 and not in Ω_1. But $d(\mathbf{x},\partial\Omega_1)=d(\mathbf{x},\Omega_1)\geqslant h_0>2d_H(\partial\Omega_1,\partial\Omega_2)$. So $\mathrm{Sk}(\Omega_1)\cap\Omega_2$ is empty and:

$$d_H(\partial\Omega_1, \partial\Omega_2) < h_0/2 \text{ and } \mathcal{P}(\Omega_1, \Omega_2)$$
$$\implies \quad d_H(\Omega_1, \Omega_2) = \sup\Big(\sup_{\mathbf{x}\in\partial\Omega_1\cap\complement\Omega_2} d(\mathbf{x}, \partial\Omega_2), \sup_{\mathbf{y}\in\partial\Omega_2\cap\complement\Omega_1} d(\mathbf{y}, \partial\Omega_1)\Big).$$

Thus, if these terms are shown to dominate in the expression of $d_H(\partial\Omega_1, \partial\Omega_2)$ when this quantity is small enough, the proof is completed.

Let us suppose that $d_H(\partial\Omega_1, \partial\Omega_2)$ is reached in the part $\sup_{\mathbf{x}'\in\partial\Omega_1} d(\mathbf{x}', \partial\Omega_2)$, or else swap Ω_1 and Ω_2. Let $\mathbf{x} \in \partial\Omega_1$ be such that $d(\mathbf{x}, \partial\Omega_2) = d_H(\partial\Omega_1, \partial\Omega_2)$. In the proof of proposition 24 it has been shown that there exists a point $\mathbf{y}$ on $\partial\Omega_2$ whose projection on $\partial\Omega_1$ is $\mathbf{x}$ if $d_H(\partial\Omega_1, \partial\Omega_2)$ is small enough. Consequently: $d_H(\partial\Omega_1, \partial\Omega_2) = d(\mathbf{x}, \partial\Omega_2) \leqslant d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \partial\Omega_1) \leqslant \sup_{\mathbf{y}'\in\partial\Omega_2} d(\mathbf{y}', \partial\Omega_1) \leqslant d_H(\partial\Omega_1, \partial\Omega_2)$. Therefore $d(\mathbf{y}, \partial\Omega_1) = d_H(\partial\Omega_1, \partial\Omega_2)$. If $\mathbf{x}$ does not belong to Ω_2 or $\mathbf{y}$ does not belong to Ω_1 then the terms that dominate in the expression of $d_H(\partial\Omega_1, \partial\Omega_2)$ are the good ones and consequently the distance between the full shapes is equal to the distances between the boundaries. If not, then $\mathbf{x} \in \Omega_2$ and $\mathbf{y} \in \Omega_1$, and the configuration is awkward. The vector $\mathbf{xy}$ is normal to both $\partial\Omega_1$ at point $\mathbf{x}$ and $\partial\Omega_2$ at point $\mathbf{y}$. Let us consider the point $\mathbf{z}$ on the straight line $(\mathbf{xy})$ such that the vector $\mathbf{xz}$ is equal to $4\mathbf{xy}$. As $\mathbf{z}$ is on the normal to $\partial\Omega_1$ at $\mathbf{x}$ and $|\mathbf{xz}| = 4d_H(\partial\Omega_1, \partial\Omega_2) < h_0$, its projection onto $\partial\Omega_1$ is unique and is $\mathbf{x}$, and consequently $d(\mathbf{z}, \partial\Omega_1) \geqslant 2d_H(\partial\Omega_1, \partial\Omega_2)$. Similarly $d(\mathbf{z}, \partial\Omega_2) = d(\mathbf{z}, \mathbf{y}) \geqslant 2d_H(\partial\Omega_1, \partial\Omega_2)$ and the property $\mathcal{P}$ implies that $\mathbf{z}$ belongs either to both Ω_1 and Ω_2, or none of them. As there can be no point of $\partial\Omega_1$ on the segment $)\mathbf{xz})$ since all the projection of all these points onto $\partial\Omega_1$ is $\mathbf{x}$, and as $\mathbf{y} \in \Omega_1$, then $\mathbf{z} \in \Omega_1$. Similarly $\mathbf{z}$ is not in Ω_2. Hence a contradiction. To sum up, either $\mathbf{x} \in \complement\Omega_2$ or $\mathbf{y} \in \complement\Omega_1$ and there exists $\varepsilon > 0$ such that:

$$d_H(\partial\Omega_1, \partial\Omega_2) < \varepsilon \text{ and } \mathcal{P}(\Omega_1, \Omega_2)$$
$$\implies \quad d_H(\partial\Omega_1, \partial\Omega_2) = \sup\Big(\sup_{\mathbf{x}\in\partial\Omega_1\cap\complement\Omega_2} d(\mathbf{x}, \partial\Omega_2), \sup_{\mathbf{y}\in\partial\Omega_2\cap\complement\Omega_1} d(\mathbf{y}, \partial\Omega_1)\Big).$$

Last and least, does this equality still stand for the condition $d_H(\Omega_1, \Omega_2) < \varepsilon$? Let $\mathbf{x} \in \partial\Omega_1$ be such that $d(\mathbf{x}, \partial\Omega_2) = d_H(\partial\Omega_1, \partial\Omega_2)$ (with a swap between Ω_1 and Ω_2 before if necessary). If $\mathbf{x} \notin \Omega_2$ then the proof is finished. Else $\mathbf{x} \in \Omega_2$. A way to solve the problem is to prove that $d_H(\partial\Omega_1, \partial\Omega_2)$ is small enough and apply the previous framework. The ball centered on $\mathbf{x}$ with radius $d_H(\partial\Omega_1, \partial\Omega_2)$ is included in Ω_2 since $\mathbf{x} \in \Omega_2$ and $d(\mathbf{x}, \partial\Omega_2) = d_H(\partial\Omega_1, \partial\Omega_2)$. Besides, the projection onto $\partial\Omega_1$ of all points belonging to the normal to $\partial\Omega_1$ at point $\mathbf{x}$ and at distance $< h_0$ from $\mathbf{x}$ is $\mathbf{x}$. The segment they form go through $\partial\Omega_1$ only at point $\mathbf{x}$ so that one of its two halves is included in $\complement\Omega_1$. For any point $\mathbf{p}$ of this half of the segment, $d(\mathbf{p}, \Omega_1) = d(\mathbf{p}, \partial\Omega_1) = d(\mathbf{p}, \mathbf{x})$. There is a point $\mathbf{y}$ on this half which is at distance $\min(h_0, d_H(\partial\Omega_1, \partial\Omega_2))/3$ from $\mathbf{x}$. To sum up: $\mathbf{y}$ is in Ω_2 and $d(\mathbf{y}, \Omega_1) = \min(h_0, d_H(\partial\Omega_1, \partial\Omega_2))/3$. So $d_H(\Omega_1, \Omega_2) \geqslant \min(h_0, d_H(\partial\Omega_1, \partial\Omega_2))/3$. If $d_H(\Omega_1, \Omega_2) < h_0/3$ then necessarily $d_H(\partial\Omega_1, \partial\Omega_2) \leqslant 3d_H(\Omega_1, \Omega_2)$ and consequently the previous paragraph can be applied again if $\mathcal{P}$ is shown to be also satisfied. As $d_H(\Omega_1, \Omega_2) < h_0$, the first implication stated in this proof gives $d_H(\Omega_1, \Omega_2) \leqslant d_H(\partial\Omega_1, \partial\Omega_2)$. But to prove $\mathcal{P}$, it is sufficient to show that any point that is in one of the full shape and not in the other is necessarily at distance $\leqslant 2d_H(\partial\Omega_1, \partial\Omega_2)$ from the boundary of the full shape which it does not belong to. Let $\mathbf{z}$ be in Ω_1 and not in Ω_2. Then $d(\mathbf{z}, \partial\Omega_2) = d(\mathbf{z}, \Omega_2) \leqslant d_H(\Omega_1, \Omega_2) \leqslant d_H(\partial\Omega_1, \partial\Omega_2)$. So $\mathcal{P}$ is satisfied and

$$d_H(\Omega_1, \Omega_2) < \varepsilon \quad \implies \quad d_H(\partial\Omega_1, \partial\Omega_2) = \sup\Big(\sup_{\mathbf{x}\in\partial\Omega_1\cap\complement\Omega_2} d(\mathbf{x}, \partial\Omega_2), \sup_{\mathbf{y}\in\partial\Omega_2\cap\complement\Omega_1} d(\mathbf{y}, \partial\Omega_1)\Big).$$

□

Remark 44. *Since the skeleton of a shape is naturally built from the boundary of the shape (and not from the "full" shape), the proposition 41 gives a simple way to compute the Hausdorff distance between the "full" shapes based on the only knowledge of their boundaries. Note that all suprema are taken among subsets of null measure (for the measure in the space which embeds the shape).*

The same framework as before (in chapter 2) can be applied to obtain a smooth, differentiable approximation of the "full" Hausdorff distance with respect to the boundary. The only new point here is the skeleton, or more precisely the subset of the (exterior) skeleton which is made of local extrema of the distance function to the boundary. These points can be expressed as the ones where the the derivative of the approximation of the distance function is zero. I do not know if the application which associates this subset to any boundary is continuous and differentiable with respect to the boundary. However, in the worst case, in order to ensure that the approximation of the "full" distance is smooth, one can use a variation on the same trick as before ($\|\cdot\|_p \to \|\cdot\|_\infty$) and replace the consideration of these extremal points by the consideration of all points, but with weights strongly depending on the norm of the gradient of the approximation of the distance function.

3.2 Incorporating shape descriptors

3.2.1 Motivation

To the notion of *shape* are associated many different kinds of considerations which have not been evoked so far: *shape* descriptions, such as the curvature, oscillations, rigidity, corners... and shape attributes, such as color and texture. The choice of the set of shapes and of the distance which is defined on it should reflect as much as possible the intuitive notion of *shape*, since they will play a decisive role in the evolution of shapes when matching one shape onto another by minimizing the distance between them. Generally speaking, the design of the set of objects and of the relations between them is the central task of the model maker. So, how to incorporate these considerations in the model of *shape* ?

In practice, the minimization of the Hausdorff distance between two shapes A and B with respect to A (see part II) sometimes leads to less than satisfying evolutions, particularly when a part of the moving contour is attracted by several different parts of the target contour, which occurs when the target contour shows large, high frequency oscillations. An example of this phenomenon can be observed when matching a rectangle to a hand, where the length of the fingers is far bigger than their thickness.

A way to try to avoid this kind of evolution consists in modifying the Hausdorff distance so that the induced correspondences reflect our intuition. Indeed, the Hausdorff distance consider implicitly that each point $\mathbf{x}$ of a shape A is attracted by its closest neighbor on the other shape B and by all the points of B whose projection on A is $\mathbf{x}$, and that the more meaningful points are those with biggest euclidean distance to their associated points. One could wonder if this straight distance is the only good one, or if one would not be wise to search for more information than the euclidean distance between any two points.

3.2.2 A first extension of the Hausdorff distance to boundary orientation

For instance, let us imagine that we would like to take into account the local orientation of the shapes when comparing two points. If we denote by $\mathbf{n}_A(\mathbf{x})$ the normal of the shape A at point $\mathbf{x}$, then the (positive) angle $\alpha(\mathbf{n}_A(\mathbf{x}), \mathbf{n}_B(\mathbf{y}))$ between them carries information about

the relative orientation of the shapes at these two points. It is then reasonable to think that, for a same euclidean distance between $\mathbf{x}$ and $\mathbf{y}$, the cost of the deformation that the shape A has to locally undergo in order to reach B should increase with the angle between the normals $\mathbf{n}_A(\mathbf{x})$ and $\mathbf{n}_B(\mathbf{y})$, since it requires a local rotation, or torsion, of the shape. Hence the idea of incorporating this information with the euclidean distance $d(\mathbf{x},\mathbf{y})$ in a new, more meaningful, distance, denoted by $d_1(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|_2 + K\alpha\big(n_A(\mathbf{x}),\mathbf{n}_B(\mathbf{y})\big)$, so that the distance between any points of any two curves depends on both euclidean distance and local orientation (with a relative importance expressed by factor K). The important point is that the new modified Hausdorff distance related to d_1 is still a distance between shapes.

Proposition 45. *Let d^* be a distance on $\mathbb{R}^n \times \mathbb{S}_{n-1}$ where $\mathbb{S}_{n-1}$ is the unit sphere of $\mathbb{R}^n$. Then the application which associates to any pair of shapes (A,B) embedded in $\mathbb{R}^n$ the real number*

$$d_H^*(A,B) \;=\; \max\Big\{ \sup_{\mathbf{x}\in A}\inf_{\mathbf{y}\in B} d^*\big((\mathbf{x},\mathbf{n}_A(\mathbf{x})),\,(\mathbf{y},\mathbf{n}_B(\mathbf{y}))\big), \quad \sup_{y\in B}\inf_{\mathbf{x}\in A} d^*\big((\mathbf{x},\mathbf{n}_A(\mathbf{x})),\,(\mathbf{y},\mathbf{n}_B(\mathbf{y}))\big) \Big\}$$

is a distance on the set of shapes.

Note that we can see it as a change of the embedding space, where a point is a pair $(\mathbf{x},\mathbf{n})$, and into which any shape can be "*lifted*". This kind of process is usual in differential geometry or more exactly in algebraic topology.

Proof. The properties that make the usual Hausdorff distance a distance are still satisfied by this extension. First, d^* is a distance on $\mathbb{R}^n \times \mathbb{S}_{n-1}$, as d was on $\mathbb{R}^n$, and secondly the syntax (with max, sup and inf) has been preserved.

1. $\forall A,\ B,\ \ d_H^*(A,B) = d_H^*(B,A)$ (straightforward)
2. $\forall A,\ \ d_H^*(A,A) = 0$ since for any $\mathbf{x}$ on A, $\inf_{\mathbf{y}\in A} d^*\big((\mathbf{x},\mathbf{n}_A(\mathbf{x})),\,(\mathbf{y},\mathbf{n}_A(\mathbf{y}))\big)$ is reached for $\mathbf{y}=\mathbf{x}$ and d^* is a distance, so for any $\mathbf{x}$ this quantity is zero. Similarly, if $d_H^*(A,B)=0$ then $A=B$ since it implies that each point of each shape also belongs to the other shape.
3. $\forall A,B,C,\ \ d_H^*(A,B)+d_H^*(B,C) \geqslant d_H^*(A,C)$: same proof as the one for the Hausdorff distance. Let us denote by $A^* = \{(\mathbf{x},\mathbf{n}_A(\mathbf{x})) \in A\times\mathbb{S}_{n-1}\}$ the "lifted" version of A. If the "*lifted*" shapes A^*,B^*,C^* are closed, bounded subsets of $\mathbb{R}^n\times\mathcal{S}_{n-1}$ (if not closed, just consider their closure), then the distance $d_H^*(A,C)$ is reached for a pair $(\mathbf{a},\mathbf{c}) \in (A^*\times C^*)$. Either $\mathbf{a}$ is the projection of $\mathbf{c}$ onto A^*, or $\mathbf{c}$ is the projection of $\mathbf{a}$ onto C^*. Let us consider the second case (the first one being symmetric). If $\mathbf{b}$ stands for the point that minimizes $\inf_{\mathbf{b}\in B^*} d^*(\mathbf{a},\mathbf{b})$, then on one hand $d^*(\mathbf{a},\mathbf{b}) \leqslant d_H^*(A,B)$ and on the other hand the distance between $\mathbf{b}$ and C^* is smaller than $d_H^*(B,C)$, so, as the projection of $\mathbf{a}$ onto C^* is the shortest path from $\mathbf{a}$ to C^*, we have:

$$d_H^*(A,C) = d^*(\mathbf{a},C^*) \leqslant d^*(\mathbf{a},\mathbf{b}) + d^*(\mathbf{b},C^*) \leqslant d_H^*(A,B) + d_H^*(B,C).$$

□

3.2.3 Extension of the Hausdorff distance to shape local descriptors

In a farther perspective, this highlights the process and the meaning of the Hausdorff distance. First, to each point x of the shapes, it implicitly associates a corresponding point on the other shape, which is the nearest neighbor of x on this other shape for the chosen distance d^*. These pairs of corresponding points are not symmetric in the sense that if $\mathbf{b} \in B^*$ is the projection onto B^* of $\mathbf{a} \in A^*$, then $\mathbf{a}$ is not necessarily itself the projection of $\mathbf{b}$ onto A^*. Then, among all these pairs, it returns the maximum distance between each point and its associated point. The design of the distance d^* should consequently be carefully studied since it defines the correspondences between the two shapes (or more exactly two sets of correspondences via the pairs of corresponding points) from which the final distance d_H^* is directly built.

A substitution of an other distance d^* for the usual euclidean distance allows to define more precisely the notion of similarity between two small parts of shapes. For instance, the previously proposed distance d_1 takes into account the local orientation of the shapes. For any local descriptor of a shape (curvature, wavelet coefficients, shape patches, color, texture, ...) and a related distance, the Hausdorff distance can be similarly adapted:

Proposition 46. *Let f be an application (which should be seen as a shape local descriptor) which associates to any shape A (embedded in $\mathbb{R}^n$) and any point x of A an element in a space named $\mathcal{D}$. Let $d_{\mathcal{D}}$ be a distance on $\mathcal{D}$. Then the application which associates to any pair of shapes (A, B) embedded in $\mathbb{R}^n$ the real number*

$$d_H^{\mathcal{D}}(A,B) = \max\left\{ \begin{array}{l} \sup\limits_{\mathbf{x}\in A} \inf\limits_{\mathbf{y}\in B} d(\mathbf{x},\mathbf{y}) + d_{\mathcal{D}}(f(\mathbf{x},A), f(\mathbf{y},B)), \\ \sup\limits_{y\in B} \inf\limits_{\mathbf{x}\in A} d(\mathbf{x},\mathbf{y}) + d_{\mathcal{D}}(f(\mathbf{x},A), f(\mathbf{y},B)) \end{array} \right\}$$

is a distance on the set of shapes.

Proof. The same as in Proposition 45. The proof of additional lemmas if $\mathcal{D}$ is not finite-dimensioned or f not bounded is left to the reader. Another (very similar) way to consider the problem is to note that $d_H^{\mathcal{D}}(A,B)$ is the "natural" Hausdorff distance between the "*lifted*" shapes $A^{\mathcal{D}}$ and $B^{\mathcal{D}}$ where $A^{\mathcal{D}} = \{(\mathbf{x}, f(\mathbf{x},A)) \in \mathbb{R}^n \times \mathcal{D} \text{ for } \mathbf{x} \in A\}$ for the distance $d_{\mathbb{R}^n \times \mathcal{D}} = d_{\mathbb{R}^n} + d_{\mathcal{D}}$. □

3.2.4 A distance in the set of local descriptions: the local cross-correlation

Descriptors such as the color require to be associated with a different kind distance from the euclidean one. The relevant information when comparing two shapes in the case of the color descriptor is indeed more related to the local variation of the intensity than to the value of the intensity itself, because of usual phenomena such as contrast variations between two images of a same object.

The local cross-correlation is well suited for such descriptors. It is introduced in next section in the case of images but can easily be adapted to compare descriptors defined on two shapes at two (different) points.

3.3 Shapes as images, images as shapes

3.3.1 Full shapes and images

The previous sections have introduced a way of incorporating local descriptors into the Hausdorff distance, and a way to go easily from the distance between boundaries to the distance between full shapes thanks to their skeletons. One could be tempted to deduce, from a mixture of these two remarks, a new method to include considerations about the color (or other information) of the interior of the shapes into the distance between their (colored) boundaries (and skeletons). The point is that, in the "lifted" spaces where one or several axes are added to take the color into account, the boundary of a shape is not its former euclidean boundary; it turns out that its new boundary is the whole shape.

Full shapes with local descriptors defined on them are strongly similar to parts of images. It is then natural to study distances and matching processes between (parts of) images. One may even think of a shape as the whole image embedding the shape. In that case the notion of boundaries disappears, making room to the notion of high-gradient zones. A natural way to reintroduce contours consists in searching for the boundaries of homogeneous regions. A first attempt in that spirit could be to consider the level curves of the intensity. However these curves hardly reflect the boundaries of the real, intuitive shapes in the images for the already previously presented reason: the local changes of contrast often makes the level value of the intensity irrelevant, without mentioning that the texture (or even variation of texture in a single object) often causes as much variation of the intensity inside the object itself as between the object and the background. Therefore it is necessary to design carefully local descriptors of the image, such as color, color variation, texture descriptors, value of local filters such as wavelets, norm of the gradient of the intensity, etc. in order to search then for homogeneous regions. If the number of descriptors is high, then this may need statistical or learning tools in order to find which descriptors (or combination of them) are relevant for the current task. The definition of *homogeneity* itself is a challenging problem of great interest.

But, for the moment, let us study image distances. Given a bounded, regular domain Ω of the space $\mathbb{R}^n$, and a destination metric space $\mathcal{D}$ ($\mathbb{R}$ if the image is grey-leveled, or something more complicated for more complicated descriptors), an image is defined as a function defined on Ω with values in $\mathcal{D}$. For the sake of simplicity, $\mathcal{D}$ will be $\mathbb{R}$ in the sequel, and the usual measure $d\Omega(x)$ on Ω will be denoted by dx.

3.3.2 Local cross-correlation

The local cross-correlation is a criterion which takes into account the local changes of contrast which appear when comparing two images. From this correlation can be extracted a distance between images, if any two images that differ just in the contrast are considered to be the same.

Given a scale σ, the cross-correlation of two images A and B at point x is defined by:

$$CC(A, B, x) = \frac{v_{AB}(x)^2}{v_A(x)\ v_B(x)}$$

where $v_A(x)$ is the local spatial variance of A in a Gaussian neighborhood of size σ centered on x, and $v_{AB}(x)$ the local covariance of A and B on the same neighborhood, i.e. we define:

$$g(x, y) = e^{-frac\|x-y\|^2 2\sigma^2}$$

$$\mu(x) = \int_{y\in\Omega} g(x,y)\ dy$$

$$\bar{A}(x) = \frac{1}{\mu(x)} \int_{y\in\Omega} A(y)\ g(x,y)\ dy$$

$$v_A(x) = \epsilon + \frac{1}{\mu(x)} \int_{y\in\Omega} (A(y) - \bar{A}(x))^2\ g(x,y)\ dy$$

$$v_{AB}(x) = \frac{1}{\mu(x)} \int_{\Omega} (A(y) - \bar{A}(x))(B(y) - \bar{B}(x))\ g(x,y)\ dy$$

The positive constant ϵ is added only not to have a null divider in the expression of $CC(A,B,x)$. The criterion $CC(A,B,x)$ returns the relative correlation of variations of the two images in the neighborhood of point x. If σ is small, i.e. if the neighborhood is small, then these variations can be approximated by a hyperplane around x and consequently CC happens to be a function of the gradients of A and B at point x; more precisely it is a function of the angle between the two directions $\nabla A(x)$ and $\nabla B(x)$. When σ increases, CC gets more and more information on the neighborhood of x to compare locally A and B, which are then considered as intensity distributions.

The local cross-correlation on the whole images are defined by [60]:

$$LCC(A,B) = \frac{1}{|\Omega|} \int_{x\in\Omega} CC(A,B,x)\ dx$$

Note that this criterion is not a distance in itself, since its value increases with the similarity of the images.

3.3.3 Distance related to the local cross-correlation

The criterion gives values between 0 and 1, so $1 - LCC(A,B)$ could be a good candidate for a distance. It is in fact the square of a distance.

Lemma 47. *For any x, $1-\sqrt{CC(\cdot,\cdot,x)}$ is the square of a pseudo-distance, or more precisely of a distance on the set of functions with zero local mean and unit local variance at this point x.*

Proof. Let x be a point, and A and B two images such that $\bar{A}(x) = \bar{B}(x) = 0$ and $v_A(x) = v_B(x) = 1$. Then

$$\begin{aligned}
\sqrt{CC(A,B,x)} &= |v_{AB}(x)| \\
&= \frac{1}{\mu(x))} \left| \int_\Omega A(y)\, B(y)\ g(x,y)\ dy \right| \\
&= \frac{1}{\mu(x)} \left| \int_\Omega \frac{1}{2} \left(A(y)^2 + B(y)^2 - (A(y) - B(y))^2 \right) g(x,y)\ dy \right| \\
&= 1 - \frac{1}{2} \left\| (A-B)^2 \frac{g(x,\cdot)}{\mu(x)} \right\|_{L^1} \\
&= 1 - \frac{1}{2} \left\| (A-B) \sqrt{\frac{g(x,\cdot)}{\mu(x)}} \right\|^2_{L^2}
\end{aligned}$$

since the local variance at point x of A and B is 1. Therefore $1 - \sqrt{CC(A,B,x)} = \frac{1}{2}\left\|(A-B)\sqrt{\frac{g(x,\cdot)}{\mu(x)}}\right\|^2_{L^2}$. As $g(x,y) > 0$ for any y, $\|I\|_{g(x,\cdot)/\mu(x)} = \left\|I\sqrt{\frac{g(x,\cdot)}{\mu(x)}}\right\|_{L^2}$ is a norm for any x. Consequently, for any x, $1 - \sqrt{CC(\cdot,\cdot,x)}$ is the square of a distance on the specified set, or of a pseudo-distance on the whole set of all images. □

Lemma 48. *If $1 - \sqrt{f(\cdot,\cdot)}$ is the square of a distance, then $1 - f(\cdot,\cdot)$ is the square of a distance also.*

Proof. Let be $f(\cdot,\cdot)$ such that $1 - \sqrt{f(\cdot,\cdot)}$ is the square of a distance. Necessarily, f is symmetric and $0 \leqslant f \leqslant 1$. Therefore $0 \leqslant 1 - f \leqslant 1$.

- $1 - f(\cdot,\cdot)$ is symmetric;
- $\forall A, B: \quad 1 - f(A,B) = 0 \iff f(A,B) = 1 \iff 1 - \sqrt{f(A,B)} = 0 \iff A = B$
- $\forall A, B, C: \quad \sqrt{1 - f(A,B)} \leqslant \sqrt{1 - f(A,C)} + \sqrt{1 - f(C,B)}$. Indeed, if the distance $\sqrt{1 - \sqrt{f}}$ is denoted by d, then $f = (1-d^2)^2 = 1 - 2d^2 + d^4$ so that $\sqrt{1-f} = d\sqrt{2-d^2}$. Consequently it is sufficient to prove that $d(A,B) \leqslant d(A,C) + d(C,B)$ implies the same inequality with $d\sqrt{2-d^2}$ instead of d. With some more concise notations: it is sufficient to prove that for any real numbers a, b, c between 0 and 1, then $a \leqslant b + c$ implies $h(a) \leqslant h(b) + h(c)$ with $h : x \mapsto x\sqrt{2-x^2}$. As h is positive, increasing and concave between 0 and 1, the last assertion is true.

□

Corollary 49. *For any x, $1 - CC(\cdot,\cdot,x)$ is the square of a pseudo-distance, or more precisely of a distance on the set of functions with zero local mean and unit local variance at this point x.*

Hence the result:

Proposition 50. *$1 - LCC(\cdot,\cdot)$ is the square of a distance on the set of images quotiented by the relation of a change of contrast.*

Proof.

$$1 - LCC(\cdot,\cdot) = \frac{1}{\Omega}\int_\Omega 1 - CC(\cdot,\cdot,x)dx = \frac{1}{\Omega}\int_\Omega d_x^2(\cdot,\cdot)dx$$

where for each x, d_x is a pseudo-distance. This implies that the application $d = \sqrt{1 - LCC}$ still satisfies the triangular inequality: indeed, if A, B and C are three images, then:

$$\begin{aligned}
d^2(A,B) &= \frac{1}{\Omega}\int_\Omega d_x^2(A,B)dx \\
&\leqslant \frac{1}{\Omega}\int_\Omega \left(d_x(A,C) + d_x(C,B)\right)^2 dx \\
&\leqslant d^2(A,C) + d^2(C,B) + 2\,\frac{1}{\Omega}\int_\Omega d_x(A,C)\, d_x(C,B)dx \\
&\leqslant d^2(A,C) + d^2(C,B) + 2\sqrt{\frac{1}{\Omega}\int_\Omega d_x^2(A,C)dx}\sqrt{\frac{1}{\Omega}\int_\Omega d_x^2(C,B)dx} \\
&\leqslant \left(d(A,C) + d(C,B)\right)^2
\end{aligned}$$

Moreover, $d(A,B) = 0 \iff d_x(A,B) = 0$ *a.e.*, $d_x(A,B)$ being seen as a function of x. But $d_x(A,B) = 0$ implies $\forall y$ *a.e.*, $\frac{1}{v_A(x)}(A(y) - \bar{A}(x)) = \frac{1}{v_B(x)}(B(y) - \bar{B}(x))$. Consequently A and B are linked together by a simple change of contrast. Even if the initially chosen neighborhood was not Gaussian and did not induce a function $g(x,y)$ that satisfies that it is strictly positive everywhere, then, despite the fact that the set of images which are at distance 0 for the pseudo-distance d_x would be much bigger, it would have been sufficient that there exists for each x an open set including x such that $g(x,y) > 0$ on this open set to infer that A and B are locally linked by an affine application, i.e. linked on each connected component of Ω by an affine application, and consequently linked by a change of contrast if Ω is a connected set. □

3.3.4 Distances and deformation

As previously in the case of contours, there exists a diffeomorphic approach, where the distance between two images is the length of the shortest path from one to the other, the distance being the integral over the path of the norm of the infinitesimal deformation field for a suitable choice of norm. This is done in the work of Trouvé and Younes [115, 49, 41, 81, 112]. On the other side, the equivalent of the Hausdorff distance, i.e. a computationally straightforward distance, is the local cross-correlation. The L^2 norm of the difference between the intensity of the two images was rejected previously as being irrelevant.

There is an hybrid method joining diffeomorphisms and straightforward distances. It consists in defining a new criterion between any two images A and B:

$$R_{LCC}(A,B) = \inf_{\text{deformations } \mathbf{f}} \Big\{ R(\mathbf{f}) - \alpha LCC(\mathbf{f}(A), B) \Big\}$$

where $R(\mathbf{f})$ is a regularizing term and α a positive number. This induces a very natural way to match the image A onto the image B, that consists in minimizing, by gradient descent to respect to $\mathbf{f}$, the quantity $R(\mathbf{f}) - \alpha LCC(\mathbf{f}(A), B)$ which takes into account the resemblance between the warped image $\mathbf{f}(A)$ and its target B (with weight α), and the cost of the deformation.

The field $\mathbf{f}$ should be smooth enough and consequently invertible, i.e. it should be a diffeomorphism from the rectangular subset Ω to itself, which can be guaranteed by the assumption that the diffeomorphism $\mathbf{f}$ equals the identity on the image boundary $\partial\Omega$. Other possibilities are offered by extending the images to a larger subset Ω_1 in order to let the boundary of Ω evolve.

In order to keep $\mathbf{f}$ continuous, the regularizing term $R(\mathbf{f})$ is necessary. It can be for example $R(\mathbf{f}) = \|\mathbf{f} - Id\|^{H^1}_{\Omega}$ where Id is the identity function on Ω and $\|a\|^{H^1}_{\Omega} = \int_{x\in\Omega} \|a(x)\|^2 + \|Da(x)\|^2 \, dx$. If one prefers to be sure that $\mathbf{f}$ remains invertible, the choice of $\|\mathbf{f} - Id\|^{H^1}_{\Omega} + \|\mathbf{f}^{-1} - Id\|^{H^1}_{\Omega}$, where $\mathbf{f}^{-1}$ is the inverse of $\mathbf{f}$, presents also the advantage of being symmetric when swapping A for B.

Note however that $R_{LCC}(A,B)$ has no reason to be symmetric (because of the Jacobian of $\mathbf{f}$ that is not present in the expression $LCC(\mathbf{f}(A), B)$) and consequently is not a distance. A simple way to correct this would be to add $LCC(A, \mathbf{f}^{-1}(B))$ inside the infimum. Concerning the triangular inequality, it still stands if R satisfies some properties concerning the composition law of deformations, more precisely $R(\mathbf{g} \circ \mathbf{f}) \leqslant R(\mathbf{g}) + R(\mathbf{f})$.

3.3.5 Multiscale

The criterion $R_{LCC}(A, B)$ depends on two parameters: α and σ. The parameter σ is the standard deviation of the local Gaussian neighborhood, and α is the weight between the smoothing and the resemblance terms. A small value of σ induces a higher precision in the comparison of $\mathbf{f}(A)$ to B but, during the matching process by minimization of the criterion with respect to f, it will lack important, global information if $f(A)$ is not very close to B: this is the everlasting problem of gradient descents which get stuck in local minima. On the contrary, a very high value of σ will allow a global comparison of the two images but will lack precision. Concerning α, the lower the value, the smoother the deformation field $\mathbf{f}$, which means that when α is low, the field f expresses a global coherent deformation, whereas a high value of α induces local, quasi-independent, irregular deformations. It is consequently relevant to let α be correlated with σ: global deformations are obtained for a high σ and a low α, whereas high-precision deformations when ending the convergence of an evolution are obtained with a low σ and a high α.

A multiscale approach appears to be suitable here. Either a new criterion is defined as a weighted sum (in practice, the successive considerations) of $R_{LCC}(A, B)$ for different values of the parameters, either the parameters are constant but the images are considered successively at different scales. The two approaches are very similar and, depending on the design of the criterion, even identical.

Chapter 4

Shape Statistics Based on the only knowledge of Distances

Abstract

This chapter deals with the statistics that can be extracted from the only knowledge of the distance matrix on a sample set of shapes. The approach is based on the Hausdorff distance between shapes, which choice is itself not fundamental since the same framework could be applied with another distance. The shape distances are used to represent the sample set of shapes by a graph, which with the technique of the graph Laplacian leads to a way of projecting shapes onto a low dimensional space.

4.1 Exact representation of a set of shapes in a finite dimensioned vector space

In the previous chapters have been introduced general sets of shapes and associated distances which can be straightforwardly computed (without building paths between shapes or searching for an infimum) and which are intrinsic in the sense that they do not depend on the parameterization of the shapes. These shapes are for instance hypersurfaces with the Hausdorff distance or images with the local cross-correlation. Now, given a real sample set of shapes, denoted by $\mathcal{D} = \{\Gamma_i \ / \ 1 \leqslant i \leqslant n\}$, included in one of these studied sets, the mutual distance $d(\Gamma_i, \Gamma_j)$ between any pair of shapes can be computed. From the only knowledge of this distance matrix, is it possible to build some shape statistics, without any shape warping ?

Without shape warping, the natural general set of shapes into which $\mathcal{D}$ is included can not be explored and its properties are not useful. Thus, the only knowledge of distances leads to a purely static geometrical point of view of the sample set $\mathcal{D}$, which is then often naturally seen as a cloud of points in a finite dimensioned vector space $\mathbb{R}^k$ embedded with the euclidean distance. For a set of n shapes, which are most often in an intrinsically infinite dimensioned space, the dimension of this embedding space is often $k \simeq n$. However, generally speaking, it is not always even possible to build such a cloud of points. For example, the set of positive integers written with the binary notation can be equipped with the distance that associates to any two such binary integers the number of digits that are different in the two writings. The distance between 0000 and any of $0010, 0100, 1000$ is one,

whereas the distance between any two of $0010, 0100, 1000$ is two. This is problematic since there is no such a quartet of points in any $\mathbb{R}^k$ for the euclidean distance L^2. Note however that it is feasible with the L^1 distance.

Anyway, even when embedding the sample set $\mathcal{D}$ in a $\mathbb{R}^k$ is possible, the question of interest is to find structure in $\mathcal{D}$, like principal axes, and to build statistics. In this perspective, a good start would be to search for the best map of $\mathcal{D}$ into a $\mathbb{R}^k$, best in the sense that the map should preserve as much as possible the distances and that the coordinate axes should be also the principal axes.

4.2 Graph Laplacian and best approximative map

It is possible to build such a map using the graph Laplacian technique [5]. The graph Laplacian technique or similar methods have already been applied recently to images for the L^2 distance between their grey levels by Tenenbaum et al [109] or Donoho and Grimes [37, 38]. However it does not appear to have been done for shapes (as boundaries).

The trick consists first in building a neighborhood graph. For a choice of a positive integer K, the search for the K nearest neighbors $N^i_{1 \leqslant l \leqslant K}$ of each shape Γ_i leads to the definition of a symmetric weight matrix W:

$$W_{i,j} = \delta_{i,j} \ e^{-\frac{d(\Gamma_i,\Gamma_j)^2}{2\sigma^2}}$$

where

$$\delta_{i,j} = \begin{cases} 1 & \text{if} \quad i \in N^j \quad \text{or} \quad j \in N^i \\ 0 & \text{otherwise} \end{cases}$$

and we have chosen for σ the mean distance between neighbors:

$$\sigma = \frac{\sum_{i,j} \ d(\Gamma_i, \Gamma_j) \ \ \delta_{i,j}}{\sum_{i,j} \ \delta_{i,j}}.$$

Then, the symmetric negative semi-definite matrix $L = W - D$ where $D_{i,j} = \sum_i \ W_{i,j} \ \delta_{i,j}$ is a discrete approximation of the Laplacian operator. Thus, as explained in [5, 4], its eigenvectors F_k of highest (negative) non-zero eigenvalues are the best functions from the shapes Γ_i to $\mathbb{R}$ that could be used as coordinate system of the set of shapes in the sense that they are the smoothest functions f from $\mathcal{D}$ to $\mathbb{R}$: they minimize $\sum_i (f(i) - f(j))^2 \, W_{i,j}$ for a fixed unit L^2 norm of f. Moreover, since they are eigenvectors of the Laplacian operator, they are orthogonal. We obtain consequently a map in $\mathbb{R}^m$ where each shape Γ_i is represented by a dot with coordinates $(F_{1 \leqslant k \leqslant m}(\Gamma_i))$, the first coordinates being the most significant ones.

4.3 Examples of maps

Let us try this approach on an artificial dataset. We build a set of rectangles with same center and width but different lengths and orientations, so there are two natural parameters we would expect the algorithm to find. The orientation varies between $-\frac{\pi}{6}$ and $+\frac{\pi}{6}$, and the length between 2 and 4 times the width. Rectangles are randomly chosen such that the distribution of their corners is the uniform law in the authorized area. Results vary depending on the distribution density and the value of K: the higher the density, the better

the results. Fig. 4.1 has been computed for 700 rectangles and shows that the representation of the rectangles by their two first coordinates $(F_{1,2}(\Gamma_i))$ leads to a natural map of the sample set. For sparse data however, the results are globally coherent in the sense that the points are relatively well ordered, but the built map shows too much irregularity.

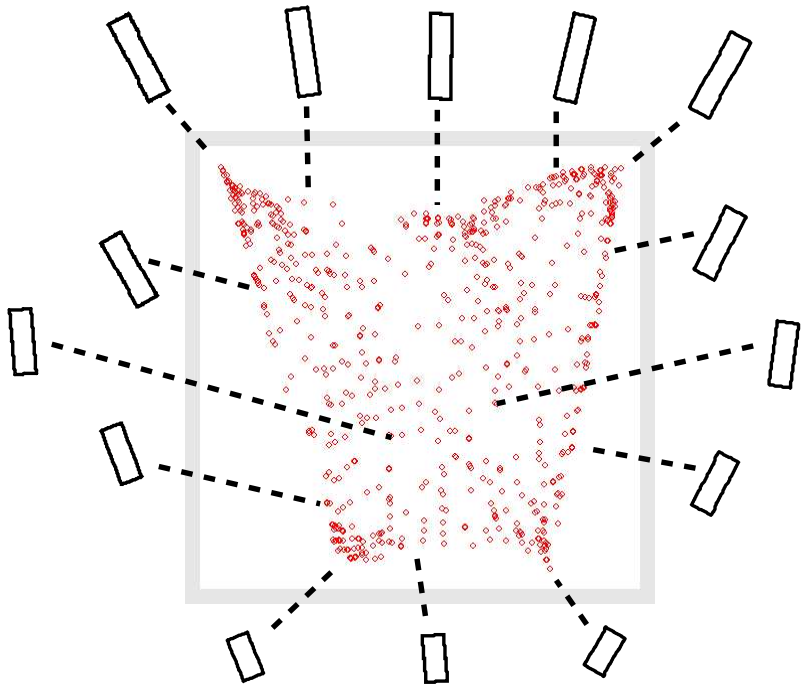


Figure 4.1: Map from the graph Laplacian method for a set of rectangles whose length and orientation have been chosen randomly ($K = 15$).

Let us now study the more complicated case of some different classes in a same connected component. We consider a set of 111 fish from the database (`www.ee.surrey.ac.uk/Research/VSSP/imagedb/demo.html`) of fish silhouettes collected by the researchers of the University of Surrey at the center for Vision, Speech and Signal Processing (`www.ee.surrey.ac.uk/Research/VSSP`). This database contains 1100 silhouettes. The resulting map for the two first coordinates (see Fig. 4.2) shows some clusters of fish families.

In conclusion, this technique is interesting in the case of high density database to build maps, and in the case of sparse data to let clusters appear. Note that there exist many techniques similar to the Laplacian eigenmaps, for instance LLE (locally linear embedding), LPP (locality preserving projections), Isomap or Hessian eigenmaps [99, 38, 37, 59, 58, 30].

To go further on in this direction, one might get inspiration from recent works by Memoli and Sapiro[79, 78] on point cloud data and distance functions on submanifolds, or by Lafon et al.[72, 71] on diffusion maps, in order to extract from spectral techniques a way to perform classification tasks, or, further, to build a shape prior for segmentation. We will not investigate the spectral direction more here; however for this last aim (image segmentation), the use of such a graph technique will also require to *move* in the set of shapes, i.e., to perform shape evolutions. This is the subject of the next part.

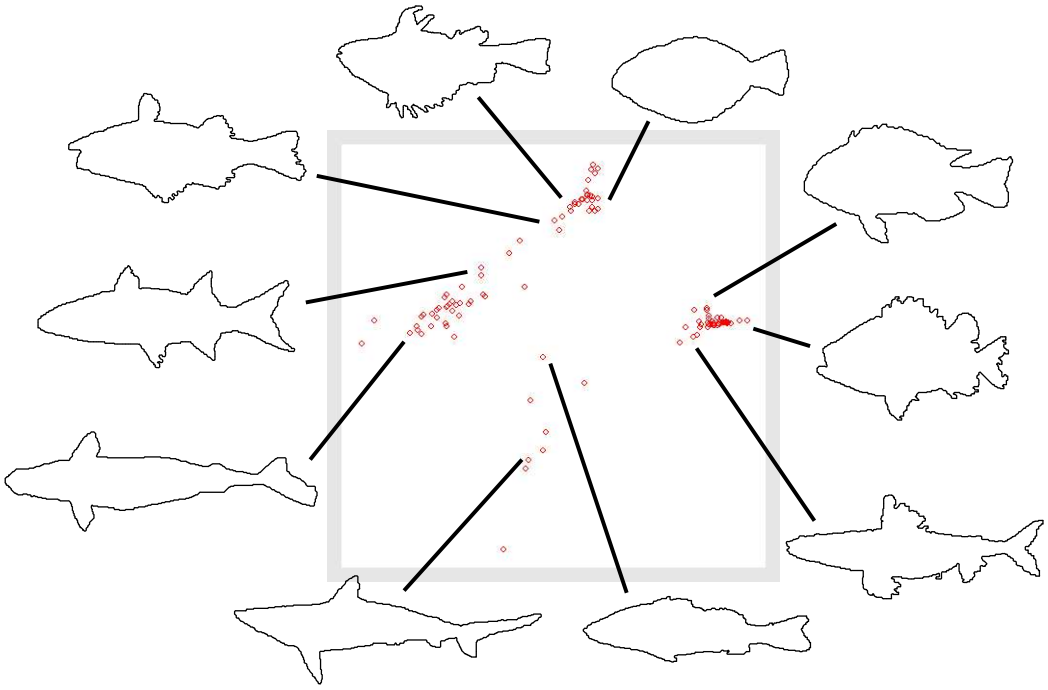


Figure 4.2: Two first coordinates for a set of 111 fish from different classes. The elements from each family are got together into clusters ($K = 25$).

Part II

Shape Warping

Chapter 5

Shape Gradient and Shape Warping

Abstract

This chapter is dedicated to warping a shape onto another one using the minimization of the distance between them with respect to the shape in a gradient descent framework. The differentiation of an energy with respect to a curve is introduced and applied to the case of the approximation of the Hausdorff distance. Results of warping are shown.

5.1 Deforming shapes

The problem of continuously deforming a shape so that it turns into another is central to this chapter. The reasons for this will become more clear in the sequel. Let us just mention here that it can be seen as an instance of the warping problem: given two shapes Ω_1 and Ω_2, how do I deform Ω_1 onto Ω_2? The applications in the field of medical image processing and analysis are immense (see for example [111, 110]). It can also be seen as an instance of the famous (in computer vision) correspondence problem: given two shapes Ω_1 and Ω_2, how do I find the corresponding point P_2 in Ω_2 of a given point P_1 in Ω_1? Note that a solution of the warping problem provides a solution of the correspondence problem if we can track the evolution of any given point during the smooth deformation of the first shape onto the second.

In order to make things more quantitative, we assume that we are given a function $E : \mathcal{C}_0 \times \mathcal{C}_0 \to \mathbb{R}^+$, called the Energy, which is continuous on $\mathcal{S} \times \mathcal{S}$ for one of the shape topologies of interest. This Energy can also be thought of as a measure of the dissimilarity between the two shapes. By smooth, we mean that it is continuous with respect to this topology and that its derivatives are well-defined in a sense we now make more precise.

5.1.1 Derivatives of an energy and gradient

We first need the notion of a normal deformation flow of a curve Γ in $\mathcal{S}$. This is a smooth (i.e. C^2) function $\beta : [0,1] \to \mathbb{R}$ (when $\Gamma \in \mathcal{S}^c$, one further requires that $\beta(0) = \beta(1)$). Let $\Gamma : [0,1] \to \mathbb{R}^2$ be a parameterization of Γ, $\mathbf{n}(p)$ the unit normal at the point $\Gamma(p)$ of Γ; the normal deformation flow β associates the point $\Gamma(p) + \beta(p)\mathbf{n}(p)$ to $\Gamma(p)$. The resulting shape is noted $\Gamma + \boldsymbol{\beta}$, where $\boldsymbol{\beta} = \beta\mathbf{n}$. There is no guarantee that $\Gamma + \boldsymbol{\beta}$ is still a shape in $\mathcal{S}$ in general but if β is such that $\beta\mathbf{n}$ is C^2 (β being C^2 is not sufficient since $\mathbf{n}$ may be only C^1) and ε is small enough, then $\Gamma + \varepsilon\boldsymbol{\beta}$ is in $\mathcal{C}_2$. Given two shapes Γ and Γ_0, the corresponding

Energy $E(\Gamma, \Gamma_0)$, and a normal deformation flow β of Γ, the Energy $E(\Gamma + \varepsilon\boldsymbol{\beta}, \Gamma_0)$ is now well-defined for ε sufficiently small. The derivative of $E(\Gamma, \Gamma_0)$ with respect to Γ in the direction of the flow $\boldsymbol{\beta}$ is then defined, when it exists, as

$$\mathcal{G}_\Gamma(E(\Gamma, \Gamma_0), \boldsymbol{\beta}) = \lim_{\varepsilon \to 0} \frac{E(\Gamma + \varepsilon\boldsymbol{\beta}, \Gamma_0) - E(\Gamma, \Gamma_0)}{\varepsilon} \tag{5.1}$$

This kind of derivative is also known as a Gâteaux semi-derivative. In our case the function $\beta \to \mathcal{G}_\Gamma(E(\Gamma, \Gamma_0), \boldsymbol{\beta})$ is linear and continuous (it is then called a Gâteaux derivative) and defines a continuous linear form on the vector space of normal deformation flows of Γ. This is a vector subspace of the Hilbert space $L^2(\Gamma)$ with the usual Hilbert product $\langle \beta_1, \beta_2 \rangle = \frac{1}{|\Gamma|} \int_\Gamma \beta_1 \beta_2 = \frac{1}{|\Gamma|} \int_\Gamma \beta_1(x)\beta_2(x)\, d\Gamma(x)$, where $|\Gamma|$ is the length of Γ. Given such an inner product, we can apply Riesz's representation theorem [100] to the Gâteaux derivative $\mathcal{G}_\Gamma(E(\Gamma, \Gamma_0), \boldsymbol{\beta})$: There exists a deformation flow, noted $\nabla E(\Gamma, \Gamma_0)$, such that

$$\mathcal{G}_\Gamma(E(\Gamma, \Gamma_0), \boldsymbol{\beta}) = \langle \nabla E(\Gamma, \Gamma_0), \beta \rangle.$$

This flow is called the gradient of $E(\Gamma, \Gamma_0)$.

We now return to the initial problem of smoothly deforming a curve Γ_1 onto a curve Γ_2. We can state it as that of defining a family $\Gamma(t)$, $t \geq 0$ of shapes such that $\Gamma(0) = \Gamma_1$, $\Gamma(T) = \Gamma_2$ for some $T > 0$ and for each value of $t \geq 0$ the deformation flow of the current shape $\Gamma(t)$ is equal to minus the gradient $\nabla E(\Gamma, \Gamma_2)$ defined previously. The usual descent gradient framework leads to the following PDE:

$$\begin{aligned} \Gamma_t &= -\nabla E(\Gamma, \Gamma_2)\mathbf{n} \\ \Gamma(0) &= \Gamma_1 \end{aligned} \tag{5.2}$$

However such a PDE supposes a definition of Γ_t, which has not been introduced yet.

5.1.2 Derivatives in the metric space $\mathcal{S}$

The set $\mathcal{S}$ of all shapes which is considered has now two different structures. It first has the structure of a metric space, with the Hausdorff distance between shapes, whose induced topology was previously shown to be equivalent to two other proposed topologies (theorem 13). But now it also has a structure similar to the one of a manifold, with for each shape Γ a set of admissible deformation fields (with associated shapes) which remembers the notion of chart or tangent space with its inner product, here $L^2(\Gamma)$. We have indeed introduced for each shape Γ a mean to associate to any deformation field $\boldsymbol{\beta}$ the resulting shape $\Gamma + \boldsymbol{\beta}$, and we can conjecture that for each shape Γ (in the interior of $\mathcal{S}$ in the sense that the inequality $reach(\Gamma) > h_0$ is strict) there exists a positive real number ε (depending on $reach(\Gamma)$) such that if the deformation field $\boldsymbol{\beta}$ is C^2 and if its amplitude $\|\boldsymbol{\beta}\|_{L^\infty}$ and its second derivative along the shape $\|d^2\boldsymbol{\beta}\|_{L^\infty}$ are upper-bounded by ε, then the resulting shape $\Gamma + \boldsymbol{\beta}$ is in $\mathcal{S}$. Anyway we will define the set of all admissible deformation fields as the set of all deformation fields which are normal to Γ and which application leads to a shape in $\mathcal{S}$. The interest of fields with a tangential component is limited here since the tangential component does not alter the shape.

We will not try here to prove that the set $\mathcal{S}$ with these structures *is* a manifold for any meaning of the term *manifold* in infinite dimensions (for example Hilbert, Banach or Finsler manifolds) but we will introduce successively the tools that we need. As it has been defined, the set of admissible deformation fields is here even not a vector space since the application

of a field with too high second derivative can induce a new shape with a too high maximum curvature which is consequently not in $\mathcal{S}$, so that the multiplication of an admissible field by a real number greater than 1 can lead to a non-admissible field. Nevertheless the set of admissible deformation fields has locally the structure of a vector space. A proper definition of a *tangent space* with considerations on directions of paths in $\mathcal{S}$ would deal with that and lead to a vector space, but it would not be complete and hence not an Hilbert space and it is out of the scope of this chapter.

The previous definition of the sum of a shape and a small-normed deformation field leads to a natural way of defining correspondences between close enough shapes. Indeed the proposition 24 in chapter 2 introduces a continuous point-to-point correspondence between any two close curves thanks to their distance functions, and this result can be extended to the case of higher-dimensioned shapes. This associates to any shape Γ' near Γ ("near" in the sense of the Hausdorff distance) a normal deformation field $\mathbf{v}_{\Gamma\to\Gamma'}$ defined on Γ such that $\Gamma + \mathbf{v}_{\Gamma\to\Gamma'} = \Gamma'$, and such a field is necessarily unique, so this induces a simple relation between the set of admissible deformation fields on Γ and the neighborhood of Γ in $\mathcal{S}$. Note that the field $\mathbf{v}_{\Gamma\to\Gamma'}$ is invertible but its inverse is rarely $\mathbf{v}_{\Gamma'\to\Gamma}$ since this last one is normal to Γ' and not to Γ.

Then, how to define the derivative with respect to a real t of a family of shapes $\Gamma(t)$ indexed by t ? The intuitive notion of being tangent to a path is the one of a direction which best fits locally the path $\Gamma(t)$. The application $\boldsymbol{\beta} \mapsto \Gamma' = \Gamma + \varepsilon\boldsymbol{\beta}$ and its inverse function $\Gamma' \mapsto \mathbf{v}_{\Gamma\to\Gamma'}$ define the notion of direction. However the notion of "best fitting" can here have several meanings: one for the Hausdorff metric and one for the norm of deformation fields, related to the inner product $L^2(\Gamma)$. In usual (euclidean) spaces, the derivative would have been defined as $\dfrac{d}{dt}\Gamma(t) = \lim\limits_{\varepsilon\to 0} \dfrac{1}{\varepsilon}\left[\Gamma(t+\varepsilon) - \Gamma(t)\right]$ but the subtraction of shapes makes no sense here, so this definition should rather be taken upside down and rewritten

$$d\left(\Gamma(t+\varepsilon),\ \Gamma(t) + \varepsilon\boldsymbol{\beta}\right) = o(\varepsilon) \tag{5.3}$$

if such a deformation field $\frac{d}{dt}\Gamma(t) = \boldsymbol{\beta}$ exists and is unique. The distance d here can either stand for the Hausdorff distance between the shapes $\Gamma(t+\varepsilon)$ and $\Gamma(t)+\varepsilon\boldsymbol{\beta}$, or for the L^2 norm of the deformation field $\left\|\mathbf{v}_{\Gamma(t+\varepsilon)\to\Gamma(t)+\varepsilon\boldsymbol{\beta}}\right\|_{L^2(\Gamma(t+\varepsilon))}$ or even $\left\|\mathbf{v}_{\Gamma(t)+\varepsilon\boldsymbol{\beta}\to\Gamma(t+\varepsilon)}\right\|_{L^2(\Gamma(t)+\varepsilon\boldsymbol{\beta})}$. Happily, all these choices lead to equivalent definitions ! This is essentially due to the equivalence of the studied topologies on $\mathcal{S}$. Indeed, by definition of the Hausdorff distance and by construction of the normal field $\mathbf{v}_{\Gamma\to\Gamma'}$ for any two close enough shapes Γ and Γ':

$$d_H(\Gamma,\Gamma') = \sup\left(\left\|\mathbf{v}_{\Gamma\to\Gamma'}\right\|_{L^\infty(\Gamma)},\ \left\|\mathbf{v}_{\Gamma'\to\Gamma}\right\|_{L^\infty(\Gamma')}\right).$$

The corollary 43 even gives:

$$d_H(\Gamma,\Gamma') = \left\|\mathbf{v}_{\Gamma\to\Gamma'}\right\|_{L^\infty(\Gamma)} = \left\|\mathbf{v}_{\Gamma'\to\Gamma}\right\|_{L^\infty(\Gamma')}.$$

But, as the curvature of all shapes $\mathcal{S}$ is bounded by a same constant number, so is the second derivative of admissible fields $\mathbf{v}_{\Gamma\to\Gamma'}$, and as the area (or length) of shapes in $\mathcal{S}$ in a neighborhood (for the Hausdorff distance) of any Γ is also bounded, the L^2 norm of admissible fields is equivalent to their L^∞ norm (same considerations as in the comparison of the three topologies on $\mathcal{S}$)!

Thus, all the introduced definitions of the derivative of a path with respect to its parameter are the same. The problem of the uniqueness of the derivative can then be

solved. If $\mathbf{v}$ and $\mathbf{v}'$ are two fields satisfying equation 5.3, then on one side $\left\|\varepsilon\mathbf{v} - \varepsilon\mathbf{v}'\right\|_{L^\infty} = O\left(d_H\big(\Gamma(t)+\varepsilon\mathbf{v}, \Gamma(t)+\varepsilon\mathbf{v}'\big)\right)$ and on the other side the triangular inequality gives

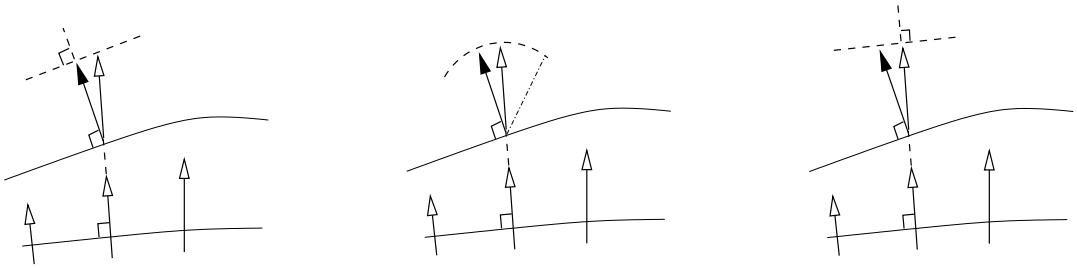
$$d_H\big(\Gamma(t)+\varepsilon\mathbf{v}, \Gamma(t)+\varepsilon\mathbf{v}'\big) \leqslant d_H\big(\Gamma(t)+\varepsilon\mathbf{v}, \Gamma(t+\varepsilon)\big) + d_H\big(\Gamma(t)+\varepsilon\mathbf{v}', \Gamma(t+\varepsilon)\big) \leqslant o(\varepsilon)$$

so that $\varepsilon \left\|\mathbf{v} - \mathbf{v}'\right\|_{L^\infty} = o(\varepsilon)$, which implies $\mathbf{v} = \mathbf{v}'$.

The derivative of a path $\Gamma(t)$ with respect to t is consequently well defined and so is the gradient descent (5.2). We do not address here the question of the existence of solutions to this equation.

5.1.3 A word on transports

Another interesting notion in usual (finite-dimensioned) differential geometry is the one of being Riemannian. This notion expresses the continuity of the inner product and its derivatives with respect to the point of the manifold. The investigation of the continuity of $L^2(\Gamma)$ or its derivatives with respect to Γ supposes a mean to "transport" any deformation field defined on Γ onto any close shape. One possibility would be to define a local transport by associating to any deformation field $\mathbf{w}$ on Γ the field $\mathbf{w} \circ \mathbf{v}_{\Gamma\to\Gamma'}^{-1}$ defined on Γ'. However the transported field $\mathbf{w} \circ \mathbf{v}_{\Gamma\to\Gamma'}^{-1}$ is supposed to be normal to the shape Γ', so a choice should be made to suitably adapt $\mathbf{w} \circ \mathbf{v}_{\Gamma\to\Gamma'}^{-1}(x)$ so that it becomes a vector which is collinear to $\mathbf{n}(x) = \mathbf{n}_{\Gamma'}(x)$, for example $\left[\mathbf{w} \circ \mathbf{v}_{\Gamma\to\Gamma'}^{-1}(x) \cdot \mathbf{n}(x)\right] \mathbf{n}(x)$, or $\left[(\mathbf{w} \cdot \mathbf{n}_\Gamma)(\mathbf{v}_{\Gamma\to\Gamma'}^{-1}(x))\right] \mathbf{n}(x)$, or even $\dfrac{\|\mathbf{w} \circ \mathbf{v}_{\Gamma\to\Gamma'}^{-1}(x)\|^2}{\mathbf{w} \circ \mathbf{v}_{\Gamma\to\Gamma'}^{-1}(x) \cdot \mathbf{n}(x)} \mathbf{n}(x)$.



1 - Projection

2 - Constant norm

3 - Component preservation

Figure 5.1: Illustration of some possible transports

Each of these choices can be justified and leads to qualitatively different transports. The second one keeps constant the norm of the normal deformation, whereas the first one projects the vector $\mathbf{w}(\mathbf{v}_{\Gamma\to\Gamma'}^{-1}(x))$ onto the new normal $\mathbf{n}(x)$. The third one keeps constant the component of the deformation in the direction of the initial field. The choice of the "local transport" is important in practice at the computational level, for example when following a "constant" direction in order to build a geodesic (for the L^2 structure) since it defines the notion of being "constant". By the way, note that the choice of transport may also act on the meaning of the second derivative with respect to the time t and consequently on the expression of the PDE which rules geodesics. Indeed, the notion of the continuity and derivatives of a field of deformation fields defined on $\mathcal{S}$ with respect to the shape, the notion of transport of a deformation field and the notion of variation of the inner product $L^2(\Gamma)$ with respect to Γ are strongly linked. The associated key-word in differential geometry

is (Levi-Civita) *connection*. Happily, the three local transports described above become arbitrarily close when the distance between the shapes tends to zero: given an admissible field $\mathbf{v}_1$ and any field $\mathbf{v}_2$ defined on Γ, and a small ε, the results of the different transports of $\mathbf{v}_2$ on $\Gamma+\varepsilon\mathbf{v}_1$ from a point $x \in \Gamma$ are all normal to the new shape and their respective length are $\mathbf{v}_2(x)$ and $\mathbf{v}_2(x)\sqrt{1 \pm \varepsilon^2\|d_x\mathbf{v}_1(x)\|^2} \simeq \mathbf{v}_2(x)(1 \pm 2\varepsilon^2\|d_x\mathbf{v}_1(x)\|^2)$. The order 2 in ε of the difference between them is important. The first consequence is that the derivative of a field of deformation fields over $\mathcal{S}$ at shape Γ in the direction $\mathbf{v}_1$ can be defined and this derivative does not depend on the choice among the three local transports. The second consequence is that the transport of any field over any (smooth enough) path can be defined and that it is also independent from the choice of the local transport. Indeed, given two fixed shapes Γ, Γ' and a path between them, any discretization of the path into a finite sequence of enough close shapes Γ_n leads to a way to associate to any field defined on Γ a new field on Γ' by successively "locally transporting" the field from Γ_n to Γ_{n+1}. Then the resulting field on Γ' depends less and less on the choice of the "local transport" when the discretization gets finer. We can conjecture that if the path is smooth enough (i.e. if its second derivative is bounded) then the resulting field converges when the discretization gets finer, and this defines the transport of the field over the given path. A third remark concerns the non-conservation of the L^2 metric. Indeed, if two points on a shape are tracked when its undergoes a deformation, the arclength between them changes, but this time the order of the variation is not 2 but 1, so that the inner product between two fields $\mathbf{v}_1, \mathbf{v}_2$ defined on Γ and transported on $\Gamma + \mathbf{v}_3$ (with $\mathbf{v}_3$ small enough) changes also and is $\simeq \displaystyle\int_\Gamma (\mathbf{v}_1 \cdot \mathbf{v}_2)\, d\Gamma + \int_\Gamma (\mathbf{v}_1 \cdot \mathbf{v}_2)(\mathbf{v}_3 \cdot \mathbf{n})\, \kappa\, d\Gamma$. If this phenomenon is not desirable, then it is sufficient to change the local transport and multiply the previous one by $(1 - \mathbf{v}_{\Gamma\to\Gamma'} \cdot \kappa\mathbf{n}/2)$, so that the L^2 energy $\int \mathbf{v}^2 d\Gamma$ is conserved when a field $\mathbf{v}$ is transported on any shape near Γ. Of course such a choice would modify the meaning of the derivative of a field of deformation fields.

5.1.4 A word on geodesics

Coming back to the gradient descent framework 5.2 that we would like to apply in order to warp any shape onto any other one, natural candidates for the Energy function E to be minimized are the distances defined in section 1.2.2. The problem we are faced with is that none of these distances are Gâteaux differentiable. This is why in the previous part a chapter was devoted to the definition of smooth approximations of some of them.

When E is a distance between shapes, it could be interesting to know whether the evolution path obtained by the gradient descent is a geodesic or not. A geodesic between two shapes Γ_0 and Γ_1 is here defined as being a path $\Gamma(t)$ with fixed extremities $\Gamma(0) = \Gamma_0$ and $\Gamma(1) = \Gamma_1$ that is a local minimum of $\int_0^1 \|\Gamma_t(t)\|_N\, dt$ for a chosen norm $\|\cdot\|_N$. The global infimum of this quantity over such paths is called the geodesic distance between Γ_0 and Γ_1:

$$d_N^G(\Gamma_0, \Gamma_1) = \inf_{\Gamma,\, \Gamma(0)=\Gamma_0,\, \Gamma(1)=\Gamma_1} \int_0^1 \|\Gamma_t(t)\|_N\, dt.$$

In mathematical morphology a different notion of geodesics is used. For instance in [101] Serra considers geodesics in a metric space, which are defined as parametrized paths Γ such that:

$$\forall t,\ \ d\big(\Gamma(0), \Gamma(t)\big) + d\big(\Gamma(t), \Gamma(1)\big) = d\big(\Gamma(0), \Gamma(1)\big).$$

Such geodesics will here be referred as geometric geodesics whereas the previously introduced ones will be refereed as path-based geodesics.

Proposition 51. *Any path-based geodesic is a geometric geodesic.*

Proof. If $\Gamma(\cdot)$ is a path-based geodesic for the norm N between Γ_0 and Γ_1, then:

$$\begin{aligned} d_N^G(\Gamma_0,\Gamma_1) &= \int_0^1 \|\Gamma_t(t)\|_N \, dt \\ &= \int_0^u \|\Gamma_t(t)\|_N \, dt + \int_u^1 \|\Gamma_t(t)\|_N \, dt \\ &= \int_0^1 \left\| \frac{1}{u}\Gamma_{t'}(ut') \right\|_N u \, dt' + \int_0^1 \left\| \frac{1}{1-u}\Gamma_{t'}((1-u)t'+u) \right\|_N (1-u) dt' \\ &= \int_0^1 \left\|\Gamma_{t'}^{0\to u}(t')\right\|_N dt' + \int_0^1 \left\|\Gamma_{t'}^{u\to 1}(t')\right\|_N dt' \\ &\geqslant d_N^G\big(\Gamma(0),\Gamma(u)\big) + d_N^G\big(\Gamma(u),\Gamma(1)\big) \end{aligned}$$

for any u. The triangular inequality upgrades this inequality into an equality. □

The reciprocal proposition is not true. For example, for $N = L^\infty$, consider two quadrilaterals that share two whole sides: only one corner differs, and moving this corner from one location to the other builds a path of quadrilaterals which is a $d_{L^\infty}^G$ geodesic for the two meanings of "geodesic". But when the geodesic evolution is stopped at middle path between the two quadrilaterals, there exist deformations that keep constant the distance to the two quadrilaterals, for instance the one consisting in moving a little the opposite corner and then moving it back to its usual position. These deformations have a non-zero L^∞ cost and adding them to the path increases the integral expression (so that the new path is not a path-based geodesic) but as the distance to the two extremities is kept constant, the new path is still a geometric geodesic.

If a distance is L^2 differentiable, that is to say if there exists a gradient for the canonical $L^2(\Gamma)$ inner product defined on any shape, then the gradient descent method 5.2 can be applied in order to build a path between two shapes. The question is: is this path a geodesic, at least in the geometrical sense ? And the answer is no in general. Indeed the gradient which is computed is related to a particular inner product (generally L^2) which is itself in general not related to the norm N from which the minimized distance $d = d_N^G$ is built. More precisely, when walking one step forward in the direction of the opposite of the gradient, the distance to the target shape decreases, and the distance to the previous shape (infinitesimally close) increases, but this distance variations have no reason to be equal one to the other. Indeed the first one is $\mathcal{G}_\Gamma\big(d(\Gamma,\Gamma_1),\ \varepsilon\Gamma_t(t)\big) = \langle \nabla_\Gamma d(\Gamma,\Gamma_1) \,|\, \varepsilon\nabla_\Gamma d(\Gamma,\Gamma_1)\rangle_{L^2} = \varepsilon \left\|\nabla_\Gamma d(\Gamma,\Gamma_1)\right\|_{L^2}^2$ whereas the second one is $\|\varepsilon\Gamma_t(t)\|_N = \varepsilon \left\|\nabla_\Gamma d(\Gamma,\Gamma_1)\right\|_N$. If N and L^2 are not obviously related, then, in the case where the geodesic to the target Γ_1 is unique and the corresponding direction is $\boldsymbol{\beta}$, it may happen that the L^2 norm of this direction $\|\boldsymbol{\beta}\|_{L^2}$ is really high so that the gradient will prefer another direction with a smaller L^2 norm. To make things more explicit, imagine that N is the norm of the Sobolev space H^1. Then the direction $\boldsymbol{\beta}$ is constrained to be smooth enough, but $\nabla_\Gamma d(\Gamma,\Gamma_1)$ is not and can have a very high H^1 norm, so it may be a typical irregular deformation from L^2 if such a deformation is shown to decrease $d(\Gamma,\Gamma_1)$ faster than $\boldsymbol{\beta}$ (seen as a function in L^2). It will be the best L^2 deformation such that once applied to Γ, the resulting shape is as close as possible to the target for $d_{H^1}^G$, but without considering that this deformation could add an heavy H^1 cost between Γ and the resulting shape. A possible solution to this problem would be to change the inner product so that it is related to N. However, if the norm N does not come from

an inner product, this is not possible. Note that the extension of the notion of gradient in chapter 8 could be applied to any norm N to solve this particular point. Besides, all distances are not path-based and such considerations happens to be not always possible. Such a direct minimization of a distance by a gradient descent is indeed interesting mainly when the distance admits an explicit expression, which allows a fast computation of the gradient without any path-based considerations.

Let us briefly study the case where a given distance is differentiable and such that all minimizing paths by gradient descents (5.2) for a particular inner product P are geodesics (if such a distance exists). A consequence of proposition 51 is that, for any such path:

$$\mathcal{G}_{\Gamma'}\Big(d(\Gamma(0),\Gamma') + d(\Gamma',\Gamma(1)),\ \ \Gamma_t(t)\Big)\Big|_{\Gamma'=\Gamma(t)} = 0. \tag{5.4}$$

If $\mathbf{a}$ stands for $\nabla_{\Gamma'} d(\Gamma(0),\Gamma')|_{\Gamma'=\Gamma(t)}$ and $\mathbf{b}$ for $\nabla_{\Gamma'} d(\Gamma',\Gamma(1))|_{\Gamma'=\Gamma(t)}$, which is equal to $\Gamma_t(t)$, then it writes: $\langle a+b\,|-b\rangle_{P(\Gamma(t))} = 0$. Moreover, a new geodesic between $\Gamma(0)$ and $\Gamma(1)$ can be defined by following the path Γ between from $\Gamma(1)$ to $\Gamma(t)$ and then by following another gradient descent from $\Gamma(t)$ to the target $\Gamma(0)$, which allows to state also that $\langle a+b\,|-a\rangle_{P(\Gamma(t))} = 0$. A sum of these equalities leads to $\mathbf{a} = -\mathbf{b}$, which means that, at any time t, the directions from $\Gamma(t)$ to the extremities $\Gamma(0)$ and $\Gamma(1)$ are exactly opposite. Consequently, given any two shapes, the path obtained by gradient descent from one shape to the other one is the same as the gradient descent path starting from the second shape and going to the first shape. Of course the hypothesis that the distance is differentiable everywhere is very important in this result. In the case of a gradient descent based on the approximation of the Hausdorff distance, this is not true at all. However, when the approximation tends to the Hausdorff distance, the property 5.4 becomes true for both directions. The explanation is that the Hausdorff distance admits directional derivatives (so 5.4 is satisfied), but the function that associates to any direction the corresponding directional derivative is not even linear, so that the gradient can not be defined and the conclusion $\mathbf{a} = -\mathbf{b}$ does not hold anymore.

5.1.5 From the theoretical framework to the practice

In the particular case that is studied here, namely a L^2 gradient descent for a family of differentiable approximations of the Hausdorff distance, things might be not so bad. I believe that the resulting minimizing path tends towards a geodesic for the Hausdorff distance when the approximation becomes nearer and nearer from the distance, if the constraint that the path should belong to $\mathcal{S}$ is relaxed. This is however not an easy thing to prove. In fact I believe that in most cases the evolution path obtained by gradient descent of any of these approximations of the Hausdorff distance is very close to a geodesic for the Hausdorff distance. A suggestion for proof would be to check how far the gradient of the approximation is from being a L^∞ gradient in the sense of equation 8.2 in chapter 8. Note that there are continuous infinities of geodesics between any two shapes for the Hausdorff distance (which helps a lot).

In practice, the constraint that any shape should belong to $\mathcal{S}$ has not been implemented. Consequently a gradient descent might go through shapes not in $\mathcal{S}$, which happens necessarily if boundaries merge to vanish or if a topological change occurs. At those particular moments, the equivalence between the three topologies studied in chapter 1 is lost, but the gradient descent framework still stands because the approximation of the Hausdorff distance is smooth enough and its expression does not change during topological changes, which is

of great practical interest. In fact, the implementation of the constraint that shapes should remain in $\mathcal{S}$ is not really desirable since it would act only when the gradient descent hits a shape on the boundary of $\mathcal{S}$ and would not modify the path before that. It might even happen that evolution gets thus stuck into a local minimum. If the constraint is really important then the distance to minimize should be changed so that a gradient descent would (more) naturally remain in $\mathcal{S}$, because the Hausdorff distance does not naturally deal with such considerations.

A few reasonable conjectures about the larger set of shapes without the Federer and smoothness constraints are that any L^2 geodesic is a L^p geodesic for any finite value of p and reciprocally; that any L^2 geodesic is also a L^∞ geodesic (but not reciprocally), that $d_H = d^G_{L^\infty}$, and that the evolution path obtained by the gradient descent of the approximation of the Hausdorff distance converges towards a L^∞ geodesic when the approximation converges towards the Hausdorff distance, that is to say when the parameters of the approximation tend to the infinite. On the opposite, it seems that the L^2 gradient descent evolution path is far from being a L^2 geodesic. By the way, Michor and Mumford showed in [80] that $d^G_{L^2} = 0$ between any two shapes. The maximum curvature hypothesis in $\mathcal{S}$ prevents $\mathcal{S}$ from this pathological behavior. However another consequence of the smoothness constraints in $\mathcal{S}$ is that the assumption $d_H = d^G_{L^\infty}$ is not true anymore. See for example figure 5.2.

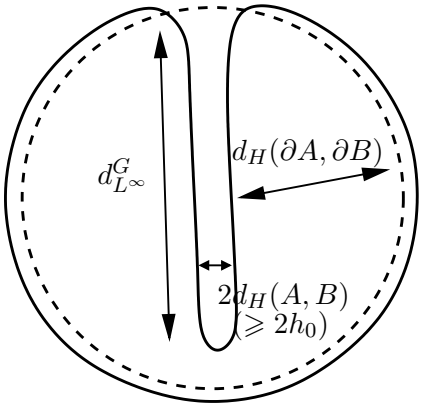


Figure 5.2: In the set of shapes $\mathcal{S}$, the L^∞ path-based geodesics do not coincide anymore with the Hausdorff distance (between full shapes or boundaries) because of the constraint $reach(\Omega) \geqslant h_0$.

5.1.6 Computing the gradient of the approximation to the Hausdorff distance

We now proceed with showing that the approximation $\tilde{\rho}_H(\Gamma, \Gamma_0)$ of the Hausdorff distance $\rho_H(\Gamma, \Gamma_0)$ is differentiable with respect to Γ and compute its gradient $\nabla \tilde{\rho}_H(\Gamma, \Gamma_0)$, in the sense of section 5.1. To simplify notations we rewrite (2.8) as

$$\tilde{\rho}_H(\Gamma, \Gamma_0) = \left\langle \left\langle \langle d(\cdot,\cdot)\rangle^{\varphi}_{\Gamma_0} \right\rangle^{\psi}_{\Gamma}, \left\langle \langle d(\cdot,\cdot)\rangle^{\varphi}_{\Gamma} \rangle^{\psi}_{\Gamma_0} \right\rangle^{\theta}, \tag{5.5}$$

and state the result, the reader interested in the proof being referred to the appendix 5.3.1.

Proposition 52. *The gradient of $\tilde{\rho}_H(\Gamma, \Gamma_0)$ at any point y of Γ is given by*

$$\nabla\tilde{\rho}_H(\Gamma, \Gamma_0)(y) = \frac{1}{\theta'(\tilde{\rho}_H(\Gamma, \Gamma_0))}\left(\alpha(y)\kappa(y) + \beta(y)\right), \tag{5.6}$$

where $\kappa(y)$ is the curvature of Γ at y, the functions $\alpha(y)$ and $\beta(y)$ are given by

$$\begin{aligned}\alpha(y) = \nu\int_{\Gamma_0}\frac{\psi'}{\varphi'}\left(\langle d(x,\cdot)\rangle_\Gamma^\varphi\right)\ \left[\ \varphi\circ\langle d(x,\cdot)\rangle_\Gamma^\varphi - \varphi\circ d(x,y)\ \right] d\Gamma_0(x)\\ + |\Gamma_0|\eta\ \left[\psi\left(\left\langle\langle d(\cdot,\cdot)\rangle_{\Gamma_0}^\varphi\right\rangle_\Gamma^\psi\right) - \psi\left(\langle d(\cdot,y)\rangle_{\Gamma_0}^\varphi\right)\ \right],\end{aligned} \tag{5.7}$$

$$\beta(y) = \int_{\Gamma_0}\varphi'\circ d(x,y)\ \left[\nu\frac{\psi'}{\varphi'}\left(\langle d(x,\cdot)\rangle_\Gamma^\varphi\right) + \eta\frac{\psi'}{\varphi'}\left(\langle d(\cdot,y)\rangle_{\Gamma_0}^\varphi\right)\right]\ \frac{y-x}{d(x,y)}\cdot n(y)\ d\Gamma_0(x), \tag{5.8}$$

where $\nu = \dfrac{1}{|\Gamma|\ |\Gamma_0|}\dfrac{\theta'}{\psi'}\left(\left\langle\langle d(\cdot,\cdot)\rangle_\Gamma^\varphi\right\rangle_{\Gamma_0}^\psi\right)$ *and* $\eta = \dfrac{1}{|\Gamma|\ |\Gamma_0|}\dfrac{\theta'}{\psi'}\left(\left\langle\langle d(\cdot,\cdot)\rangle_{\Gamma_0}^\varphi\right\rangle_\Gamma^\psi\right)$.

Note that the function $\beta(y)$ is well-defined even if y belongs to Γ_0 since the term $\frac{y-x}{d(x,y)}$ is of unit norm.

The first two terms of the gradient show explicitly that minimizing the energy implies homogenizing the distance to Γ_0 along the curve Γ, that is to say the algorithm will take care in priority of the points of Γ which are the furthest from Γ_0.

Also note that the expression of the gradient in proposition 52 still stands when Γ and Γ_0 are two surfaces (embedded in $\mathbb{R}^3$), if κ stands for the mean curvature.

5.1.7 Computation of the gradient of the approximation to the $W^{1,2}$ norm

The gradient $\nabla\tilde{\rho}_D(\Gamma, \Gamma_0)$, of our approximation $\tilde{\rho}_D(\Gamma, \Gamma_0)$ of the distance $\rho_D(\Gamma, \Gamma_0)$ given by (2.14) in the sense of section 5.1 can be computed. The interested reader is referred to appendix 5.3.2. We simply state the result in the

Proposition 53. *The gradient of $\tilde{\rho}_D(\Gamma, \Gamma_0)$ at any point y of Γ is given by*

$$\begin{aligned}\nabla\tilde{\rho}_D(\Gamma, \Gamma_0)(y) =\\ \int_D\left[B(x,y)\left(C_1(x) - \frac{\varphi^{"}}{\varphi'}(\tilde{d}_\Gamma(x))\left(\mathbf{C}_2(x)\cdot\nabla\tilde{d}_\Gamma(x)\right)\right) + \mathbf{C}_2(x)\cdot\nabla B(x,y)\right]\ dx,\end{aligned} \tag{5.9}$$

where

$$B(x,y) = \kappa(y)\left(\langle\varphi\circ d(x,\cdot)\rangle_\Gamma - \varphi\circ d(x,y)\right) + \varphi'(d(x,y))\frac{y-x}{d(x,y)}\cdot\mathbf{n}(y),$$

$\kappa(y)$ is the curvature of Γ at y,

$$C_1(x) = \frac{1}{|\Gamma|\ \varphi'(\tilde{d}_\Gamma(x))}\|\tilde{d}_\Gamma - \tilde{d}_{\Gamma_0}\|_{L^2(D)}^{-1}\ \ \left(\tilde{d}_\Gamma(x) - \tilde{d}_{\Gamma_0})(x)\right),$$

and

$$\mathbf{C}_2(x) = \frac{1}{|\Gamma|\ \varphi'(\tilde{d}_\Gamma(x))}\|\nabla(\tilde{d}_\Gamma - \tilde{d}_{\Gamma_0})\|_{\mathbf{L}^2(D)}^{-1}\ \ \nabla(\tilde{d}_\Gamma - \tilde{d}_{\Gamma_0})(x),$$

5.1.8 Direct minimization of the $W^{1,2}$ norm

An alternative to the method presented in the previous section is to evolve not the curve Γ but its distance function d_Γ. Minimizing $\rho_D(\Gamma, \Gamma_0)$ with respect to d_Γ implies computing the corresponding Euler-Lagrange equation EL. The reader will verify that the result is

$$EL = \frac{d_\Gamma - d_{\Gamma_0}}{\|d_\Gamma - d_{\Gamma_0}\|_{L^2(D)}} - div\left(\frac{\nabla\left(d_\Gamma - d_{\Gamma_0}\right)}{\|\nabla(d_\Gamma - d_{\Gamma_0})\|_{\mathbf{L}^2(D)})}\right) \tag{5.10}$$

To simplify notations we now use d instead of d_Γ. The problem of warping Γ_1 onto Γ_0 is then transformed into the problem of solving the following PDE

$$\begin{aligned} d_t &= -EL \\ d(0,\cdot) &= d_{\Gamma_1}(\cdot). \end{aligned}$$

The problem that this PDE does not preserve the fact that d is a distance function is alleviated by "reprojecting" at each iteration the current function d onto the set of distance functions by running a few iterations of the "standard" restoration PDE [108]

$$\begin{aligned} d_t &= (1 - |\nabla d|)sign(d) \\ d(0,\cdot) &= d_0 \end{aligned}$$

5.2 Application to curve evolutions: Hausdorff warping

In this section we show a number of examples of solving equation (5.2) with the gradient given by equation (5.6).

5.2.1 Applying the theory

In practice, the Energy that we minimize is not $\tilde{\rho}_H$ but in fact a "regularized" version obtained by combining $\tilde{\rho}_H$ with a term E_L which depends upon the lengths of the two curves. A natural candidate for E_L is $\max(|\Gamma|, |\Gamma'|)$ since it acts only if $|\Gamma|$ becomes larger than $|\Gamma'|$, thereby avoiding undesirable oscillations. To obtain smoothness, we approximate the max with a Ψ-average:

$$E_L(|\Gamma|, |\Gamma'|) = \left\langle |\Gamma|, |\Gamma'| \right\rangle^{\Psi} \tag{5.11}$$

We know that the function $\Gamma \to |\Gamma|$ is in general l.s.c.. It is in fact continuous on $\mathcal{S}$ (see the proof of proposition 23) and takes its values in the interval $[0, 2c_0]$, hence

Proposition 54. *The function $\mathcal{S} \to \mathbb{R}^+$ given by $\Gamma \to E_L(\Gamma, \Gamma')$ is continuous for the Hausdorff topology.*

Proof. It is clear since E_L is a combination of continuous functions. □

We combine E_L with $\tilde{\rho}_H$ the expected way, i.e. by computing their $\tilde{\Psi}$ average so that the final energy is

$$E(\Gamma, \Gamma') = \left\langle \tilde{\rho}_H(\Gamma, \Gamma'), E_L(|\Gamma|, |\Gamma'|) \right\rangle^{\tilde{\Psi}} \tag{5.12}$$

The function $E : \mathcal{S} \times \mathcal{S} \to \mathbb{R}^+$ is continuous for the Hausdorff metric because of propositions 23 and 54 and therefore

Proposition 55. *The function* $\Gamma \to E(\Gamma, \Gamma')$ *defined on the set of shapes* $\mathcal{S}$ *has at least a minimum in the closure* $\overline{\mathcal{S}}$ *of* $\mathcal{S}$ *in* $\mathcal{L}_0$.

Proof. This is a direct application of proposition 22 applied to the function E. □

We call the resulting warping technique the *Hausdorff* warping. A first example, the Hausdorff warping of two circles, is shown in figure 5.3. A second example, the Hausdorff warping of two hand silhouettes, is shown in figure 5.4

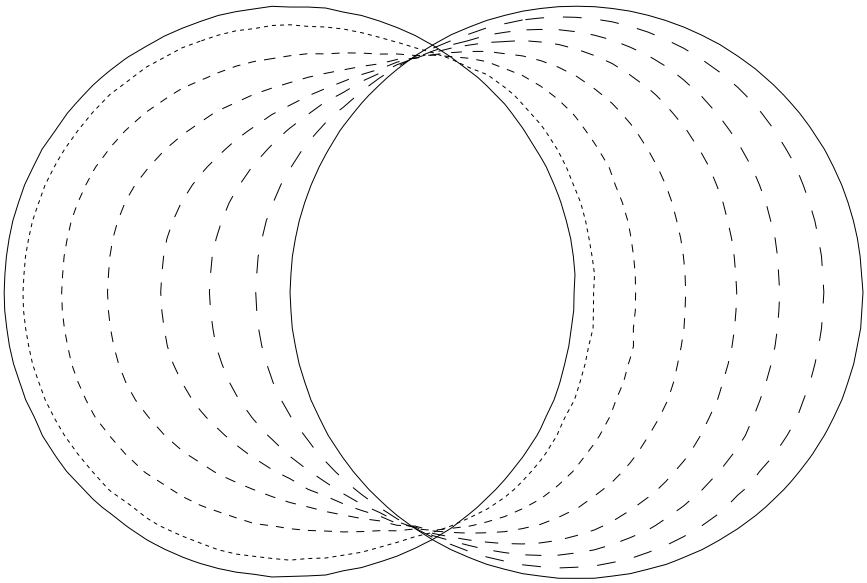


Figure 5.3: The result of the Hausdorff warping of two circles. The two circles are represented in continuous line while the intermediate shapes are represented in dotted lines.

We have borrowed the example in figure 5.5 from the database (`www.ee.surrey.ac.uk/Research/VSSP/imagedb/demo.html`) of fish silhouettes collected by the researchers of the University of Surrey at the center for Vision, Speech and Signal Processing (`www.ee.surrey.ac.uk/Research/VSSP`). This database contains 1100 silhouettes. A few steps of the result of the Hausdorff warping of one of these silhouettes onto another are shown in figure 5.5. Another similar example is shown in figure 5.6. Note that, prior to warping, the two shapes have been normalized in such a way as to align their centers of gravity and their principal axes.

Figures 5.7 and 5.8 give a better understanding of the behavior of Hausdorff warping. A slightly displaced detail "warps back" to its original place (figure 5.7). Displaced further, the same detail is considered as another one and disappears during the warping process while the original one reappears (figure 5.8).

Finally, figures 5.9 and 5.11 show the Hausdorff warping between two open curves and between two closed surfaces, respectively. Figure 5.11 and the whole 3D framework has been computed by Pierre Maurel.

Note also that other warpings are given by the minimization of other approximations of the Hausdorff distance. Figure 5.10 shows the warping of a rough curve to the silhouette of a fish and bubbles given by the minimization of the $W^{1,2}$ norm as explained in section

Figure 5.4: The result of the Hausdorff warping of two hand silhouettes. The two hands are represented in continuous line while the intermediate shapes are represented in dotted lines.

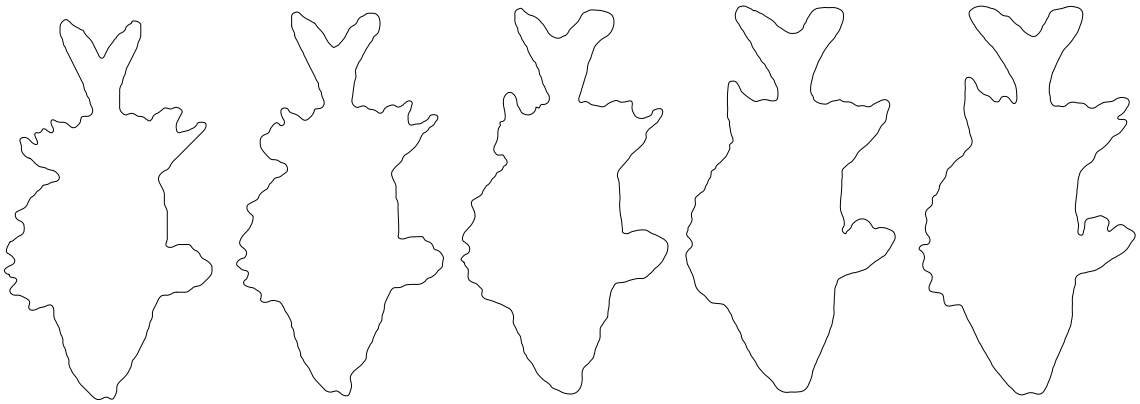


Figure 5.5: Hausdorff warping of a fish onto another.

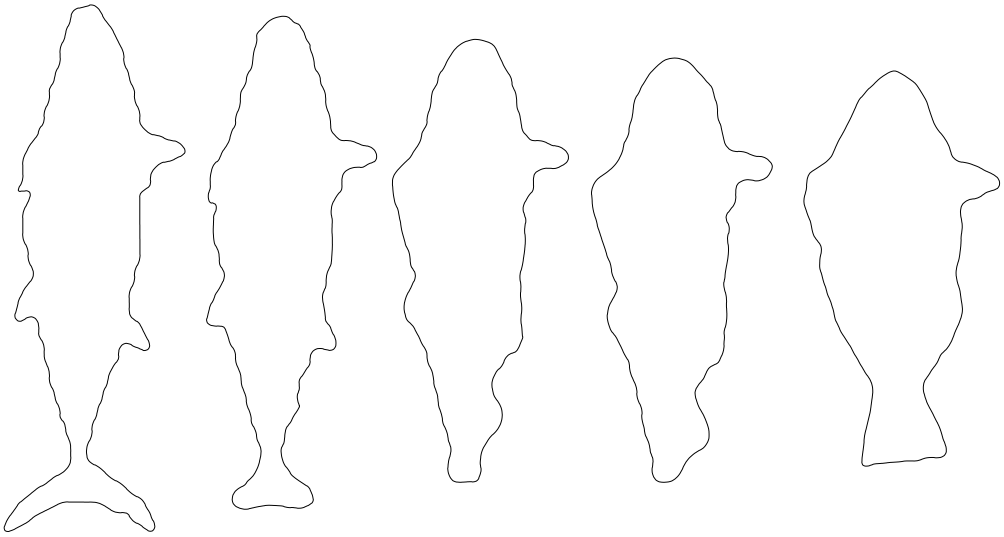


Figure 5.6: Another example of fish Hausdorff warping.

5.1.8. Our "level sets" implementation (see section 5.2.2) can deal with the splitting of the source curve while warping onto the target one.

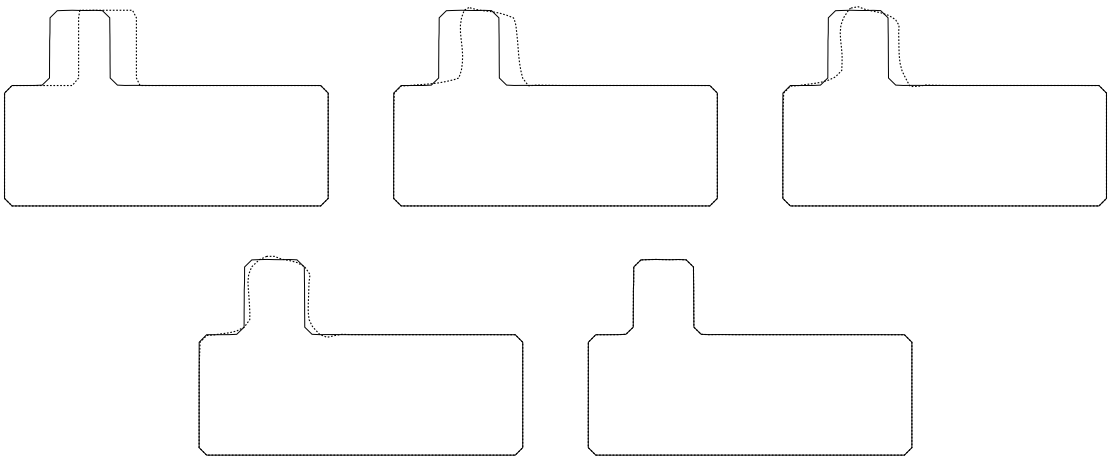


Figure 5.7: Hausdorff warping boxes (i). A translation-like behavior.

Heuristically a Hausdorff warping never gets stuck into a local minimum and always converges towards its target. However it happens that the evolving path is sometimes not satisfying, especially when warping a straight line onto large oscillations (see figure 5.12). Happily, for relatively close shapes, this phenomenon does not appear. The question of setting priors on the paths to avoid this kind of evolution is the subject of several following chapters.

5.2.2 Some remarks about our implementation

There are several choices for the definition of $\tilde{\rho}_H$ and the implementation of the motion of a curve Γ under a velocity field v: $\Gamma_t = v$. When Γ is composed of one or more closed

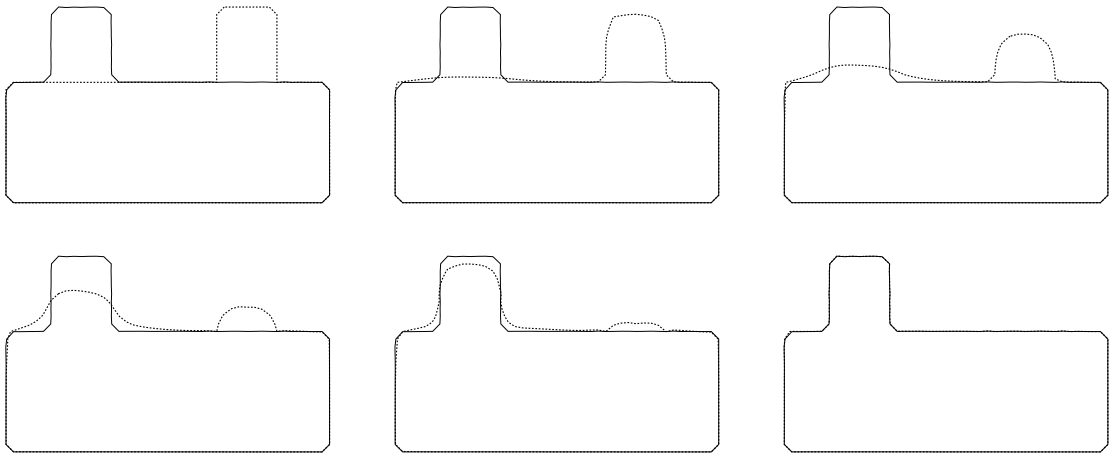


Figure 5.8: Hausdorff warping boxes (ii). A different behavior: a detail disappears while another one appears.

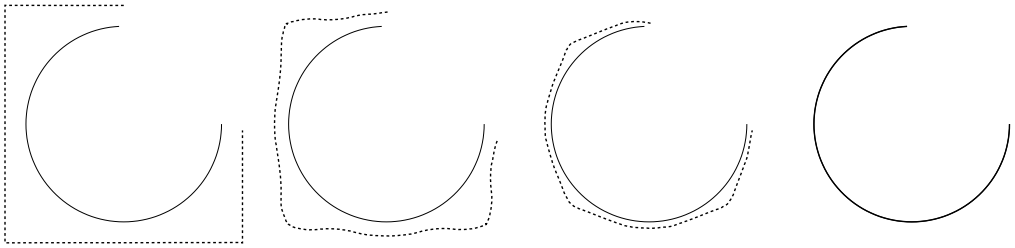


Figure 5.9: Hausdorff warping an open curve to another one.

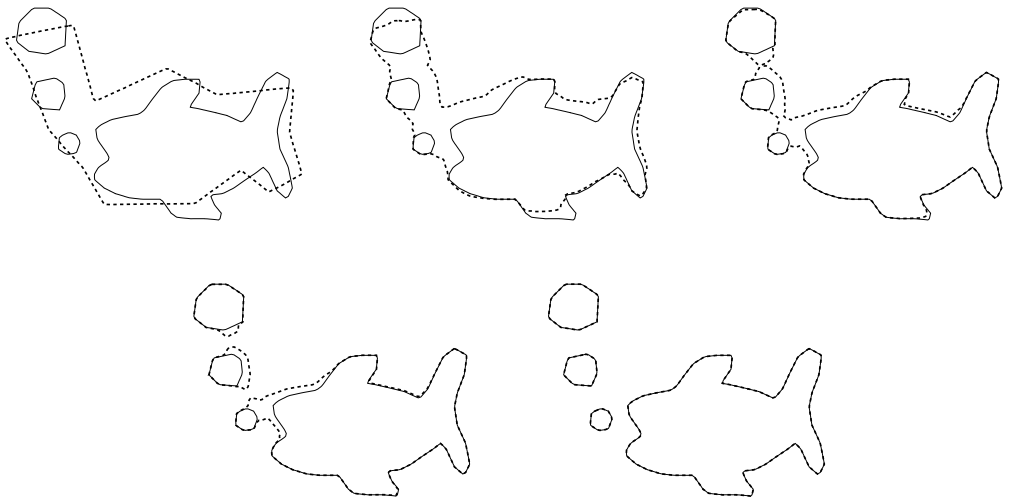


Figure 5.10: Splitting while warping.

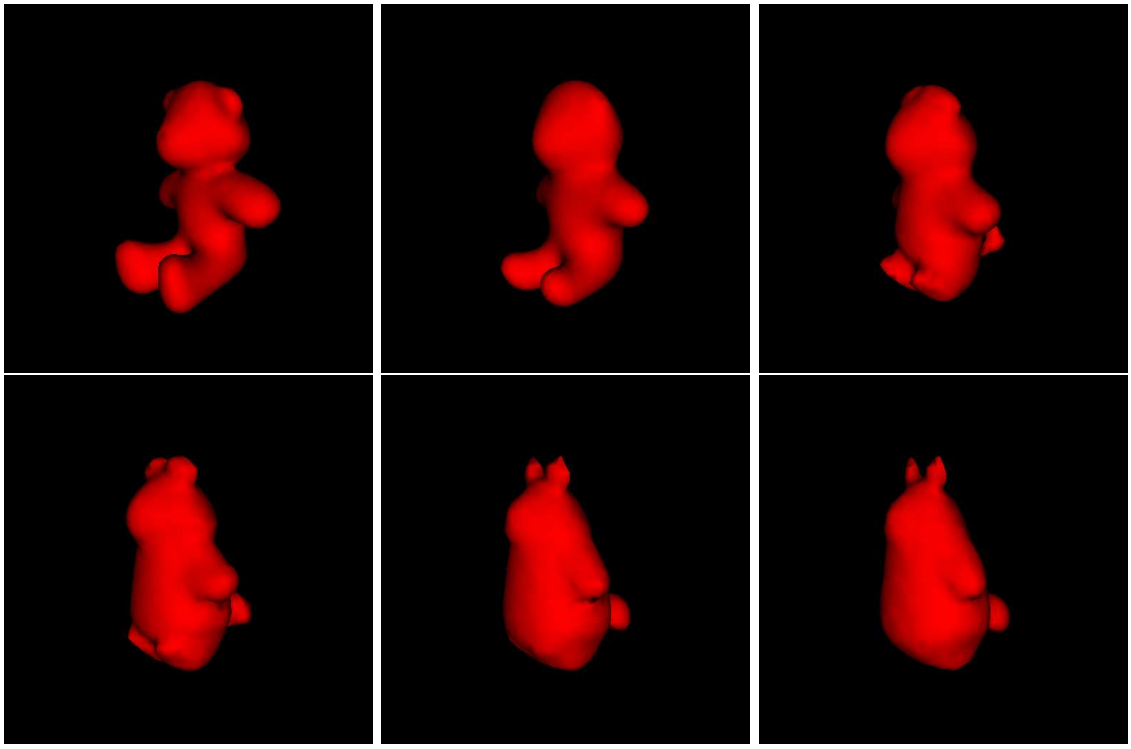


Figure 5.11: Hausdorff warping a closed surface to another one.

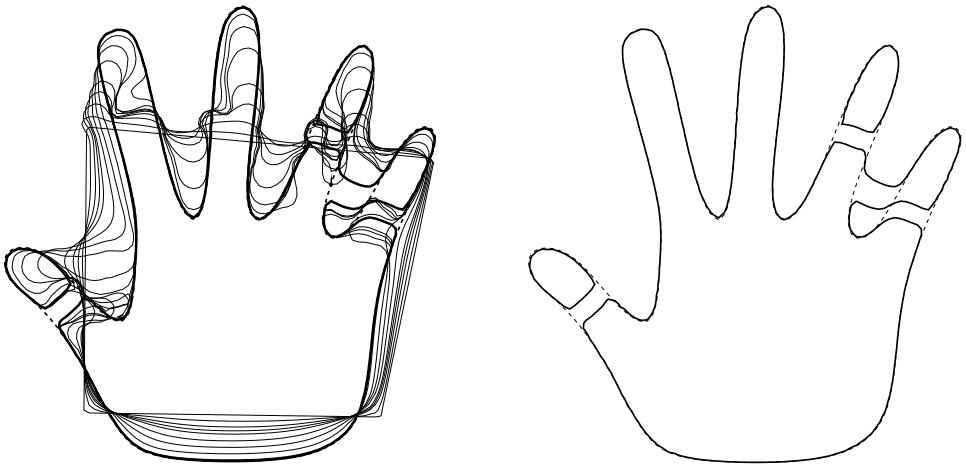


Figure 5.12: Example of an evolution, from a rectangle to a hand, which converges to the target but whose path is not very satisfying. A snapshot of the evolving shape just before convergence is shown on the right part.

connected components, we use the level set method introduced by Osher an Sethian in 1987 [86, 103, 85].

The approximation of the Hausdorff distance

We need to decide on the parameters appearing in the definition (2.8) of $\tilde{\rho}_H$, i.e. the values of α, β and γ, and on the function φ. γ controls how well we approximate the max operator that occurs in the definition of the Hausdorff distance (1.4). In our implementation we have used the value of 2. The parameters α and β control the accuracy with which we approximate the sup and inf operators that appear in equation (1.4). The higher they are are, the better the approximation, but the better the approximation, the higher the numerical difficulties due to the "stiffness" of the function which is of course due the non-differentiability of the Hausdorff distance. As explained in section 2.2.1, the values of α and β, together with the function φ, can be ultimately related to the coarseness with which one analyzes the values of the distance function of one curve at points on the other curve, see figures 2.3 and 2.4. In our implementation we use the values $\alpha = \beta = 4$ which we found to be a good compromise between the amount of smoothness of $\tilde{\rho}_H$ and the quality of the approximation of ρ_H. The function φ is equal to φ_1 given by equation (2.7).

Closed curves: the level set method

The key idea of the level set method is to represent the curve $\Gamma(t)$ implicitly, i.e. as the zero level of some function $u(x,t)$ defined for $x \in D$. Usually, u if negative inside Γ and positive outside. It can be easily proved that, if u evolves according to

$$\frac{\partial u(x,y)}{\partial t} + v\nabla u = 0$$

then, its zero level $\{x|\Gamma(x,t) = 0\}$ evolves according to the required equation $\Gamma_t = v$. Here, v is the desired velocity on Γ and is arbitrary elsewhere (see below).

Often, only the normal velocity field $\beta\mathbf{n}$ is important. As $\mathbf{n} = \nabla u/|\nabla u|$, the evolution of u becomes:

$$\frac{\partial u(x,y)}{\partial t} + \beta|\nabla u| = 0$$

The advantages of the level set method are well known: stability, accuracy, convergence to the correct solution, easy extension to higher dimensions, correct handling of topological changes such as breaking and merging.

An important issue is that β is only defined on curve Γ in the partial differential equation $\Gamma_t = \beta\mathbf{n}$. In many cases, β has a natural extension everywhere in domain D, so that equation $u_t + \beta|\nabla u| = 0$ is defined. For instance, when $\beta(x)$ is the curvature of Γ at point $x \in \Gamma$, one could choose, at each point $x \in D$, $\beta(x)$ equal to the curvature of the level set of u going through x. In some other cases, like ours, β can only be computed on Γ and some extension procedure has to be used to get β everywhere in D. This is now classical [95, 1, 52].

It should be noted that the zero level set of function $u(.,t)$ is only extracted to visualize $\Gamma(t)$, usually with the Marching Cubes algorithm [75] which interpolates its position and gives a nice polygonal approximation of it (a triangulated mesh in 3D). For obvious speed and accuracy reasons, it is important not to rely on this approximation to compute the velocity β. Useful quantities can generally be computed directly from u. So it is for the

normal and the curvature:

$$\mathbf{n} = \nabla u/|\nabla u|$$
$$\kappa = -\nabla \cdot (\nabla u/|\nabla u|)$$

and for the integral of some quantity $f(x)$ over Γ:

$$\int_\Gamma f(x)d\Gamma(x) = \int_D f(x)\delta(u(x))|\nabla u|dx$$

where $\delta(.)$ is a one dimensional Dirac function. Our β also involves the distance function to the curve Γ which has to be known without extracting the zero level set of function u. Usually, one takes the signed distance to the initial curve $\Gamma(0)$ as an initial value of u. Thus, $u(.,0)$ can be used to compute β at time $t = 0$. Yet, $u(.,t)$ has no reason to remain the distance to $\Gamma(t)$... except in some implementations where β is extended in such a way that the distance function is preserved during the evolution of u (see [52]): this is exactly what we need. And what we use!

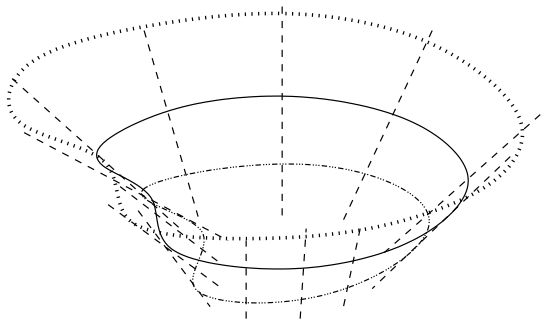


Figure 5.13: Graphic representation of the signed distance to a curve

As a conclusion, in the case of closed curves, the Hausdorff warping (section 5.2) and the shape statistics (chapter 9) are implemented with a Level Set Method with: (i) velocity extension, (ii) distance function preserving, and (iii) no need to extract the zero level set (except for visualization).

The minimization of the $W^{1,2}$ norm (section 5.1.8 and figure 5.10) is also implemented with the Level Set Method. As already mentioned, the reprojection on the set of distance functions is a "standard" Level Set technique [108].

Open curves

For open curves (figure 5.9), the level set method cannot be used. A straight Lagrangian approach and polygonal approximations of the curves were used as a first step toward more refined methods like the ones described in [8].

5.2.3 Comparison with other approaches

It is interesting to recall the fact that our approach can be seen as the opposite of that consisting in first building a Riemannian structure on the set of shapes, i.e. going from an infinitesimal metric structure to a global one. The infinitesimal structure is defined by an inner product in the tangent space (the set of normal deformation fields) and has to vary

continuously from point to point, i.e. from shape to shape. As mentioned before, this is mostly dealt with in the work of Miller, Trouvé and Younes [81, 112, 120]. The problem with these approaches, beside that of having to deal with parameterizations of the shapes, is that there exist global metric structures on the set of shapes (see section 1.2.2) which are useful and relevant to the problem of the comparison of shapes but that do not arise from an infinitesimal structure.

Our approach can be seen as taking the problem from exactly the opposite viewpoint from the previous one: we start with a global metric on the set of shapes (ρ_H or the $W^{1,2}$ metric) and build smooth functions (in effect smooth approximations of these metrics) that we use as dissimilarity measures or energy functions and minimize using techniques of the calculus of variation by computing their gradient and performing infinitesimal gradient descent. We have seen that in order to compute the gradients we need to define an inner-product of normal deformation flows and the choice of this inner-product may influence the way our algorithms evolve from one shape to another. This last point is related to, but different from, the choice of the Riemannian metric in the first approach. Its investigation is the topic of a next chapter.

5.3 Computations

5.3.1 Computation of $\nabla \tilde{\rho}_H(\Gamma, \Gamma_0)$

We prove proposition 52.

Proof. We make a few definitions to simplify notations:

$$m^{\varphi,\psi}_{\Gamma_0,\Gamma} = \left\langle \langle d(\cdot,\cdot)\rangle^{\varphi}_{\Gamma_0} \right\rangle^{\psi}_{\Gamma}$$
$$m^{\varphi,\psi}_{\Gamma,\Gamma_0} = \left\langle \langle d(\cdot,\cdot)\rangle^{\varphi}_{\Gamma} \right\rangle^{\psi}_{\Gamma_0}$$

We also define the corresponding functions

$$m^{\varphi}_{\Gamma}(x) = \langle d(x,\cdot)\rangle^{\varphi}_{\Gamma}$$
$$m^{\varphi}_{\Gamma_0}(x) = \langle d(x,\cdot)\rangle^{\varphi}_{\Gamma_0}.$$

We then proceed with

$$\mathcal{G}(\tilde{\rho}_H(\Gamma, \Gamma_0), \boldsymbol{\beta}) = \frac{1}{2\theta'(\tilde{\rho}_H(\Gamma, \Gamma_0))}\left[\theta'\left(m^{\varphi,\psi}_{\Gamma_0,\Gamma}\right)\ \mathcal{G}\left(m^{\varphi,\psi}_{\Gamma_0,\Gamma}, \boldsymbol{\beta}\right) + \theta'\left(m^{\varphi,\psi}_{\Gamma,\Gamma_0}\right)\ \mathcal{G}\left(m^{\varphi,\psi}_{\Gamma,\Gamma_0}, \boldsymbol{\beta}\right)\right],$$

because of (2.3).

Computation of the first term $\mathcal{G}\left(m^{\varphi,\psi}_{\Gamma_0,\Gamma}, \boldsymbol{\beta}\right)$

We apply the chain rule and (2.2) to obtain

$$\mathcal{G}\left(m^{\varphi,\psi}_{\Gamma_0,\Gamma}, \boldsymbol{\beta}\right) = \frac{1}{\psi'\left(m^{\varphi,\psi}_{\Gamma_0,\Gamma}\right)}\left[\frac{1}{|\Gamma|}\ \mathcal{G}\left(\int_\Gamma \psi\left(\langle d(\cdot,\cdot)\rangle^{\varphi}_{\Gamma_0}\right), \boldsymbol{\beta}\right) + \left(\int_\Gamma \psi\left(\langle d(\cdot,\cdot)\rangle^{\varphi}_{\Gamma_0}\right)\right)\ \mathcal{G}\left(\frac{1}{|\Gamma|}, \boldsymbol{\beta}\right)\right].$$

We now compute $\mathcal{G}\left(\frac{1}{|\Gamma|}, \boldsymbol{\beta}\right)$:

$$\begin{aligned}
\mathcal{G}\left(\frac{1}{|\Gamma|}, \boldsymbol{\beta}\right) &= -\frac{1}{|\Gamma|^2}\ \mathcal{G}\left(|\Gamma|, \boldsymbol{\beta}\right) \\
&= -\frac{1}{|\Gamma|^2}\mathcal{G}\left(\int_0^1 |\Gamma'(p)|dp, \boldsymbol{\beta}\right) \\
&= -\frac{1}{|\Gamma|^2}\int_0^1 \frac{\Gamma'(p)}{|\Gamma'(p)|}\cdot\boldsymbol{\beta}'(p)\ dp \\
&= \frac{1}{|\Gamma|^2}\int_0^1 \kappa(p)\mathbf{n}(p)\cdot\boldsymbol{\beta}(p)\ \ |\Gamma'(p)|dp.
\end{aligned}$$

The last line is obtained by integrating by parts and using the hypothesis that $\boldsymbol{\beta}$ is parallel to $\mathbf{n}$. $\kappa(p)$ is the curvature of Γ at the point $\Gamma(p)$. This yields

$$\left(\int_\Gamma \psi\left(\langle d(\cdot,\cdot)\rangle^{\varphi}_{\Gamma_0}\right)\right)\ \mathcal{G}\left(\frac{1}{|\Gamma|}, \boldsymbol{\beta}\right) = \frac{1}{|\Gamma|}\psi(m^{\varphi,\psi}_{\Gamma_0,\Gamma})\int_\Gamma \kappa(y)\mathbf{n}(y)\cdot\boldsymbol{\beta}(y)\ d\Gamma(y) \tag{5.13}$$

We continue with

$$\begin{aligned}
&\mathcal{G}\left(\int_\Gamma \psi\left(\langle d(\cdot,\cdot)\rangle^{\varphi}_{\Gamma_0}\right), \boldsymbol{\beta}\right) = \mathcal{G}\left(\int_0^1 \psi\left(\langle d(\Gamma(p),\cdot)\rangle^{\varphi}_{\Gamma_0}\right)\ |\Gamma'(p)|dp, \boldsymbol{\beta}\right) = \\
&\int_0^1 \frac{\psi'}{\varphi'}\left(\langle d(\Gamma(p),\cdot)\rangle^{\varphi}_{\Gamma_0}\right)\ \frac{1}{|\Gamma_0|}\ \lim_{\tau\to 0}\frac{\int_{\Gamma_0}\left(\varphi(d(\Gamma(p)+\tau\boldsymbol{\beta}(p),\cdot)-\varphi(d(\Gamma(p),\cdot))\right)}{\tau}\ |\Gamma'(p)|dp + \\
&\qquad\qquad \int_0^1 \psi\left(\langle d(\Gamma(p),\cdot)\rangle^{\varphi}_{\Gamma_0}\right)\ \lim_{\tau\to 0}\frac{|\Gamma'(p)+\tau\boldsymbol{\beta}'(p)|-|\Gamma'(p)|}{\tau}dp.
\end{aligned}$$

The last term is equal to (using the hypothesis that $\boldsymbol{\beta}(p)$ is parallel to $\mathbf{n}(p)$ for all p's):

$$-\int_0^1 \psi\left(\langle d(\Gamma(p),\cdot)\rangle^{\varphi}_{\Gamma_0}\right)\ \kappa(p)\mathbf{n}(p)\cdot\boldsymbol{\beta}(p)\ |\Gamma'(p)|dp.$$

The first term can be written:

$$\int_0^1 \frac{\psi'}{\varphi'}\left(\langle d(\Gamma(p),\cdot)\rangle^{\varphi}_{\Gamma_0}\right)\frac{1}{|\Gamma_0|}\left(\int_{\Gamma_0}\varphi'(d(\Gamma(p),x))\ \ \frac{\Gamma(p)-x}{d(\Gamma(p),x)}\cdot\boldsymbol{\beta}(p)\ d\Gamma_0(x)\right)\ |\Gamma'(p)|dp$$

Combining them we obtain

$$\int_\Gamma\left(\frac{\psi'}{\varphi'}\left(m^{\varphi}_{\Gamma_0}(y)\right)\langle\varphi'(d(y,\cdot))\ \ \frac{y-\cdot}{d(y,\cdot)}\rangle_{\Gamma_0} - \psi\left(m^{\varphi}_{\Gamma_0}(y)\right)\kappa(y)\mathbf{n}(y)\right)\cdot\boldsymbol{\beta}(y)\ d\Gamma(y) \tag{5.14}$$

We finally combine (5.13) and (5.14)

$$\begin{aligned}
\mathcal{G}\left(m^{\varphi,\psi}_{\Gamma_0,\Gamma}, \boldsymbol{\beta}\right) = \frac{1}{\psi'\left(m^{\varphi,\psi}_{\Gamma_0,\Gamma}\right)|\Gamma|}\int_\Gamma\left(\frac{\psi'}{\varphi'}\left(m^{\varphi}_{\Gamma_0}(y)\right)\left\langle\varphi'(d(y,\cdot))\ \ \frac{y-\cdot}{d(y,\cdot)}\right\rangle_{\Gamma_0} + \right. \\
\left.\left(\psi\left(m^{\varphi,\psi}_{\Gamma_0,\Gamma}\right) - \psi\left(m^{\varphi}_{\Gamma_0}(y)\right)\right)\kappa(y)\mathbf{n}(y)\right)\cdot\boldsymbol{\beta}(y)\ d\Gamma(y)
\end{aligned} \tag{5.15}$$

Computation of the second term $\mathcal{G}\left(m^{\varphi,\psi}_{\Gamma,\Gamma_0},\boldsymbol{\beta}\right)$

Because of (2.2) we can write

$$\mathcal{G}\left(m^{\varphi,\psi}_{\Gamma,\Gamma_0},\boldsymbol{\beta}\right) = \frac{1}{\psi'\left(m^{\varphi,\psi}_{\Gamma,\Gamma_0}\right)}\frac{1}{|\Gamma_0|} \times \int_{\Gamma_0}\frac{\psi'}{\varphi'}\left(\langle d(x,\cdot)\rangle^{\varphi}_{\Gamma}\right)\ \mathcal{G}\left(\frac{1}{|\Gamma|}\int_0^1 \varphi\left(d(x,\Gamma(p))\right)\ |\Gamma'(p)|dp,\boldsymbol{\beta}\right)\ d\Gamma_0(x).$$

Using the chain rule

$$\mathcal{G}\left(\frac{1}{|\Gamma|}\int_0^1 \varphi\left(d(x,\Gamma(p))\right)\ |\Gamma'(p)|dp,\boldsymbol{\beta}\right) = \left(\int_\Gamma \varphi\circ d(x,\cdot)\right)\frac{1}{|\Gamma|^2}\int_0^1 \kappa(p)\mathbf{n}(p)\cdot\boldsymbol{\beta}(p)\ |\Gamma'(p)|dp+ \frac{1}{|\Gamma|}\int_0^1 \varphi'(d(\Gamma(p),x))\ \frac{\Gamma(p)-x}{d(\Gamma(p),x)}\cdot\boldsymbol{\beta}(p)\ |\Gamma'(p)|dp+ \frac{1}{|\Gamma|}\int_0^1 \varphi\left(d(x,\Gamma(p))\right)\ \frac{\Gamma'(p)}{|\Gamma'(p)|}\cdot\boldsymbol{\beta}'(p)dp\ .$$

Under the same hypothesis for $\boldsymbol{\beta}$, the last term is equal to:

$$-\frac{1}{|\Gamma|}\int_0^1 \varphi\left(d(\Gamma(p),x)\right)\ \kappa(p)\mathbf{n}(p)\cdot\boldsymbol{\beta}(p)\ |\Gamma'(p)|dp$$

The expression of $\mathcal{G}(\tilde{\rho}_H(\Gamma,\,\Gamma_0),\boldsymbol{\beta})$ is obtained by reordering these terms. This yields

$$\begin{aligned}
&\mathcal{G}(\tilde{\rho}_H(\Gamma,\,\Gamma_0),\boldsymbol{\beta}) = \\
&\frac{1}{2\theta'(\tilde{\rho}_H(\Gamma,\,\Gamma_0))}\int_\Gamma\Bigg[\Bigg[\nu\kappa(p)\int_{\Gamma_0}\frac{\psi'}{\varphi'}\left(\langle d(x,\cdot)\rangle^{\varphi}_{\Gamma}\right)\ \left[\ \varphi\circ\langle d(x,\cdot)\rangle^{\varphi}_{\Gamma}-\varphi\circ d(x,y)\ \right]d\Gamma_0(x)\\
&\qquad+|\Gamma_0|\eta\kappa(p)\ \left[\psi\left(\left\langle\langle d(\cdot,\cdot)\rangle^{\varphi}_{\Gamma_0}\right\rangle^{\psi}_{\Gamma}\right)-\psi\left(\langle d(\cdot,y)\rangle^{\varphi}_{\Gamma_0}\right)\ \right]\Bigg]\mathbf{n}(p)\\
&+\int_{\Gamma_0}\frac{\varphi'\circ d(x,y)}{d(x,y)}\ \left[\nu\frac{\psi'}{\varphi'}\left(\langle d(x,\cdot)\rangle^{\varphi}_{\Gamma}\right)+\eta\frac{\psi'}{\varphi'}\left(\langle d(\cdot,y)\rangle^{\varphi}_{\Gamma_0}\right)\right]\ (y-x)\ d\Gamma_0(x)\Bigg]\cdot\boldsymbol{\beta}(p)d\Gamma(y)
\end{aligned}$$

where $\nu = \dfrac{1}{|\Gamma|\ |\Gamma_0|}\dfrac{\theta'}{\psi'}\left(\left\langle\langle d(\cdot,\cdot)\rangle^{\varphi}_{\Gamma}\right\rangle^{\psi}_{\Gamma_0}\right)$ and $\eta = \dfrac{1}{|\Gamma|\ |\Gamma_0|}\dfrac{\theta'}{\psi'}\left(\left\langle\langle d(\cdot,\cdot)\rangle^{\varphi}_{\Gamma_0}\right\rangle^{\psi}_{\Gamma}\right)$.
The gradient $\nabla\,\tilde{\rho}_H(\Gamma,\,\Gamma_0)$ is obtained by identifying the previous expression as an inner product of normal deformation flows $\int_\Gamma \nabla\,\tilde{\rho}_H(\Gamma,\,\Gamma_0)(y)\beta(y)\,d\Gamma(y)$

$$\begin{aligned}
&\nabla\,\tilde{\rho}_H(\Gamma,\,\Gamma_0)(y) = \\
&\frac{1}{\theta'(\tilde{\rho}_H(\Gamma,\,\Gamma_0))}\Bigg[\nu\kappa(p)\int_{\Gamma_0}\frac{\psi'}{\varphi'}\left(\langle d(x,\cdot)\rangle^{\varphi}_{\Gamma}\right)\ \left[\ \varphi\circ\langle d(x,\cdot)\rangle^{\varphi}_{\Gamma}-\varphi\circ d(x,y)\ \right]d\Gamma_0(x)\\
&\qquad+|\Gamma_0|\eta\kappa(p)\ \left[\psi\left(\left\langle\langle d(\cdot,\cdot)\rangle^{\varphi}_{\Gamma_0}\right\rangle^{\psi}_{\Gamma}\right)-\psi\left(\langle d(\cdot,y)\rangle^{\varphi}_{\Gamma_0}\right)\ \right]\\
&+\int_{\Gamma_0}\frac{\varphi'\circ d(x,y)}{d(x,y)}\ \left[\nu\frac{\psi'}{\varphi'}\left(\langle d(x,\cdot)\rangle^{\varphi}_{\Gamma}\right)+\eta\frac{\psi'}{\varphi'}\left(\langle d(\cdot,y)\rangle^{\varphi}_{\Gamma_0}\right)\right]\ (y-x)\cdot\mathbf{n}(p)\ d\Gamma_0(x)\Bigg]
\end{aligned}$$

We should note that all these results hold independently of the fact that the curves are open or closed since we only used in the integration by parts the fact that the field $\boldsymbol{\beta}$ was parallel to the normal field $\mathbf{n}$. □

5.3.2 Computation of $\nabla\tilde{\rho}_D(\Gamma,\,\Gamma_0)$

We prove proposition 53

Proof. From the definitions

$$\mathcal{G}(\tilde{\rho}_D(\Gamma,\,\Gamma_0),\boldsymbol{\beta}) = \mathcal{G}(\|\tilde{d}_\Gamma - \tilde{d}_{\Gamma_0}\|_{L^2(D)},\,\boldsymbol{\beta}) + \mathcal{G}(\|\nabla(\tilde{d}_\Gamma - \tilde{d}_{\Gamma_0}\|_{\mathbf{L}^2(D)},\,\boldsymbol{\beta}),$$

and

$$\mathcal{G}(\|\tilde{d}_\Gamma - \tilde{d}_{\Gamma_0}\|_{L^2(D)},\,\boldsymbol{\beta}) = \frac{1}{\|\tilde{d}_\Gamma - \tilde{d}_{\Gamma_0}\|_{L^2(D)}}\int_D |\tilde{d}_\Gamma(x) - \tilde{d}_{\Gamma_0}(x)|\mathcal{G}(\tilde{d}_\Gamma(x),\,\boldsymbol{\beta})\;dx,$$

and

$$\mathcal{G}(\|\nabla(\tilde{d}_\Gamma - \tilde{d}_{\Gamma_0})\|_{\mathbf{L}^2(D)},\,\boldsymbol{\beta}) = \frac{1}{\|\nabla(\tilde{d}_\Gamma - \tilde{d}_{\Gamma_0})\|_{\mathbf{L}^2(D)}}\int_D \nabla(\tilde{d}_\Gamma(x) - \tilde{d}_{\Gamma_0}(x))\cdot\nabla(\mathcal{G}(\tilde{d}_\Gamma(x),\,\boldsymbol{\beta}))\;dx.$$

We now compute $\mathcal{G}(\tilde{d}_\Gamma(x),\,\boldsymbol{\beta})$ and its gradient. Starting with (2.6), we readily obtain

$$\mathcal{G}(\tilde{d}_\Gamma(x),\,\boldsymbol{\beta}) = \frac{1}{|\Gamma|\;\varphi'(\tilde{d}_\Gamma(x))}\Big[\left(\int_\Gamma \kappa(y)\mathbf{n}(y)\cdot\boldsymbol{\beta}(y)\;d\Gamma(y)\right)\langle\varphi\circ d(x,\cdot)\rangle_\Gamma + \int_\Gamma \varphi'(d(x,y))\frac{y-x}{d(x,y)}\cdot\boldsymbol{\beta}(y)\;d\Gamma(y) - \int_\Gamma \varphi(d(x,y))\kappa(y)\mathbf{n}(y)\cdot\boldsymbol{\beta}(y)\;d\Gamma(y)\Big]$$

According to our initial hypothesis, $\boldsymbol{\beta}(y) = \beta(y)\,\mathbf{n}(y)$. We define

$$B(x,y) = \kappa(y)\left(\langle\varphi\circ d(x,\cdot)\rangle_\Gamma - \varphi\circ d(x,y)\right) + \varphi'(d(x,y))\frac{y-x}{d(x,y)}\cdot\mathbf{n}(y),$$

so that

$$\mathcal{G}(\tilde{d}_\Gamma(x),\,\boldsymbol{\beta}) = \frac{1}{|\Gamma|\;\varphi'(\tilde{d}_\Gamma(x))}\int_\Gamma B(x,y)\beta(y)\;d\Gamma(y).$$

Let us compute the gradient of this expression with respect to the x variable:

$$\nabla\mathcal{G}(\tilde{d}_\Gamma(x),\,\boldsymbol{\beta}) = -\frac{\varphi"(\tilde{d}_\Gamma(x))}{|\Gamma|\;\varphi'^2(\tilde{d}_\Gamma(x))}\left(\int_\Gamma B(x,y)\beta(y)\;d\Gamma(y)\right)\nabla\tilde{d}_\Gamma(x) + \frac{1}{|\Gamma|\;\varphi'(\tilde{d}_\Gamma(x))}\int_\Gamma \nabla B(x,y)\beta(y)\;d\Gamma(y)$$

After some manipulation, we find that

$$\nabla B(x,y) =$$
$$\kappa(y)\left(\left\langle \varphi'\circ d(x,\cdot)\,\frac{x-\cdot}{d(x,\cdot)}\right\rangle_\Gamma - \varphi'\circ d(x,y)\,\frac{x-y}{d(x,y)}\right) + \left(\varphi^{"}(d(x,y)) - \frac{\varphi'(d(x,y)}{d(x,y)}\right)\mathbf{n}(y),$$

where we have used in particular the fact that

$$\nabla\tilde{d}_\Gamma(x) = \frac{1}{\varphi'(\tilde{d}_\Gamma(x))}\left\langle \varphi'\circ d(x,\cdot)\,\frac{x-\cdot}{d(x,\cdot)}\right\rangle_\Gamma.$$

We also define

$$C_1(x) = \frac{1}{|\Gamma|\;\varphi'(\tilde{d}_\Gamma(x))}\|\tilde{d}_\Gamma - \tilde{d}_{\Gamma_0}\|^{-1}_{L^2(D)}\;\;\left(\tilde{d}_\Gamma(x) - \tilde{d}_{\Gamma_0})(x)\right),$$

so that

$$\mathcal{G}(\|\tilde{d}_\Gamma - \tilde{d}_{\Gamma_0}\|_{L^2(D)},\boldsymbol{\beta}) = \int_D\int_\Gamma B(x,y)C_1(x)\beta(y)\;d\Gamma(y)\;dx = \int_\Gamma\left(\int_D B(x,y)C_1(x)\;dx\right)\beta(y)\;d\Gamma(y). \tag{5.16}$$

We then define the vector quantity

$$\mathbf{C}_2(x) = \frac{1}{|\Gamma|\;\varphi'(\tilde{d}_\Gamma(x))}\|\nabla(\tilde{d}_\Gamma - \tilde{d}_{\Gamma_0})\|^{-1}_{\mathbf{L}^2(D)}\;\;\nabla(\tilde{d}_\Gamma - \tilde{d}_{\Gamma_0})(x),$$

so that

$$\begin{aligned}\mathcal{G}(\|\nabla(\tilde{d}_\Gamma - \tilde{d}_{\Gamma_0})\|_{\mathbf{L}^2(D)},\boldsymbol{\beta}) =& \\ -\int_D\int_\Gamma \frac{\varphi^{"}}{\varphi'}(\tilde{d}_\Gamma(x))\left(\mathbf{C}_2(x)\cdot\nabla\tilde{d}_\Gamma(x)\right)B(x,y)\beta(y)\;d\Gamma(y)\;dx+& \\ \int_D\int_\Gamma\left(\mathbf{C}_2(x)\cdot\nabla B(x,y)\right)\beta(y)\;d\Gamma(y)\;dx =& \\ \int_\Gamma\left(\int_D\left(\mathbf{C}_2(x)\cdot\nabla B(x,y) - \frac{\varphi^{"}}{\varphi'}(\tilde{d}_\Gamma(x))\left(\mathbf{C}_2(x)\cdot\nabla d(x,\Gamma)\right)B(x,y)\right)\;dx\right)\beta(y)\;d\Gamma(y)&\end{aligned} \tag{5.17}$$

Combining equations (5.16) and (5.17) we obtain the corresponding gradient $\nabla\tilde{\rho}_D(\Gamma,\,\Gamma_0)$:

$$\nabla\tilde{\rho}_D(\Gamma,\,\Gamma_0)(y) = \int_D\left[B(x,y)\left(C_1(x) - \frac{\varphi^{"}}{\varphi'}(\tilde{d}_\Gamma(x))\left(\mathbf{C}_2(x)\cdot\nabla\tilde{d}_\Gamma(x)\right)\right) + \mathbf{C}_2(x)\cdot\nabla B(x,y)\right]\;dx,$$

and this completes the proof. □

Chapter 6

Intrinsic Differentiation

Abstract

This chapter is dedicated to preliminary remarks on how to compute the gradient of a shape. For a given representation of a shape (often a finite set of parameters), what should one take the derivative with respect to ? How should one use the computed gradient to let the shape evolve ? The intuitive answers are sometimes misleading.

6.1 Differentiation with respect to parameters

In practice, an evolving shape S is often modeled as a polyhedron. The coordinates (p_i) of its vertices are easy, intuitive parameters to deal with. Shapes represented by splines are also easily characterized by their spline parameters. Shapes defined as plots or zeros of polynomials have for natural parameters the coefficient of these polynomials. One could also consider, that, concerning the level-sets method, the shapes are embedded in a higher dimension space and their "parameters" are the values of their distance function on a grid. In a word, all practical representations of shapes are based on some kind of parameters which are often confused with the shapes themselves. This should not be a problem since giving the values of the parameters is equivalent to giving the shape.

However, when it comes to take the derivative of an energy E with respect to the shape in a minimization framework, the temptation to differentiate with respect to the parameters p_i is very attractive. But one should be aware that this way of differentiating is absolutely not a neutral choice or without consequences.

In order to set things clearly, let S be a shape, entirely characterized (possibly redundantly) by n parameters $(p_i)_{1 \leqslant i \leqslant n}$. The energy to be minimized is a function of the shape S which can be seen as a function of the parameters: $E(S) = E(S((p_i)_i)) = E((p_i)_i)$. Three notions of differentiation arise:

1. $\nabla_{L^2} E(S)$
2. $\partial_{p_i} E(S((p_j)_j))$
3. $d_{p_i} E(S((p_j)_j))$.

The meaning of the first one is that, for any infinitesimal deformation field δS on the shape S, the variation of the energy is $DE(S)(\delta S) = \langle \nabla_{L^2} E(S) \,|\delta S\rangle_{L^2}$.

The meaning of the second derivative, which is the partial derivative with respect to the i-th argument of S and which should be noted ∂_i to avoid confusion, is that a variation δa of this argument increases the function by $DE(S)\left(\partial_i S(p_1, p_2, ..., p_i, ...) \cdot \delta a\right)$.

The meaning of the third derivative, which is the total derivative with respect to parameter p_i, is that the variation of the energy for a variation δp_i of the parameter p_i increases the energy by $DE(S)\left(d_{p_i} S((p_j)_j) \cdot \delta p_i\right)$. The point is that, if there is a constraint, like $p_1 + p_i = 0$, which has to be taken into account, then this derivative involves also ∂_1. More generally, when the parameters p_i are "fake" parameters in the sense that they can themselves be described as functions of other, subjacent, independent parameters (which are possibly part of the mentioned p_i), then the total derivative with respect to these "real", independent parameters should be carefully computed.

What are the links between those derivatives ? If the parameters are not independent, the partial and total derivatives differs in that:

$$d_{p_i} S((p_j)_j) \cdot \delta p_i = \sum_j \partial_j S(p_1, p_2, ..., p_j, ...) \cdot d_{p_i} p_j \cdot \delta p_i$$

For the next point, let us suppose that the parameters are independent, for the sake of readability. The deformation field on S induced by a variation (δp_i) of all parameters simultaneously is $\sum_i d_{p_i} S \cdot \delta p_i$. Consequently, the induced variation of the energy is:

$$DE(S)\left(\sum_i d_{p_i} S \cdot \delta p_i\right) = \sum_i DE(S)\left(d_{p_i} S \cdot \delta p_i\right)$$

which can be seen directly as a linear function of the parameter variation:

$$L \ : \ (\delta p_i)_i \in \mathbb{R}^n \ \mapsto \ L((\delta p_i)_i) = \sum_i DE(S)\left(d_{p_i} S \cdot \delta p_i\right)$$

Until now, nothing really surprising. But now, let us consider the respectively associated inner products. The intuitive gradient descent method would consist in considering the variation of the parameters as a element of $\mathbb{R}^n$, naturally embedded with the L^2 inner product. The Riesz theorem is then applied to the continuous linear form L from $\mathbb{R}^n$ to $\mathbb{R}$ and gives an existing, unique vector L^* of $\mathbb{R}^n$ such that, for any parameter variation $(\delta p_i)_i$:

$$\langle L^* \, | (\delta p_i)_i \rangle_{L^2(\mathbb{R}^n)} = DE((p_j)_j)((\delta p_i)_i).$$

It seems then very natural to choose $(\delta p_i)_i = L^*$, i.e. $\delta p_i = DE(S)(d_{p_i} S)^*$. So now, if there is a problem, where is it ? It lies on the fact that the L^2 inner product on $\mathbb{R}^n$ gives the same importance to all parameters p_i, that is to say, that all deformations caused by the same variation of any parameter p_i are considered to have the same importance, without taking into account that this same variation could induce small or big shape variations (for a given norm in the tangent space, set of all shape variations) depending on the chosen parameter p_i which undergoes this parameter variation. This intuitive method is consequently not neutral at all, but on the contrary sets a significant prior on the typical deformations that the evolving shape will undergo, and this prior comes from the choice of the parameterization of the shape. Consequently, two different representation methods (for instance, polyhedron versus splines) will lead to qualitatively different evolutions.

Is it feasible to build a "neutral" gradient, which would depend as little as possible on the shape representation method ?

6.2 Intrinsic differentiation

Let us consider the tangent space of the shape S, set of all admissible deformation fields on S, embedded with an inner product, typically the $L^2(S \to \mathbb{R})$ one. The associated gradient $\nabla_{L^2} E(S)$ does not need any introduction anymore: it is the unique field satisfying, for all deformation fields δS defined on the shape:

$$DE(S)(\delta S) = \langle \nabla_{L^2} E(S) \,|\delta S\rangle_{L^2} .$$

$\nabla_{L^2} E(S)$ is the best direction to maximize the energy E, best direction being taken in the sense of the L^2 inner product on S. This notion of best direction does not depend on any shape representation, it is an entirely intrinsic notion (oppositely to L^*); that's why this is the one which deserves consideration.

The gradient $\nabla_{L^2} E(S)$ is the deformation field that the shape should undergo if its representation method could allow it to. However, due to discretization problems, the shape is characterized by a finite number of parameters. The natural thing to do then is to search for the best deformation field which is admissible for the current representation, that is to say the projection of the gradient $\nabla_{L^2} E(S)$ onto the set of computable deformations $\mathcal{D} = \left\{ \sum_i d_{p_i} S(p_1, p_2..., p_n) \cdot \delta p_i, \text{ with } (\delta p_i)_i \in \mathbb{R}^n \right\} = \text{Span}\left(d_{p_i} S(p_1, p_2..., p_n)\right)$, which is the vector subspace of $L^2(S \to \mathbb{R})$ generated by the n deformations $d_{p_i} S$.

In order words, if δS is a deformation field of $\mathcal{D}$, associated to a parameter variation $(\delta p_i)_i$ by $\delta S = \sum_i d_{p_i} S \cdot \delta p_i = DS(\,(\delta p_i)_i\,)$, and if the projection of any field defined on S onto the subspace $\mathcal{D}$ (for the L^2 inner product) is denoted by $P_{\mathcal{D}}$, then:

$$\begin{aligned}
DE(\,(\delta p_i)_i\,) &= DE(S)(\delta S) \\
&= \langle\, \nabla_{L^2} E(S) \mid \delta S \,\rangle_{L^2(S)} \\
&= \langle\, P_{\mathcal{D}}\left(\nabla_{L^2} E(S)\right) \mid \delta S \,\rangle_{L^2(S)} \\
&= \langle\, P_{\mathcal{D}}\left(\nabla_{L^2} E(S)\right) \mid \delta S \,\rangle_{L^2(\mathcal{D})} \\
&= \left\langle\, DS\Big(\, (DS)^{-1} \circ P_{\mathcal{D}}\left(\nabla_{L^2} E(S)\right) \,\Big) \;\middle|\; DS\left((\delta p_i)_i\right) \right\rangle_{L^2(\mathcal{D})} \\
&= \left\langle\, (DS)^{-1} \circ P_{\mathcal{D}}\left(\nabla_{L^2} E(S)\right) \mid (\delta p_i)_i \,\right\rangle_{DS(\mathbb{R}^n)}
\end{aligned}$$

with the new inner product defined on $\mathbb{R}^n$:

$$\left\langle\, (\delta p_i)_i \;\middle|\; (\delta' p_i)_i \,\right\rangle_{DS(\mathbb{R}^n)} = \left\langle\, DS(\,(\delta p_i)_i\,) \mid DS(\,(\delta' p_i)_i\,) \,\right\rangle_{L^2(\mathcal{D})}$$

This inner product takes into account the real cost of the shape variations induced by a parameter variation. The natural, shape-intrinsic "parameter gradient" that appears is consequently

$$(DS)^{-1} \circ P_{\mathcal{D}}\left(\nabla_{L^2} E(S)\right),$$

where DS^{-1} is the inverse function of the derivative of S (seen as a function of $(\delta p_i)_i$) and associates to each realizable field (i.e. in $\mathcal{D}$) the corresponding parameter variation $(\delta p_i)_i$. The observed deformation of S when applying this gradient will be the projection of the usual L^2 gradient into the set of all admissible shape deformations for the given shape representation, which fits the intuition of "best deformation field".

This gradient is to be compared to the previous, non-intrinsic "parameter gradient" $L^* = DE(S) \circ DS^* = (\nabla_{L^2} E(S) \cdot DS(\cdot))^*$ which leads to the shape deformation $DS \circ (\nabla_{L^2} E(S) \cdot DS(\cdot))^*$.

Discretization

Some people compute the intrinsic shape gradient $\nabla_{L^2} E(S)$ with respect to the shape S and then evaluate it at the vertices v_i of the polyhedron that results from the discretization of the shape. The obtained field $\mathbf{f} = (\nabla_{L^2} E(S)(v_i))_i$ differs from the one obtained by the projection of the shape gradient onto the subspace $\mathcal{S}$ of admissible, discretized deformation fields, and without additional hypotheses there is no reason why this field $\mathbf{f}$ should increase the energy, on the opposite of the field L^* which is a gradient (for a different inner product that the recommend, intrinsic one). However, with a smoothness hypothesis, the field $\mathbf{f}$ can be shown to increase the energy. For example, in the case of planar curves, if the absolute value of the second derivative (with respect to the arc length) of the gradient $\mathbf{g} = \nabla_{L^2} E(S)$ is known to be bounded, then it is sufficient that the maximum distance between two successive vertices is smaller than $\sqrt{\frac{\|\mathbf{g}\|_{L^2}}{\sup_S |\mathbf{g}''|}}$ to guarantee the positivity of $DE(S)(\mathbf{f})$.

Conclusion

The intuition, which at first glance may guide to directly differentiate with respect to the parameters (for the usual L^2 inner product in the space of these parameters), is misleading. Happily there happens to be a "more intrinsic" way to compute the gradient, whose only restriction towards the usual shape gradient is that it belongs to the set of admissible deformations imposed by the choice of shape representation.

Chapter 7

Generalized Gradients: Priors on Minimization Flows

Abstract

This chapter tackles an important aspect of the variational problem underlying active contours: optimization by gradient flows. Classically, the definition of a gradient depends directly on the choice of an inner product structure. This consideration is largely absent from the active contours literature. Most authors, explicitly or implicitly, assume that the space of admissible deformations is ruled by the canonical L^2 inner product. The classical gradient flows reported in the literature are relative to this particular choice. Here, we investigate the relevance of using (i) other inner products, yielding other gradient descents, and (ii) other minimizing flows not deriving from any inner product. In particular, we show how to induce different degrees of spatial consistency into the minimizing flow, in order to decrease the probability of getting trapped into irrelevant local minima. We report numerical experiments indicating that the sensitivity of the active contours method to initial conditions, which seriously limits its applicability and efficiency, is alleviated by our application-specific spatially coherent minimizing flows. We show that the choice of the inner product can be seen as a prior on the deformation fields.

7.1. Introduction

Many problems in computer vision can advantageously be cast in a variational form, i.e. as a minimization of an energy functional. In this chapter, we focus on variational methods dedicated to the recovery of contours. In this case, the problem amounts to finding a contour which corresponds to a global minimum of the energy. Unfortunately, in most cases, the exact minimization of the energy functional is computationally unfeasible due to the huge number of unknowns.

The *graph cuts* method is a powerful energy minimization method which allows to find a global minimum or a strong local minimum of an energy. In the last few years, this method has been successfully applied to several problems in computer vision, including stereovision [69] and image segmentation [11]. However, it has a severe limitation: it cannot be applied to an arbitrary energy function [70], and, when applicable, is computationally expensive.

Hence, in most cases, a suboptimal strategy must be adopted. A common minimization procedure consists in evolving an initial contour, positioned by the user, in the direction of

steepest descent of the energy. This approach, known in the literature as *active contours* or *deformable models*, was pioneered by Kass. et al. in [64] for the purpose of image segmentation. Since, it has been applied in many domains of computer vision and image analysis (image segmentation [14], surface reconstruction [121, 40], stereo reconstruction [44, 62, 50], etc.).

However, due to the highly non-convex nature of most energy functionals, a gradient descent flow is very likely to be trapped in a local minimum. Also, this local minimum depends on the position of the initial contour. If the latter is far from the expected final configuration, the evolution may be trapped in a completely irrelevant state. This sensitivity to initial conditions seriously limits the applicability and efficiency of the active contours method.

We detail in section 7.2 the general gradient descent process so as to emphasize the crucial role of the inner product. After an abstract study in section 7.3 on how to handle inner products and minimizing flows, we propose, in section 7.4, various inner products and show how they induce different degrees of spatial coherence in the minimizing flow with numerical examples of shape warping in section 7.5.

7.2. Minimization and inner product

In the following we consider a shape Γ, seen as a manifold of dimension k embedded in $\mathbb{R}^n$, for example a planar curve or a surface in the space $\mathbb{R}^3$. We denote by $E(\Gamma)$ the energy functional to be minimized. In order to define the gradient of the energy functional, the first step is to compute its Gâteaux derivatives $\delta E(\Gamma, v)$ in all directions, i.e. for all admissible velocity fields v defined on the shape Γ with values in $\mathbb{R}^n$. The deformation space, set of all these fields v, can be seen as the tangent space of Γ, considered itself as a point in the manifold of all admissible shapes.

$$\delta E(\Gamma, v) \stackrel{def}{=} \lim_{\epsilon \to 0} \frac{E(\Gamma + \epsilon v) - E(\Gamma)}{\epsilon} \ . \tag{7.1}$$

where $\Gamma + \epsilon v$ is the shape defined, for any parametrization of Γ, say $\mathcal{P}_\Gamma : \sigma \in S \subset \mathbb{R}^k \mapsto \mathcal{P}_\Gamma(\sigma) \in \Gamma \subset \mathbb{R}^n$, by the parametrization $\mathcal{P}_{\Gamma+\epsilon v} : \sigma \in S \mapsto \mathcal{P}_{\Gamma+\epsilon v}(\sigma) = \mathcal{P}_\Gamma(\sigma) + \epsilon v(\mathcal{P}_\Gamma(\sigma)) \in \mathbb{R}^n$.

Then, we would like to pick the gradient as the direction of steepest descent of the energy. However, it is not yet possible at this stage: to be able to assess the steepness of the energy, the deformation space needs additional structure, namely an inner product introducing the geometrical notions of angles and lengths. This consideration is largely absent from the active contours literature: most authors, explicitly or implicitly, assume that the deformation space is ruled by the canonical L^2 inner product on Γ, which is, for two deformation fields u and v:

$$\langle u \,|\, v \rangle_{L^2} = \frac{1}{|\Gamma|} \int_\Gamma u(\mathbf{x}) \cdot v(\mathbf{x}) \, d\Gamma(\mathbf{x}) \ ,$$

where $d\Gamma(\mathbf{x})$ stands for the area element of the contour so that the integral over Γ is intrinsic and does not depend on the parametrization.

Here, for sake of generality, we model the space of admissible deformations as an inner product space $(F, \langle | \rangle_F)$. If there exists a deformation field $u \in F$ such that

$$\forall v \in F, \ \delta E(\Gamma, v) = \langle u \,|\, v \rangle_F \ ,$$

then u is unique, we call it the gradient of E *relative to the inner product* $\langle | \rangle_F$, and we denote by $u = \nabla_F E(\Gamma)$. The existence of u is related to the smoothness of E, or more exactly to the continuity of $\delta E(\Gamma, v)$ with respect to v (Riesz representation theorem, see [100] for more details).

Clearly, each choice of inner product yields its own gradient. This is often neglected and most authors improperly refer to *the* gradient of the energy. Thus, the classical gradient flows reported in the literature (mean curvature flow, geodesic active contours [14, 51, 106], multi-view 3D reconstruction [44, 62, 50]) are relative to the L^2 inner product.

The gradient descent method consists in deforming an initial contour Γ_0 in the opposite direction of the gradient.

$$\left\{ \begin{array}{l} \Gamma(0) = \Gamma_0 \\ \dfrac{d\Gamma}{dt} = -\nabla_F E(\Gamma) \end{array} \right. \tag{7.2}$$

The problem of the existence and the uniqueness of this minimizing flow is out of the scope of this chapter. Indeed, it is highly dependent on the properties of each particular energy functional. If this evolution exists, it decreases the energy:

$$\frac{dE(\Gamma)}{dt} = - \left\| \nabla_F E(\Gamma) \right\|_F^2 \leq 0 \; .$$

The standard choice for F is the Hilbert space of square integrable velocity fields $L^2(\Gamma, \mathbb{R}^n)$ equipped with its canonical inner product. Very few authors in the active contours area have considered using other inner products, whereas this is an established technique in image registration, as in [112] where Trouvé designs specifically an inner product to help a gradient descent. Very recently, in the context of shape representation and analysis, Michor and Mumford [80] and Yezzi and Mennucci [116] have shown that slightly modifying the L^2 inner product allows to build well-behaved metrics in the space of curves; the particular case of the H^1 inner product has been simultaneously and independently investigated by us [21] and by Sundaramoorthi et al. [107].

The variations on the gradient descent theme, as in [9], will still be applicable to the new gradients we propose, since these methods are in fact not specific to the particular L^2 gradient.

Minimizing flows not deriving from any inner product, that is to say evolutions that decrease the energy, without any gradient interpretation, have also been overlooked so far. Note that any evolution fulfilling the condition

$$\frac{dE(\Gamma)}{dt} = \left\langle \nabla_F E(\Gamma) \left| \frac{d\Gamma}{dt} \right. \right\rangle_F \leq 0 \tag{7.3}$$

is a candidate to solve the minimization problem. This idea, proposed in [106], is applied by the same authors to the alignment of curve in images in [87]: a complicated term in the gradient is safely neglected after checking that the evolution still decreases the energy.

The spirit of our work is different. We do not focus either on a specific inner product or on a particular energy functional. We rather explore general procedures to build some new inner products and to compute the associated gradients. We also address the design of non-gradient minimizing flows.

Our motivation is also different. Our primary concern in this work is the sensitivity of the active contours method to initial conditions. There are essentially two ways of dealing with

this problem: positioning the initial contour very close to the expected final configuration, or using a multiresolution coarse-to-fine strategy, in other words running the optimization on a series of smoothed and subsampled contours and input data. In this chapter, we pioneer a third way to tackle the problem of unwanted local minima: the careful design of the minimizing flow.

We do not modify the energy, hence the relief of the energy landscape and in particular the "number" of local minima remains unchanged. But by using an evolution that favors certain types of directions, we expect to decrease the probability of falling into unwanted energy basins.

Typically, in many applications, spatially coherent motions are to be preferred over erratic evolutions. For example, in the tracking problem, the object of interest is likely to have similar shapes in consecutive frames. So if we initialize the contour with the result of the previous frame, it makes sense to encourage the motions which preserve its overall appearance. This way, it may be easier to dodge unexpected local low-energy configurations. A traditional L^2 gradient descent definitely does not have this desirable property since the L^2 inner product completely disregards the spatial coherence of the velocity field.

7.3. New Inner Products and New Flows

In this section, we suppose that the space F of all admissible deformations of the shape Γ is initially equipped with the inner product $\langle | \rangle_F$, for example L^2 in the standard case, and we study how to build new inner products or new minimizing flows from the given one.

7.3.1. Designing new inner products

Definition 56. *For any symmetric positive definite linear operator $\mathcal{L} : F \to F$, a new inner product can be defined by*

$$\langle u \,|v\rangle_{\mathcal{L}} = \langle \mathcal{L}u \,|v\rangle_F \ . \tag{7.4}$$

Here, for simplicity, we assume that the domain and the range of $\mathcal{L}$ are equal to F. A similar study is possible if they are strictly smaller than F, under certain conditions, using the Friedrichs extension of $\mathcal{L}$ (see [2] for details). But these technical details are out of the scope of this chapter.

The following observation is central to our work:

Proposition 57. *If $\nabla_F E(\Gamma)$ exists and if $\mathcal{L}$ is also invertible, then $\nabla_{\mathcal{L}} E(\Gamma)$ also exists and we have*

$$\nabla_{\mathcal{L}} E(\Gamma) = \mathcal{L}^{-1} \left(\nabla_F E(\Gamma) \right) \ . \tag{7.5}$$

Proof. Indeed:

$$\begin{aligned} \forall v \in F,\ \delta E(\Gamma, v) &= \langle \nabla_F E(\Gamma) \,|v\rangle_F \\ &= \left\langle \mathcal{L}\mathcal{L}^{-1} \nabla_F E(\Gamma) \,|v \right\rangle_F \\ &= \left\langle \mathcal{L}^{-1} \nabla_F E(\Gamma) \,|v \right\rangle_{\mathcal{L}} \ . \end{aligned}$$

□

The above procedure is of great practical interest because it allows to upgrade any existing L^2 gradient flow. However, it is not completely general in the sense than all inner products cannot be expressed in this form.

Nevertheless, if F is a separable Hilbert space (i.e. complete with respect to the norm $\|\cdot\|_F$), the Riesz representation theorem tells us that any inner product $\langle | \rangle_{\mathcal{L}}$ such that

$$\exists C > 0,\ \forall u \in F,\ \|u\|_{\mathcal{L}} \leqslant C \|u\|_F$$

can be written in the form of equation (7.4). This suggests that our procedure accounts for a wide range of inner products.

7.3.2. Designing new minimizing flows

In this subsection, we follow the inverse approach. Instead of working with the inner product, we apply a linear operator $\mathcal{L} : F \to F$ to the gradient, and we study the properties of the resulting flow:

$$\frac{d\Gamma}{dt} = -\mathcal{L}\big(\nabla_F E(\Gamma)\big)\ . \tag{7.6}$$

This naturally sets up a hierarchy among different types of operators:

- if $\mathcal{L}$ is positive, the energy is non-increasing along the flow (7.6). Indeed,

$$\frac{dE(\Gamma)}{dt} = -\left\langle \nabla_F E(\Gamma) \,|\mathcal{L}\nabla_F E(\Gamma)\right\rangle_F \leqslant 0\ .$$

- if $\mathcal{L}$ is positive definite, the energy strictly decreases along the flow (7.6) until a critical point of the original gradient flow (7.2) is reached.

- if $\mathcal{L}$ is symmetric positive definite and invertible, the flow (7.6) coincides with a gradient descent relative to the inner product $\langle | \rangle_{\mathcal{L}^{-1}}$, as defined in equation (7.4).

The third case is contained in Subsection 7.3.1. A useful example of the second case is given in Subsection 7.4.3.

7.3.3. Adding an orthogonal term

The rate of decrease of the energy when following the direction of descent $\frac{d\Gamma}{dt}$ is given by:

$$\frac{dE(\Gamma)}{dt} = \left\langle \nabla_F E(\Gamma) \,\middle|\, \frac{d\Gamma}{dt} \right\rangle_F \leqslant 0\,.$$

In particular, for the usual evolution $\frac{d\Gamma}{dt} = -\nabla_F E(\Gamma)$, we have:

$$\frac{dE(\Gamma)}{dt} = -\|\nabla_F E(\Gamma)\|_F^2$$

If we denote by v any vector field defined on Γ such as $\langle \nabla_F E(\Gamma)\,|v\rangle_F = 0$, then adding such a vector field v to the usual gradient descent term will not change the amount of decreased energy:

$$\frac{dE(\Gamma)}{dt} = \langle \nabla_F E(\Gamma)\,|-\nabla_F E(\Gamma) + v\rangle_F = -\|\nabla_F E(\Gamma)\|_F^2$$

so we can choose the field v which we would like to add to the initial gradient. Rather than choosing $v = 0$ as usual, we could for example choose one, noted $\hat{v}$, that minimizes a regularizing criterion $R(-\nabla_F E(\Gamma) + v)$:

$$\hat{v} \;=\; \underset{v \perp \nabla_F E(\Gamma)}{\arg\min}\; R(-\nabla_F E(\Gamma) + v) \tag{7.7}$$

In fact this remark still stands when the choice of the direction of descent is not the gradient itself. If we denote by u the initially proposed deformation field $\frac{d\Gamma}{dt}$, then adding a vector field which is orthogonal to the gradient $\nabla_F E(\Gamma)$ will not change the amount of decreased energy at this step of the gradient descent (but will change the evolution):

$$\frac{dE(\Gamma)}{dt} = \langle \nabla_F E(\Gamma) \,|-u+v\rangle_F = \langle \nabla_F E(\Gamma) \,|-u\rangle_F$$

Note that the notion of being orthogonal to the gradient is independent from the chosen inner product. Indeed, if F and G are two different inner products, $\nabla_F E$ and $\nabla_G E$ the associated gradients, and $\perp_F$ and $\perp_G$ the associated notions of orthogonality, we have:

$$\langle \nabla_F E(\Gamma) \,|v\rangle_F = \delta E(\Gamma, v) = \langle \nabla_G E(\Gamma) \,|v\rangle_G$$

so, consequently:

$$\begin{aligned} \langle \nabla_F E(\Gamma) \,|v\rangle_F = 0 \;&\iff\; \langle \nabla_G E(\Gamma) \,|v\rangle_G = 0 \\ \nabla_F E(\Gamma) \;\perp_F\; v \;&\iff\; \nabla_G E(\Gamma) \;\perp_G\; v\,. \end{aligned}$$

7.4. Some Spatially Coherent Minimizing Flows

This theoretical study has brought us the tools we need to better apprehend minimizing flows and build new ones. We now propose some minimizing flows yielding different degrees of spatial coherence. We insist on the fact that this spatial coherence has nothing to do with an eventual regularity term in the energy functional. We do not modify the energy, so the regularity constraint on the contour remains unchanged. We modify the trajectory of the minimizing flow, by favoring spatially coherent motions, but this does not condition the regularity of the final contour.

In the following, we sometimes use differential geometry. We refer the reader to [36] for the basic notions.

7.4.1. Motion decomposition

A simple and useful procedure, to design new inner products yielding spatially coherent flows, is to decompose the deformation space into a sum of several mutually orthogonal linear subspaces, and to apply different penalty factors to the different types of motions. Typically, the subspaces are chosen according to an application-specific hierarchy of the motions. For example, rigid/non-rigid, affine/non-affine, etc.

We suppose that such an orthogonal (with respect to $\langle|\rangle_F$) decomposition of the deformation space F into N closed linear subspaces is available:

$$F = F_1 \perp F_2 \perp \cdots \perp F_N\,.$$

Then a new inner product is derived from $\langle | \rangle_F$ by applying the procedure of Subsection 7.3.1 with

$$\mathcal{L} = \bigoplus_{i=1}^{N} \lambda_i \, \mathrm{Id}_{F_i} \ ,$$

where $\forall i,\ \lambda_i > 0$. The lower is λ_i, the shorter is the norm of the velocity fields of subspace F_i, and the stronger will be this type of motion in the new gradient flow.

Obviously, $\mathcal{L}$ is symmetric positive definite and invertible. If $\nabla_F E$ exists, so does $\nabla_{\mathcal{L}} E$ and

$$\nabla_{\mathcal{L}} E \ = \ \sum_{i=1}^{N} \frac{1}{\lambda_i} \ \Pi_{F_i} \left(\nabla_F E \right) \ , \tag{7.8}$$

where Π_{F_i} denotes the orthogonal projection on the i^{th} subspace F_i. Of course, if all λ_i are equal to 1, $\nabla_{\mathcal{L}} E$ coincides with $\nabla_F E$.

We apply this general construction to two useful cases. In the first case, we decompose the velocity field into a translation, an instantaneous rotation, a rescaling motion and a non-rigid residual. In the second case, we isolate the instantaneous affine motion.

In the following, we denote by $\mathbf{G} = \left(\int_\Gamma \mathbf{x} \, d\Gamma(\mathbf{x}) \right) / \int_\Gamma d\Gamma(\mathbf{x})$ the center of mass of the shape.

Translation, rotation and scaling

In this paragraph, we focus on the two-dimensional and three-dimensional cases. The expressions below are for the 3D case, but can easily be adapted to 2D.

We denote by T, R and S the subspaces of the translations, the instantaneous rotations around the centroid, and the scaling motions centered on the centroid, respectively, defined on the shape Γ:

$$\begin{aligned} T &= \left\{ v : \mathbf{x} \in \Gamma \mapsto \mathbf{t} \mid \mathbf{t} \in \mathbb{R}^3 \right\} \ , \\ R &= \left\{ v : \mathbf{x} \mapsto (\mathbf{x} - \mathbf{G}) \wedge \omega \mid \omega \in \mathbb{R}^3 \right\} \ , \\ S &= \left\{ v : \mathbf{x} \mapsto s(\mathbf{x} - \mathbf{G}) \mid s \in \mathbb{R} \right\} \ . \end{aligned}$$

These subspaces are mutually orthogonal for the L^2 inner product. Indeed, the L^2 product of any two fields of any two different subspaces (among T, R and S) is zero. For instance, if $v_1 : \mathbf{x} \mapsto \mathbf{t}_1$ is an element of T and $v_2 : \mathbf{x} \mapsto (\mathbf{x} - \mathbf{G}) \wedge \omega_2$ an element of R, then:

$$\langle v_1 \, | v_2 \rangle_{L^2} \ = \ \frac{1}{|\Gamma|} \int_\Gamma \mathbf{t}_1 \cdot \left((\mathbf{x} - \mathbf{G}) \wedge \omega_2 \right) d\Gamma(\mathbf{x}) \ = \ \mathbf{t}_1 \cdot \left(\left[\frac{1}{|\Gamma|} \int_\Gamma \mathbf{x} \, d\Gamma(\mathbf{x}) - \mathbf{G} \right] \wedge \omega_2 \right) \ = \ 0.$$

We suppose that these subspaces are included in the space of admissible deformations F, and that the latter is ruled by the L^2 inner product. We denote by N the orthogonal complement of these subspaces: $F = T \perp R \perp S \perp N$. The orthogonal projection of a velocity field u on one of these subspaces can be found by minimizing $\|u - v\|_F$ with respect to v in the considered subspace. As an example, we detail the computation of $(\Pi_R \ u)$.

As for each element v of R there exists an ω such that $v(\mathbf{x}) = (\mathbf{x} - \mathbf{G}) \wedge \omega$ for all $\mathbf{x}$, we minimize the quantity $\|u - (\cdot - \mathbf{G}) \wedge \omega\|_{L^2}$ with respect to ω.

$$\partial_\omega \left(\int_\Gamma \|u(\mathbf{y}) - (\mathbf{y} - \mathbf{G}) \wedge \omega\|^2 \, d\Gamma(\mathbf{y}) \right) \ = \ \int_\Gamma -\big(u(\mathbf{y}) - (\mathbf{y} - \mathbf{G}) \wedge \omega \big) \wedge (\mathbf{y} - \mathbf{G}) \, d\Gamma(\mathbf{y})$$

$$= -\int_\Gamma u(\mathbf{y}) \wedge (\mathbf{y}-\mathbf{G})\, d\Gamma(\mathbf{y}) + \left(\int_\Gamma \|\mathbf{y}-\mathbf{G}\|^2 d\Gamma(\mathbf{y})\right)\omega - \left[\int_\Gamma (\mathbf{y}-\mathbf{G})(\mathbf{y}-\mathbf{G})^T d\Gamma(\mathbf{y})\right]\omega$$

As this quantity is zero for the ω_u which minimizes $\|u - (\cdot - \mathbf{G}) \wedge \omega\|_{L^2}$, we have:

$$\omega_u = \left[\left(\int_\Gamma \|\mathbf{y}-\mathbf{G}\|^2 d\Gamma(\mathbf{y})\right)\mathrm{Id} - \int_\Gamma (\mathbf{y}-\mathbf{G})(\mathbf{y}-\mathbf{G})^T d\Gamma(\mathbf{y})\right]^{-1} \left(\int_\Gamma u(\mathbf{y}) \wedge (\mathbf{y}-\mathbf{G})\, d\Gamma(\mathbf{y})\right)$$

To guarantee that the linear application between brackets is invertible, we prove it is a symmetric positive definite matrix M. We have indeed for any $\mathbf{x}$:

$$\mathbf{x}^T M \mathbf{x} \;=\; \int_\Gamma \|\mathbf{x}\|^2 \|\mathbf{y}-\mathbf{G}\|^2 - \big(\mathbf{x}\cdot(\mathbf{y}-\mathbf{G})\big)^2\, d\Gamma(\mathbf{y})$$

As for any $\mathbf{x}$ and $\mathbf{z}$ we have $(\mathbf{x}\cdot\mathbf{z}) \leqslant \|\mathbf{x}\|\|\mathbf{z}\|$, with equality only if the two vectors are collinear, and as $\mathbf{x}$ cannot be collinear with all $\mathbf{y}-\mathbf{G}$ for $\mathbf{y}$ in Γ, we obtain $\mathbf{x}^T M \mathbf{x} > 0$ for any $\mathbf{x}$, so M is positive definite and consequently invertible.

Note that if we had not taken for u the L^2 gradient but the gradient for another inner product F, we would have to ensure the subspaces are orthogonal for that inner product F, and compute new projections by minimizing $\|u-v\|_F$.

We apply the method we detailed for the subspace R to the other subspaces T and S, and obtain:

$$\begin{aligned}
(\Pi_T\; u)(\mathbf{x}) &= \overline{u} \;:=\; \frac{1}{|\Gamma|}\int_\Gamma u(\mathbf{y})\, d\Gamma(\mathbf{y}),\\
(\Pi_R\; u)(\mathbf{x}) &= (\mathbf{x}-\mathbf{G}) \wedge \omega_u\,,\\
(\Pi_S\; u)(\mathbf{x}) &= \frac{\int_\Gamma u(\mathbf{y})\cdot(\mathbf{y}-\mathbf{G})\, d\Gamma(\mathbf{y})}{\int_\Gamma \|\mathbf{y}-\mathbf{G}\|^2\; d\Gamma(\mathbf{y})}(\mathbf{x}-\mathbf{G})\,,\\
(\Pi_N\; u)(\mathbf{x}) &= u(\mathbf{x}) - (\Pi_T + \Pi_R + \Pi_S)\,(u)(\mathbf{x})\,.
\end{aligned}$$

In the two-dimensional case, the expressions of the projections are the same, and the expression of ω_u can be simplified in:

$$\omega_u \;=\; \frac{\int_\Gamma (\mathbf{y}-\mathbf{G}) \wedge u(\mathbf{y})\, d\Gamma(\mathbf{y})}{\int_\Gamma \|\mathbf{y}-\mathbf{G}\|^2\; d\Gamma(\mathbf{y})}\,.$$

The new gradient is deduced from the L^2 gradient by equation (7.5) with

$$\mathcal{L}^{-1} = \mathrm{Id} + \left(\frac{1}{\lambda_T} - 1\right)\Pi_T + \left(\frac{1}{\lambda_R} - 1\right)\Pi_R + \left(\frac{1}{\lambda_S} - 1\right)\Pi_S\,.$$

The weights λ_T, λ_R and λ_S are adapted to the user's needs in each particular application. For example:

- Boost rigid+scaling motions: $\lambda_T, \lambda_R, \lambda_S \ll 1$,
- Boost rigid motions: $\lambda_T, \lambda_R \ll 1$, $\lambda_S = 1$,
- Boost translations: $\lambda_T \ll 1$, $\lambda_R = \lambda_S = 1$.

Affine motion

We can apply this same idea to the subspace $\mathcal{A}$ of instantaneous affine motions:

$$\mathcal{A} = \left\{ v : \mathbf{x} \in \Gamma \mapsto \mathbf{A}\mathbf{x} + \mathbf{b} \mid \mathbf{A} \in \mathbb{R}^{n \times n},\ \mathbf{b} \in \mathbb{R}^n \right\} .$$

The L^2 orthogonal projection on this subspace writes:

$$\left(\Pi_A\ u\right)(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b} ,$$

where

$$\begin{aligned} \mathbf{A} &= \left[\int_\Gamma u(\mathbf{y})(\mathbf{y}-\mathbf{G})^T\ d\Gamma(\mathbf{y})\right]\left[\int_\Gamma (\mathbf{y}-\mathbf{G})(\mathbf{y}-\mathbf{G})^T\ d\Gamma(\mathbf{y})\right]^{-1} , \\ \mathbf{b} &= \overline{u} - \mathbf{A}\ \mathbf{G} . \end{aligned}$$

7.4.2. The Sobolev H^1 gradient flow

We consider the canonical inner product of the Sobolev space $H^1(\Gamma, \mathbb{R}^n)$ of square integrable velocity fields with square integrable derivatives, defined on the shape Γ with values in $\mathbb{R}^n$. For two such fields u and v its expression is:

$$\langle u \,|\, v \rangle_{H^1} = \frac{1}{|\Gamma|}\int_\Gamma u(\mathbf{x}) \cdot v(\mathbf{x}) d\Gamma(\mathbf{x}) \ +\ \frac{1}{|\Gamma|}\, l^2 \int_\Gamma D_{\mathbf{x}} u(\mathbf{x}) \cdot D_{\mathbf{x}} v(\mathbf{x}) d\Gamma(\mathbf{x}) ,$$

where $D_{\mathbf{x}}$ denotes the intrinsic derivatives on the contour and l is a characteristic length for the derivation which acts as a weight between the two integrals. The second term of this expression introduces a notion of spatial coherence: not only the length of the velocity field, but also its variations along the contour are penalized. Indeed, $D_{\mathbf{x}}u(\mathbf{x})$ stands for the matrix of the derivative of the vector field u at the point $\mathbf{x}$ on the manifold Γ and consequently expresses how much the field u varies at point $\mathbf{x}$. In the two-dimensional case, $D_{\mathbf{x}}u(\mathbf{x})$ is simply a vector. In the general case, $D_{\mathbf{x}}u(\mathbf{x}) \cdot D_{\mathbf{x}}v(\mathbf{x}) = \sum_{i,j} \left(D_{\mathbf{x}}u(\mathbf{x})\right)_{i,j} \left(D_{\mathbf{x}}v(\mathbf{x})\right)_{i,j}$ is the usual inner product between matrices.

By definition of the gradients of $E(\Gamma)$, and then by integrating by parts on the manifold Γ, we have:

$$\begin{aligned} \forall\ v\,, \quad \langle \nabla_{L^2} E(\Gamma) \,|\, v \rangle_{L^2} &= \delta E(\Gamma, v) \ =\ \langle \nabla_{H^1} E(\Gamma) \,|\, v \rangle_{H^1} \\ &= \langle \nabla_{H^1} E \,|\, v \rangle_{L^2} + l^2\ \langle D_{\mathbf{x}} \nabla_{H^1} E \mid\ D_{\mathbf{x}} v\ \rangle_{L^2} \\ &= \left\langle \nabla_{H^1} E - l^2 \Delta \nabla_{H^1} E \,\middle|\, v \right\rangle_{L^2} \end{aligned}$$

Thus the H^1 inner product is related to the L^2 inner product as proposed in Subsection 7.3.1 through the linear operator $\mathcal{L}(u) = u - l^2\Delta u$, where Δ denotes the intrinsic Laplacian operator on the contour, often called the Laplace-Beltrami operator. As a consequence, the H^1 gradient can be obtained from the L^2 gradient by solving an intrinsic heat equation with a data attachment term:

$$l^2\, \Delta u = u - \nabla_{L^2} E . \tag{7.9}$$

Interestingly, the solution of equation (7.9) coincides with

$$\arg\min_u \int_\Gamma \left\| u(\mathbf{x}) \ -\ \nabla_{L^2} E(\Gamma)(\mathbf{x}) \right\|^2\ d\Gamma(\mathbf{x}) \ +\ l^2 \int_\Gamma \left\| D_{\mathbf{x}} u(\mathbf{x}) \right\|^2\ d\Gamma(\mathbf{x}) \tag{7.10}$$

Intuitively, the H^1 gradient is a smoothed version of the L^2 gradient and can be obtained by a process similar to the image restoration process on a manifold Γ, a problem familiar to the image processing community. The factor l^2 acts as a parameter balancing the influences of the data term and the regularizing term. Actually, smoothing a gradient using this particular inner product is a standard "trick", well-known in numerical analysis. As we mentioned previously, this idea has been introduced in computer vision simultaneously by us [21] and by Sundaramoorthi et al. [107]. However, the main point remains that, introducing this smoothing via a modification of the gradient rather than directly from equation (7.10), warrants that the gradient descent will decrease the energy.

In the two-dimensional case, the shape is a curve which can be parametrized by its arc length σ, so that any field u defined on Γ can be seen as an application from $[\,0, |\Gamma|\,]$ to $\mathbb{R}^2$, where $|\Gamma|$ is the length of the curve. The explicit solution of the equation $\Delta u = u - v$ is then known and given by:

$$u(\sigma) \;=\; \frac{1}{2l}\left(e^{\sigma/l}\left(A - \int_0^{\sigma} e^{-\tau/l}\, v(\tau)\, d\tau\right) \;+\; e^{-\sigma/l}\left(B + \int_0^{\sigma} e^{\tau/l}\, v(\tau)\, d\tau\right)\right) \tag{7.11}$$

$$\text{with } A = \frac{e^{|\Gamma|/l}}{e^{|\Gamma|/l} - 1} \oint_{\Gamma} e^{-\tau/l}\, v(\tau)\, d\tau$$
$$\text{and } B = \frac{1}{e^{|\Gamma|/l} - 1} \oint_{\Gamma} e^{\tau/l}\, v(\tau)\, d\tau\,.$$

Of course, the choice of the initial point on Γ in order to define its parametrization by the arc length does not interfere with the resulting solution considered as an application from Γ into $\mathbb{R}^2$.

In greater dimensions, we can obtain in practice the H^1 gradient, solution of equation (7.9), from an iterative minimization induced by (7.10). Since the work introduced in [7], implementing a PDE on a surface is affordable in the implicit framework with the level set method [34, 86].

7.4.3. Intrinsic Gaussian smoothing

We apply the procedure of Subsection 7.3.2 to design a useful minimizing flow: it is a smoothed version of the L^2 gradient flow. Hence, to some extent, it resembles the H^1 gradient flow of Subsection 7.4.2. However, here, we apply an *ad hoc* procedure to the L^2 gradient without resorting to an inner product.

We define a linear intrinsic smoothing operator which may be seen as the counterpart on the contour of Gaussian smoothing in $\mathbb{R}^{n-1}$, by considering the solution $\tilde{u}$ of the intrinsic heat equation on Γ with initial condition u:

$$\left\{ \begin{array}{l} \tilde{u}(.,0) = u \\ \dfrac{\partial \tilde{u}}{\partial \tau} = \Delta\, \tilde{u} \end{array} \right. , \tag{7.12}$$

where Δ denotes the Laplace-Beltrami operator. We then denote by $\mathcal{L}_\tau\, u$ its solution $\tilde{u}(.,\tau)$ at time $\tau \geq 0$.

On the one hand, $\mathcal{L}_\tau$ is symmetric positive. In particular, a flow (7.6) based on this operator decreases the energy. The larger is τ, the smoother is the flow.

$\mathcal{L}_\tau$ is symmetric:

$$\begin{aligned} \langle \mathcal{L}_0(u) \,|v\rangle_{L^2} &= \langle u \,|\mathcal{L}_0(v)\rangle_{L^2} &&= \langle u \,|v\rangle_{L^2} \ , \\ \frac{\partial}{\partial \tau} \langle \mathcal{L}_\tau(u) \,|v\rangle_{L^2} &= \frac{\partial}{\partial \tau} \langle u \,|\mathcal{L}_\tau(v)\rangle_{L^2} &&= -\langle D_{\mathbf{x}} u \,|D_{\mathbf{x}} v\rangle_{L^2} \end{aligned}$$

$\mathcal{L}_\tau$ is positive:

$$\langle \mathcal{L}_\tau(u) \,|u\rangle_{L^2} = \left\langle \mathcal{L}_{\tau/2}\mathcal{L}_{\tau/2}(u) \,|u\right\rangle_{L^2} = \left\|\mathcal{L}_{\tau/2}(u)\right\|_{L^2} \geq 0$$

But on the other hand, the inversion of $\mathcal{L}_\tau$ for $\tau > 0$ is an ill-posed anti-diffusive process. So a gradient interpretation is not available.

7.5. Numerical Experiments With The New Inner Products

The approach presented in this chapter can be applied to virtually any active contour evolution. In this section, we have chosen some particular applications to demonstrate the interest of our contribution.

Moreover, the content of this chapter is not specific to a particular implementation of the contour evolution. In our experiments, we have used the level set framework [34, 86, 103, 84, 85], motivated by its numerical stability and its ability to handle topological changes automatically. The implicit framework also offers an elegant formulation of the Laplace-Beltrami operator [6] and of the average of a quantity along the contour [95].

The additional computational cost of our approach depends on the type of minimizing flow we consider. The extra time is barely noticeable for the rigid plus scaling and affine flows of paragraphs 7.4.1 and 7.4.1. The latter only require to compute a handful of integrals on the contour. The smooth minimizing flows of Subsections 7.4.2 and 7.4.3 are more demanding. In 2D, the implicit diffusion equations (7.9) and (7.12) are equivalent to some convolutions with respect to the curvilinear coordinate on Γ. In 3D and more, they must be solved with some iterative methods, for each time step.

7.5.1. Shape warping

We illustrate our approach in the problem of shape warping. In this context, the energy functional to be minimized is a measure of dissimilarity between the evolving contour and a target contour. The study of shape metrics is still an active research area [118, 117, 18, 116], and there are many candidates for the dissimilarity measure. We use the differentiable approximation of the Hausdorff distance to warp the contours of two different hands.

Figure 7.1 compares the evolution of the contour when using the L^2 gradient descent *(top row)* and a modified gradient descent favoring rigid plus scaling motions *(bottom row)* as in paragraph 7.4.1. Both evolutions achieve a perfect warping. However, despite the similarity of the two input shapes, the L^2 gradient flow goes through some states of completely different appearances. The trajectory followed by this flow looks particularly inefficient and unnatural, because the notion of length contained in the L^2 inner product is very far from our intuition. In contrast, the behavior of our gradient flow is natural and visually pleasing. Some movies of these evolutions are available in our additional submission data.

In Figure 7.2, we show a three-dimensional warping example from a teddy bear to Hayao Miyazaki's character Totoro. We use here the $W^{1,2}$-norm of the distance functions as proposed in section 5.1.8 in chapter 5. Despite an initial rigid registration, the L^2 gradient descent is unable to give satisfying results. A modified gradient descent favoring rigid plus scaling motions leads to better results.

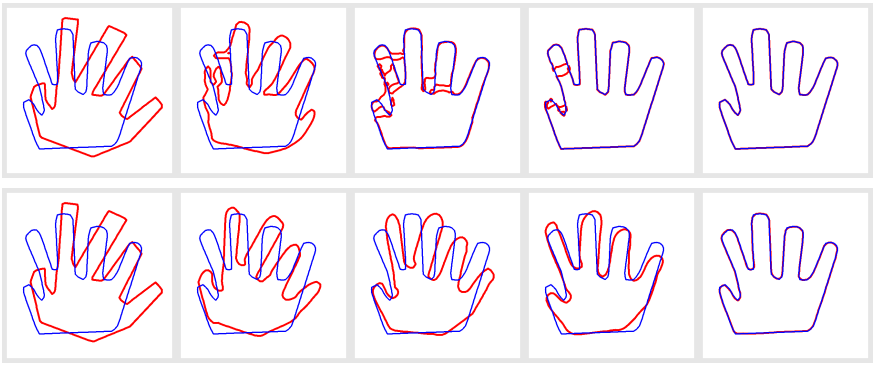


Figure 7.1: Shape warping with the L^2 gradient descent *(top)* and with a modified gradient descent favoring rigid plus scaling motions *(bottom)*: $\lambda_T = \lambda_R = \lambda_S = 0.025$.

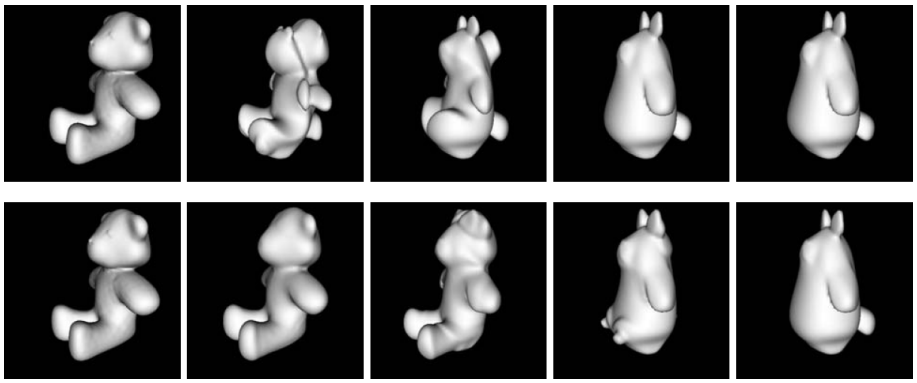


Figure 7.2: 3D shape warping with the L^2 gradient descent *(top)* and with a modified gradient descent favoring rigid plus scaling motions *(bottom)*: $\lambda_T = \lambda_R = \lambda_S = 0.025$.

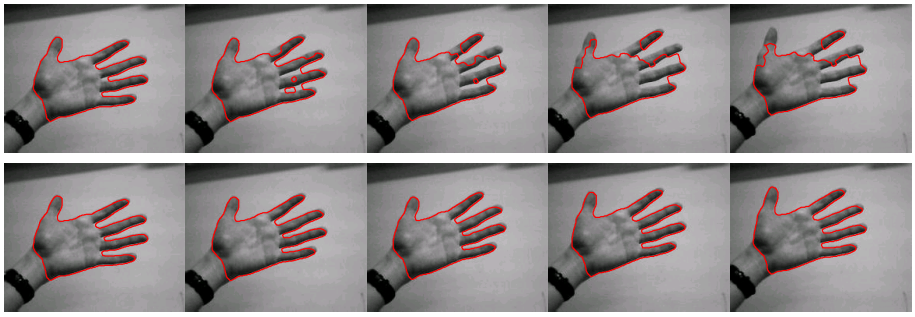


Figure 7.3: Tracking a hand in a video sequence with the L^2 gradient descent *(top)* and with a modified gradient descent favoring affine motions *(bottom)*: $\lambda_A = 0.025$.

This suggests that our approach can infer relevant correspondences between the two contours, as a byproduct of the warping process. This point-to-point matching is obtained by tracking the points along the evolution. It does not make much sense with a L^2 gradient flow, because the latter yields a strictly normal velocity field. But when using our approach, the velocity field has a meaningful tangential part. Maintaining point correspondences during the evolution is straightforward in an implementation with meshes. It is also feasible in a level set implementation, with an extension proposed in [97].

7.5.2. Tracking

We now illustrate the better robustness to local minima of spatially coherent minimizing flows with a naive experiment. We insist on the fact that this example is illustrative: we did not look for the method and the energy that gave the best results of tracking for the particular sequence we worked on; we focus more on the improvements brought by our change of inner product rather than on the results themselves.

We track a moving hand in a monocular video sequence. For each frame, we minimize the contour-based energy of the original geodesic active contours method [14], starting from the result of the segmentation of the previous frame. Note that a region-based approach [91] or a background subtraction method would give better results on our particular test sequence.

Figure 7.3 compares the evolution of the contour when using the L^2 gradient descent *(top row)* and a modified gradient descent favoring affine motions *(bottom row)* as in paragraph 7.4.1. Due to large displacements between consecutive frames, the L^2 gradient flow fails and the contour finally locks between two fingers, whereas our gradient flow manages to dodge this unwanted low-energy configuration.

7.5.3. Landmarks-guided shape warping

Let us study the case of an energy which does not admit a usual L^2 gradient because its formal computation leads to an expression with Dirac peaks in the space of distributions. The problem with such a gradient is that it is not implementable in practice. However, with a suitable choice of another inner product, we naturally obtain a smoothed version of this gradient.

The following work in the next two pages about landmarks has been done by Pierre Maurel and is given here as an example of application of the change of inner product. A

more complete version of his work can be found in [77].

We consider two shapes Γ_1 and Γ_2, and we would like to warp Γ_1 onto Γ_2. It can be useful to consider a small set of landmarks in order to improve the evolution. Provided by the user (anatomical landmarks), or automatically extracted (geometric landmarks), we assume that we are given p pairs of corresponding points on the initial and on the target shapes, $\{(\mathbf{x}_{1i}, \mathbf{x}_{2i}) \in \Gamma_1 \times \Gamma_2, 1 \leq i \leq p\}$. We would like to use the information given by these correspondences to guide the evolution.

Choice of the energy and of the inner product

The usual variational approach consists in minimizing the distance between the evolving shape $\Gamma(t)$ and the target one Γ_2, with initial condition $\Gamma(0) = \Gamma_1$. This distance E_d could be for example the approximation of the Hausdorff distance presented in chapter 2 or the $W^{1,2}$ norm of the signed distance functions over the embedding space. We would like to add to this distance a landmark term $E_{\mathscr{L}}$; the energy to minimize would be consequently:

$$E(\Gamma, \Gamma_2) = E_d(\Gamma, \Gamma_2) + E_{\mathscr{L}}(\Gamma, \Gamma_2)$$

We follow each landmark $\mathbf{x}_{1i}$ from the initial shape Γ_1 during the evolution and denote by $\mathbf{x}_i(t)$ its corresponding point on $\Gamma(t)$. We simply choose for the landmark term:

$$E_{\mathscr{L}} = \sum_i d(\mathbf{x}_i(t), \mathbf{x}_{2i})^2 \tag{7.13}$$

Formally, the energy given by equation (7.13) yields Dirac peaks in the expression of the gradient of the energy:

$$\forall x \in \Gamma, \quad \nabla_{L^2} E(\Gamma)(\mathbf{x}) = \nabla_{L^2} E_d(\Gamma)(\mathbf{x}) + \sum_i \delta_{\mathbf{x}_i(t)}(\mathbf{x})(\mathbf{x}_i(t) - \mathbf{x}_{2i}) \tag{7.14}$$

where $\delta_{\mathbf{x}}$ denotes the Dirac function centered at point $\mathbf{x}$. This is indeed not a good candidate for a gradient descent.

The trick consists in changing the inner product which appears in the definition of the gradient. We use $H^1(\Gamma, \mathbb{R}^n)$, the Sobolev space of square integrable velocity fields with square integrable derivatives, defined and studied in section 7.4.2.

Starting from the irregular gradient $\nabla_{L^2} E(\Gamma)$ given by equation (7.14), we obtain a smooth gradient $\nabla_{H^1} E(\Gamma)$, given by the PDE (7.9) and mathematically justified by an adapted choice of inner product that guarantees a decrease of the energy. In practice and in detail, when solving (7.9), we substitute for each Dirac peak in the expression of $\nabla_{L^2} E(\Gamma)$ a Gaussian with a very small standard deviation.

In the two-dimensional case, the equation (7.11) gives us an explicit expression of the H^1 gradient from the L^2 one thanks to a convolution. In the three-dimensional case we have to deal with the minimization process proposed in section 7.4.2.

Experiments

As a benchmark, we warp some artificial two-dimensional curves with the original energy $E_d = d_{W^{1,2}}$ and test how our landmark-guided force modifies the warping and the final matching. To begin with a simple example, we show in figure 7.4 the warping of a rectangle

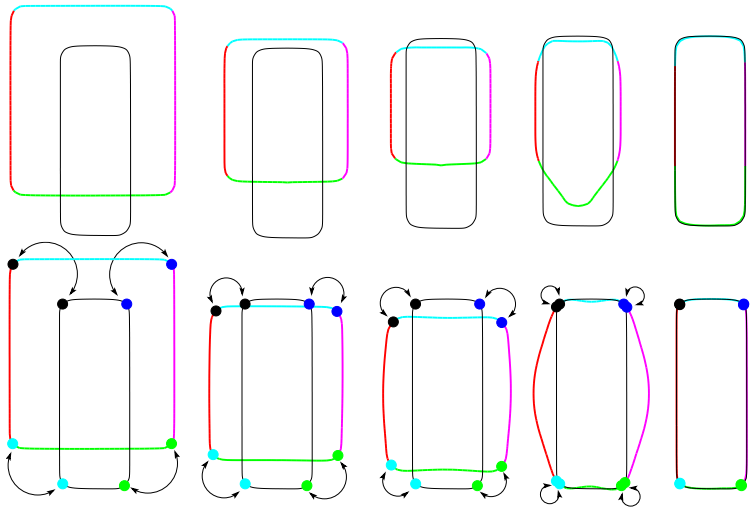


Figure 7.4: Warping of a rectangle shape onto another one. Top row: evolution with $E = d_{W^{1,2}}$. Bottom row: evolution with the same energy, augmented with four provided landmarks, marked by color spots. The colors on the evolving curve shows the evolution of different parts of it. See text for comments.

onto another one. The different parts of the curves are shown with different colors, so that their respective evolution can be followed. Although the initial warping without any landmark seems natural, it fails to discover the matching between the edges of the rectangles, a matching indeed recovered when providing landmarks. Let us now study the case of some real, complex shapes. Figure 7.5 shows the warping between two hands. The energy $E = d_{W^{1,2}}$ yields an unnatural warping. Adding spatially coherent flows improves the warping but still fails in some parts, mainly because the difference between the two shapes can not be summed up to a global motion. With three landmarks only, both a satisfying warping and a good matching are recovered. Figure 7.6 shows the warping of a teddy bear onto a cartoon character. Without any landmarks, the top row evolution fails matching the ears and arms of the characters. The bottom row shows the evolution with four landmarks. Red spots allow to check a good matching between landmarks.

7.6 Combination of the effects of two different inner products

As introduced in part 7.3.1, a gradient related to an other inner product that the usual L^2 inner product can often be obtained by applying a symmetric positive definite linear operator on the usual gradient. More generally, the application of such an operator on the usual gradient guarantees that its results is strictly positively correlated with the usual gradient. The choice of a particular operator allows to set priors on the desired field. For instance, a convolution with a Gaussian function will smooth the field with a characteristic length given by its standard deviation, and the rigidification process previously studied will increase the importance of translations, rotations and scalings.

A question one could be interested in would be the combination of two of these symmetric positive definite operators, in order to benefit from effects of both operators. Does the successive application of two different operators make sense ? If not, how to combine these

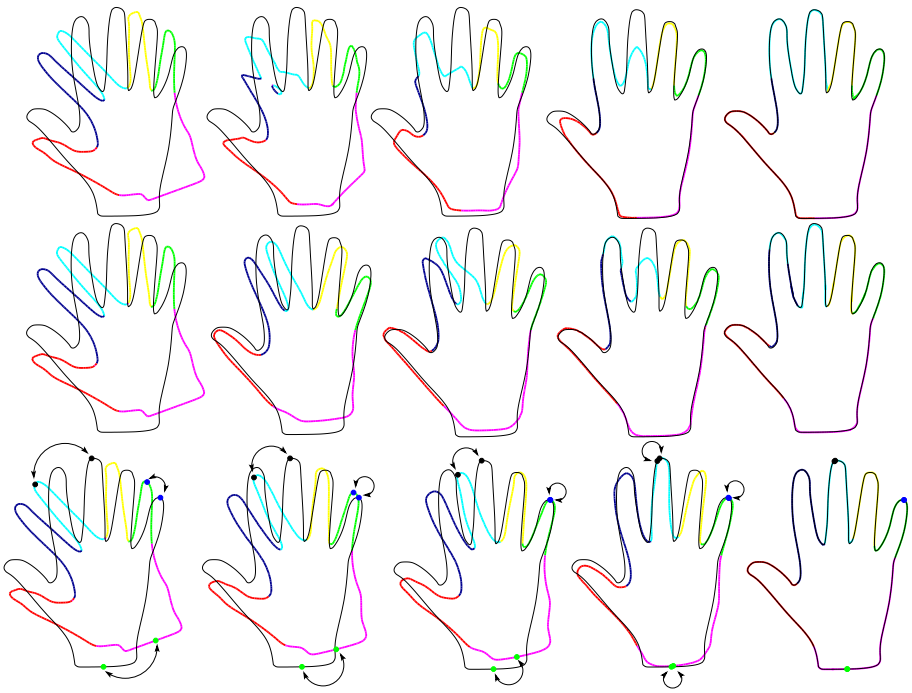


Figure 7.5: Warping of a hand shape onto another one. Top row: evolution with $E = d_{W^{1,2}}$. Middle row: evolution with the same energy plus spatially coherent flows. Bottom row: evolution with the same energy plus coherent flows plus three provided landmarks. See text for comments.

operators ?

The answer to the first question is negative in the general case. If R and N are symmetric positive definite operators defined on the set of deformation fields (the tangent space of the current shape), if E is the energy to be minimized and $\mathbf{v} = \nabla_{L^2} E$ denotes the usual gradient, then $\langle \mathbf{v} \,|R(\mathbf{v})\rangle_{L^2} = DE(R(\mathbf{v})) > 0$ since R is positive definite, and so on for N, but there is no reason why $DE(N \circ R(\mathbf{v})) = \langle \mathbf{v} \,|N \circ R(\mathbf{v})\rangle_{L^2}$ should be positive since $N \circ R$ has no reason to be itself symmetric positive definite. For an example of such a phenomenon, it is sufficient to consider the simple case of symmetric positive definite operators on the plane $\mathbb{R}^2$. They can be written as symmetric matrices 2×2, and the product of two of them is generally not even symmetric:

$$\begin{pmatrix} a & b \\ b & d \end{pmatrix} \times \begin{pmatrix} e & f \\ f & h \end{pmatrix} = \begin{pmatrix} ae + bf & af + bh \\ be + df & bf + dh \end{pmatrix}$$

7.6.1 Nothing but a symmetric way to preserve the symmetry

However, there is many ways to combine two such operators R and N in a new one with the same properties:

Proposition 58. *Let R and N be two symmetric positive definite operators on a linear space $\mathcal{X}$. Then $R \circ N \circ R$ and $N \circ R \circ N$ are also symmetric positive definite operators on $\mathcal{X}$.*

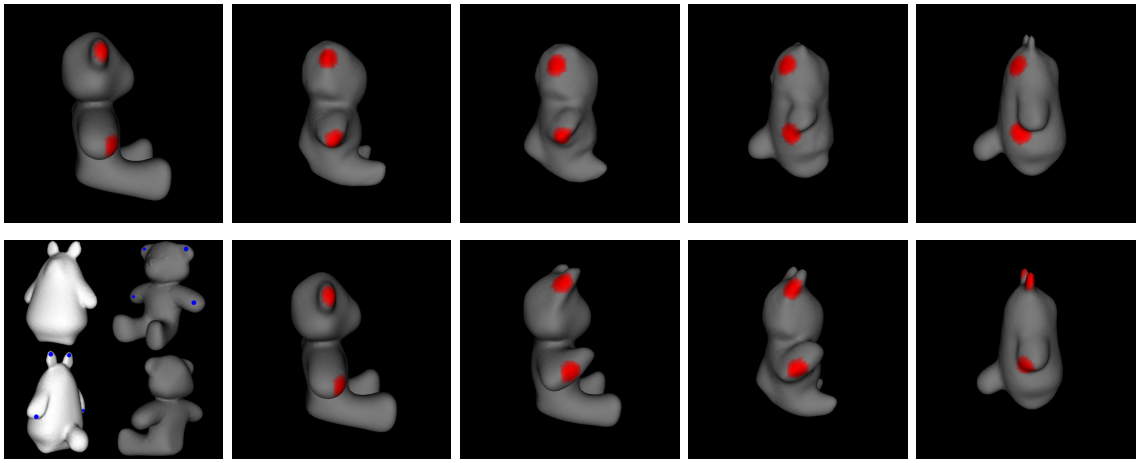


Figure 7.6: Warping of a teddy bear onto a cartoon character. Top row: evolution with $E = d_{W^{1,2}}$. Bottom row, first image: four landmarks provided on the two shapes, indicated by blue spots. Bottom row, remaining images: evolution with $E = d_{W^{1,2}}$ plus the provided landmarks. In red, some parts of the shapes are tracked. See text for comments.

These new operators can consequently be used to build new minimizing flows incorporating some aspects of both transformations R and N.

Proof. In two words, the symmetry of operator R implies:

$$\forall \mathbf{v}, \mathbf{w} \in \mathcal{X}, \ \langle \mathbf{v} \,|R \circ N \circ R(\mathbf{w})\rangle_{\mathcal{X}} \ = \ \langle R(\mathbf{v}) \,|N \circ R(\mathbf{w})\rangle_{\mathcal{X}}$$

so that $R \circ N \circ R$, as a bilinear operator, is:

1. symmetric: $\langle \mathbf{v} \,|R \circ N \circ R(\mathbf{w})\rangle_{\mathcal{X}} = \langle \mathbf{w} \,|R \circ N \circ R(\mathbf{v})\rangle_{\mathcal{X}}$ using also the symmetry of N.

2. positive and definite: $\langle \mathbf{v} \,|R \circ N \circ R(\mathbf{v})\rangle_{\mathcal{X}} = \langle \mathbf{u} \,|N(\mathbf{u})\rangle_{\mathcal{X}}$ with $\mathbf{u} = R(\mathbf{v})$.
$\langle \mathbf{u} \,|N(\mathbf{u})\rangle_{\mathcal{X}}$ is always positive (since N is positive) and zero if and only if $\mathbf{u} = 0$, which happens if and only if $\mathbf{v} = 0$ (since N and R are definite).

□

7.6.2 Better weights in the symmetry for parameterized groups

The operators which are proposed to be applied to the usual gradient are often parameterized, as members of operator families. For example, the Gaussian convolutions are parameterized by their standard deviation, and the rigidification operators by their projection coefficient. These families are generally semigroups for the composition product law, which expresses itself as an addition of the parameters of the two multiplied operators (or sometimes a product of the parameters, which is the same as an addition if the logarithm of the parameters is considered instead) . In this case, the square root of a member O_λ of an operator family $(O_\lambda)_{0<\lambda<+\infty}$ is easily obtained, since $O_{\lambda/2} \circ O_{\lambda/2} = O_\lambda$. The same way, the n-root is $O_{\lambda/n}$. The notation $R^{1/2}$ or $R^{1/n}$ can consequently be used if the operator R belongs to such a family.

The first difference between $R \circ N \circ R$ and $N \circ R \circ N$ is the relative (and absolute) importance of R and N. If N stands for a convolution with a Gaussian of a given standard deviation, one feels that the second possibility is an operator that smooths twice too much compared to the action of N alone, and compared to R which appears only once. The solution simply consists in considering $R^{1/2} \circ N \circ R^{1/2}$ and $N^{1/2} \circ R \circ N^{1/2}$ instead.

Now, the difference between $R^{1/2} \circ N \circ R^{1/2}$ and $N^{1/2} \circ R \circ N^{1/2}$ is the order of the successive application of fragments of N and R. Should one rather smooth a little, rigidify, and smooth a little again, or rather rigidify a little, smooth, and rigidify a little again ?

7.6.3 Extension towards homogeneity

If there is no particular reason to use N or R first and last, then it may make sense to consider the sequences

$$A_n = R^{1/n} N^{1/n} R^{1/n} N^{1/n} ... N^{1/n} R^{1/n}$$

(with n occurrences of $R^{1/n}$ and $n-1$ of $N^{1/n}$) and

$$A'_n = N^{1/n} R^{1/n} N^{1/n} R^{1/n} ... R^{1/n} N^{1/n},$$

or, in order to keep constant and equal to one the total weight of each operator, as said previously,

$$B_n = R^{1/n} N^{1/(n-1)} R^{1/n} N^{1/(n-1)} ... N^{1/(n-1)} R^{1/n}$$

and

$$B'_n = N^{1/n} R^{1/(n-1)} N^{1/n} R^{1/(n-1)} ... R^{1/(n-1)} N^{1/n}.$$

The hope is that these four sequences of symmetric positive definite operators may converge together and towards the same asymptotic operator. A first remark is that, as $R^{1/n}$ and $N^{1/n}$ converge towards the identity operator, the sequences A_n and A'_n, if they converge, converge towards the same operator, since $N^{1/n} A_n = A'_n R^{1/n}$. The sequences B_n and B'_n are more difficult to deal with, and, more generally, it is not easy to state any convergence result, all the more in such non-countable infinite dimensioned spaces if the operators do not commute or are not simultaneously diagonalizable.

However, this case can be solved with some inspiration from the basis of the spectral theory and semigroups. If the norm of the operator $R - Id$ is strictly less than 1, that is to say if there exists a constant $C < 1$ such that for any deformation field $\mathbf{v}$, $\|R\mathbf{v} - \mathbf{v}\|_{\mathcal{X}} \leqslant C\|\mathbf{v}\|_{\mathcal{X}}$, then the logarithm of the operator R can be defined by using the series expansion of the analytic logarithm function. Indeed, for any field $\mathbf{v}$, since the series $-\sum_{i=1}^{\infty} \frac{x^n}{n}$ converges towards $\ln(1-x)$ if $|x| < 1$, then the series $-\sum_{i=1}^{\infty} \frac{1}{n}(Id - R)^n(\mathbf{v})$, where the notation $(Id - R)^n$ stands for the composition of the operator $(Id - R)$ n times, converges for any $\mathbf{v}$ towards a field named $\ln(R)(\mathbf{v})$. The convergence comes from the inequality $\|(Id - R)^n(\mathbf{v})\|_{\mathcal{X}} \leqslant C^n \|\mathbf{v}\|_{\mathcal{X}}$. The application $\ln(R)$ thus defined for any $\mathbf{v}$ is a linear operator on $\mathcal{X}$. Moreover, as for any $\mathbf{v}$, $\|\ln(R)(\mathbf{v})\| \leqslant |\ln(C)|\ \|v\|_{\mathcal{X}}$, the operator $\ln(R)$ is bounded with norm $|||\ln(R)||| \leqslant |\ln(|||R - Id|||)|$.

Symmetrically, the exponentiation of a (bounded) linear operator can be defined with the serer $\sum_{i=0}^{\infty} \frac{x^n}{n!}$, and it satisfies $\exp(\ln(R)) = R$. The interest of these operators is that

they also satisfy $\exp\left(\frac{1}{n}\ln(R)\right) = R^{1/n}$, which brings us back to the studied sequences of operators. The sequence A_n writes:

$$A_n = R^{1/n}N^{1/n}R^{1/n}N^{1/n}...N^{1/n}R^{1/n} = \left(R^{1/n}N^{1/n}\right)^{n-1}R^{1/n}$$

Since $R^{1/n}N^{1/n} = \exp\left(\frac{1}{n}\ln(R)\right)\ \exp\left(\frac{1}{n}\ln(N)\right)$, it could be useful to consider the properties of the exponentiation. Unluckily, the exponentials of two operators that do no commute do not commute themselves, and more generally $\exp(R+N)$ is not equal to $\exp(R)\ \exp(N)$. However, as, for high values of n, $R^{1/n}$ is near the identity operator, and $\frac{1}{n}\ln(R)$ near the zero operator, the development into series of exp and ln can show itself useful. Denoting the operator $\ln(R)$ by r, we obtain:

$$\begin{aligned}
\exp\left(\frac{1}{n}r\right) &= \sum_{k=0}^{\infty}\frac{1}{k!}\left(\frac{1}{n}r\right)^k \\
&= Id + \frac{1}{n}r + \frac{1}{n^2}\left(\sum_{k=2}^{\infty}\frac{1}{k!}\frac{1}{n^{k-2}}r^k\right) \\
&= Id + \frac{1}{n}r + O\left(\frac{1}{n^2}\right)
\end{aligned}$$

and consequently:

$$\begin{aligned}
R^{1/n}N^{1/n} &= \left(Id + \frac{1}{n}\ln(R) + O\left(\frac{1}{n^2}\right)\right)\circ\left(Id + \frac{1}{n}\ln(N) + O\left(\frac{1}{n^2}\right)\right) \\
&= Id + \frac{1}{n}\ln(N) + \frac{1}{n}\ln(R)\circ\left(Id + O\left(\frac{1}{n}\right)\right) + O\left(\frac{1}{n^2}\right) \\
&= Id + \frac{1}{n}\left(\ln(R)+\ln(N)\right) + O\left(\frac{1}{n^2}\right) \qquad \text{(linearity of } \ln(R)\text{)} \\
&= \exp\left(\frac{1}{n}\left(\ln(R)+\ln(N)\right) + O\left(\frac{1}{n^2}\right)\right) \qquad \text{(for a different } O\text{).}
\end{aligned}$$

Hence:

$$\begin{aligned}
\left(R^{1/n}N^{1/n}\right)^{n-1} &= \exp\left((n-1)\left(\frac{1}{n}\left(\ln(R)+\ln(N)\right) + O\left(\frac{1}{n^2}\right)\right)\right) \\
&= \exp\left(\ln(R)+\ln(N) + O\left(\frac{1}{n}\right)\right)
\end{aligned}$$

To conclude, we need the continuity of the exponentiation.

Lemma 59. *The exponentiation of bounded operators is continuous.*

Proof. If L and M are bounded linear operators, and $\varepsilon \leqslant 1$ a small real number, then:

$$\begin{aligned}
\exp(L+\varepsilon M) - \exp(L) &= \sum_{n=0}^{\infty} \frac{1}{n!} \left((L+\varepsilon M)^n - L^n\right) \\
&= \sum_{n=1}^{\infty} \frac{1}{n!} \left(L^n + \sum_{\varepsilon M \text{appears at least once}} [\text{product of } n \text{ terms}] \; - \; L^n \right) \\
&= \varepsilon \sum_{n=1}^{\infty} \frac{1}{n!} \sum_{M \text{appears at least once}} [\text{product of } n \text{ terms}]
\end{aligned}$$

so that

$$\begin{aligned}
|||\exp(L+\varepsilon M) - \exp(L)||| &\leqslant |\varepsilon| \sum_{n=1}^{\infty} \frac{1}{n!} \sum_{M \text{appears at least once}} \max(|||L|||, \, |||M|||)^n \\
&\leqslant |\varepsilon| \sum_{n=1}^{\infty} \frac{1}{n!} 2^n \max(|||L|||, \, |||M|||)^n \\
&\leqslant |\varepsilon| \, \exp\left(2 \max(|||L|||, \, |||M|||)\right)
\end{aligned}$$

and the exponentiation of bounded operators is consequently continuous. □

The conclusion states:

$$A_n = \exp\left(\ln(R) + \ln(N) \, + \, O\left(\frac{1}{n}\right)\right) \circ \exp\left(\frac{1}{n}\ln(R)\right) \underset{n\to\infty}{\longrightarrow} \exp\left(\ln(R) + \ln(N)\right)$$

The proof of the convergence of the three other sequences (A'_n, B_n and B'_n) is similar. To sum up:

Proposition 60. *If two linear operators R and N are such that $|||R - Id||| < 1$ and $|||N - Id||| < 1$, then the previously introduced sequences A_n, A'_n, B_n and B'_n all converge towards the same limit, which is* $\exp\left(\ln(R) + \ln(N)\right)$.

Corollary 61. *If $(R_\lambda)_{0<\lambda<\infty}$ and $(N_\sigma)_{0<\sigma<\infty}$ are two semigroups of linear operators, such that $|||R_1 - Id||| < 1$ and $|||N_1 - Id||| < 1$, then $\forall \lambda, \sigma > 0$, the following sequence converges:*

$$A_n = (R_\lambda^{1/n} N_\sigma^{1/n})^{n-1} R_\lambda^{1/n} \underset{n\to\infty}{\longrightarrow} \exp\left(\lambda \ln(R) + \sigma \ln(N)\right).$$

and so do the associated sequences A'_n, B_n and B'_n.

Remark 62. *If $\bigoplus_{0<i\leqslant n} V_i$ is an orthogonal decomposition of the current vector space into a finite number of subspaces V_i, and $(\lambda_i)_{0<i\leqslant n}$ a corresponding set of n strictly positive real numbers, then the logarithm of the linear operator $R = \sum_{i=1}^{n} \lambda_i \, P_{V_i}$ (where P_W stands for the projection onto the subspace W) is*

$$\ln(R) = \sum_{i=1}^{n} \ln(\lambda_i) \, P_{V_i}$$

and it satisfies

$$\exp(t\ln(R)) = \sum_{i=1}^{n} \lambda_i^t P_{V_i}.$$

This last remark is of course dedicated to the family of operators introduced in part 7.4.1.

7.6.4 The smoothed rigidification case

Let us consider the particular case of the combination of the rigidification operator R and the Gaussian convolution N. If the initial L^2 gradient is a little too noisy, it may be relevant not only to favor rigid motion due to the rigidification operator, but also to smooth the rest of the deformation. However, the smoothing of a rigid part, such as a rotation or a scaling, would not be appropriate, since a pure rigid motion should not be altered. Therefore, instead of considering the operator $R^{1/2} \circ N \circ R^{1/2}$ or $N^{1/2} \circ R \circ N^{1/2}$, one might think of splitting the initial gradient into two parts, the rigid one and the remaining noise, and of applying the rigidification magnification process to the first part, and a smoothing to the other part. The problem is that the obtained operator is not symmetric, and this comes from the fact that the smoothing of the remaining part can make a new rigid motion arise, which is not very desirable. The solution consists in the projection of the smoothed remaining part orthogonally to the rigid motions. Let us denote by P_V the projection onto the set V of rigid motions, and P_W the projection to the orthogonal W of V, with consequently $P_V + P_W = Id$. Then the rigidification operator, introduced in part 7.4.1, writes $R = M_V \circ P_V + \alpha\, P_W$ with $0 < \alpha < 1$; the application M_V is an endomorphism on V that potentially changes the weights of the different parts V_i of the rigid motion and writes $M_V = \sum_i \lambda_i P_{V_i}$. The proposed new operator is:

$$R_N = M_V \circ P_V \ + \ P_W \circ N \circ P_W.$$

It writes more symmetrically, since M is an endomorphism on V:

$$\begin{aligned} R_N &= P_V \circ M_V \circ P_V \ + \ P_W \circ N \circ P_W \\ &= M_V \otimes N|_W \ . \end{aligned}$$

M_V and the restriction $N|_W$ of N to W act independently as two endomorphisms on the two subspaces V and W, which are orthogonal complements.

Lemma 63. *Let V and W be two orthogonal complements in a given linear space $\mathcal{X}$, and M_V and M_W two symmetric positive definite endomorphisms of V and W, respectively. Then the linear operator*

$$M_V \otimes M_W \ : \ \mathbf{x} \in \mathcal{X} \ \mapsto \ M_V(P_V(\mathbf{x})) \ + \ M_W(P_W(\mathbf{x}))$$

is a symmetric positive definite operator on $\mathcal{X}$.

The proof is straightforward but it implies that:

Corollary 64. *The previously introduced rigidification and smoothing operator*

$$R_N = M_V \circ P_V \ + \ P_W \circ N \circ P_W$$

is symmetric positive definite and can consequently be used to help the gradient descent process.

Chapter 8

Extended Gradient: more General Priors

Abstract

This chapter is dedicated to an extension of the definition of the gradient towards more general priors. The gradient is then not related to an inner product structure but to an energy that can be seen as a prior on the cost of deformation fields. An example is shown for an energy that favors locally rigid motions.

8.1. The meaning of the gradient

In this section, we go further and consider the definition of the gradient of an energy from a new point of view, which leads us to a larger class of minimization algorithms. The thread we follow is the fact that the gradient of the energy can be obtained as the result of another minimization problem.

To help in developing the reader's intuition let us recall that the usual gradient descent method can be seen, up to first order, as minimizing $E(\Gamma+u)$ with respect to the deformation field u through the linearization of the energy E in the neighborhood of the shape Γ:

$$E(\Gamma + u) \simeq E(\Gamma) + \delta E(\Gamma, u)$$

But since $\delta E(\Gamma, u)$ is linear with respect to the deformation field u, there is no minimum. This is of course a direct consequence of the first-order approximation. It is therefore more sensible to speak in terms of the direction of the deformation field u. The notion of direction implies the choice of a norm: the set of all directions is the set of all fields with norm equal to 1. Once a norm F has been chosen (related to an inner product preferably), a natural solution appears as the direction u_F that minimizes the energy $\delta E(\Gamma, v)$:

$$u_F = \underset{\{v \text{ s.t. } \|v\|_F=1\}}{\arg\min} [\ \delta E(\Gamma, v)\ ] \ = \ -\frac{\nabla_F E(\Gamma)}{\|\nabla_F E(\Gamma)\|_F} \tag{8.1}$$

The main point here is that the opposite of the gradient $-\nabla_F E(\Gamma)$ of the energy E for the inner product related to the norm F is precisely in the direction u_F. This gradient has been introduced previously as the deformation field linked to the continuous linear form $\delta E(\Gamma, \cdot)$ for the inner product F thanks to the Riesz theorem. Note that the influence of

the inner product F upon the best direction u_F lies in the fact that it changes the shape of the unit sphere (the set of all directions u with unit norm $\|u\|_F = 1$).

It turns out that the gradient itself (not only its direction) can be obtained as the solution of a minimization problem. This also explicits the link between the norm F and the gradient. This is shown in the following

Theorem 65. *The gradient $\nabla_F E(\Gamma)$ for the inner product F satisfies:*

$$-\nabla_F E(\Gamma) = \arg\min_v \left[\delta E(\Gamma, v) \; + \; \frac{1}{2}\|v\|_F^2 \right]$$

Proof. We have indeed, for any v:

$$\begin{aligned} \delta E(\Gamma, v) \; + \; \frac{1}{2}\|v\|_F^2 &= \frac{1}{2}\left[\|v\|_F^2 + 2\,\langle v \,|\nabla_F E(\Gamma)\rangle_F\,\right] \\ &= \frac{1}{2}\left[\|v + \nabla_F E(\Gamma)\|_F^2 - \|\nabla_F E(\Gamma)\|_F^2\right] \end{aligned}$$

So that:

$$\begin{aligned} \arg\min_v \left[\delta E(\Gamma, v) \; + \; \frac{1}{2}\|v\|_F^2 \right] &= \arg\min_v \left[\|v + \nabla_F E(\Gamma)\|_F^2\right] \\ &= -\nabla_F E(\Gamma). \end{aligned}$$

□

The expression between brackets breaks up into two parts: the first one, $\delta E(\Gamma, v)$, comes from the energy $E(\Gamma)$ and stands for the quantity to minimize, whereas the second one, $R(v) = \frac{1}{2}\|v\|_F^2$, is a regularizing term which imposes to the solution to be smooth and small enough in the sense of the norm F. Different choices of the smoothing term due to different choices of the norm F imply different final gradients $\nabla_F E(\Gamma)$.

For example, the choice of the H^1 inner product leads to the regularizing term $R(v) = \frac{1}{2}\|v\|_{L^2}^2 + \frac{1}{2}l^2\|Dv\|_{L^2}^2$ and consequently the gradient $\nabla_{H^1} E(\Gamma)$ is the deformation field which minimizes $\delta E(\Gamma, v) \; + \; \frac{1}{2}\|v\|_{L^2}^2 \; + \; \frac{1}{2}l^2\|Dv\|_{L^2}^2$. This leads us to an elegant proof of a property of the H^1 gradient stated in section 7.4.2, without considering PDEs:

Proposition 66. *The opposite of the H^1 gradient is the solution of:*

$$\arg\min_v \left[\|u - v\|_{L^2}^2 + l^2\|Dv\|_{L^2}^2\right]$$

where $u = -\nabla_{L^2} E(\Gamma)$ is the opposite of the usual gradient.

Proof. Indeed, for any v:

$$\|u - v\|_{L^2}^2 = \|u\|_{L^2}^2 - 2\,\langle u \,|v\rangle_{L^2} + \|v\|_{L^2}^2$$

hence

$$\|u - v\|_{L^2}^2 + l^2\|Dv\|_{L^2}^2 = \|u\|_{L^2}^2 + 2\,\delta E(\Gamma, v) + \|v\|_{H^1}^2$$

since by definition of u, $\delta E(\Gamma, v) = \langle -u \,|v\rangle_{L^2}$; so the H^1-gradient can naturally be seen as a smoothed version of the standard gradient u, thanks to theorem 65:

$$\arg\min_v \left[\|u - v\|_{L^2}^2 + l^2\|Dv\|_{L^2}^2\right] = \arg\min_v \left[\delta E(\Gamma, v) + \frac{1}{2}\|v\|_{H^1}^2\right] = -\nabla_{H^1} E(\Gamma)$$

□

8.2. Generalization of the regularizing term

We have stressed the influence of the choice of an inner product $\langle | \rangle_F$ on the associated gradient:

$$-\nabla_F E(\Gamma) = \arg\min_v [\ \delta E(\Gamma, v)\ +\ R(v)\ ]$$

where $R(v) = \frac{1}{2}\|v\|_F^2$, and $\|\cdot\|_F$ is the norm related to the chosen inner product. Since the choice of the inner product is equivalent to the choice of the regularizing term $R(v)$ and acts qualitatively upon the gradient descent paths, we can see $R(v)$ as a prior on the deformation fields.

Let us now generalize our framework and allow $R(v)$ to be (almost) any positive real function, not necessarily related to an inner product, and compute (when it exists) the associated field which we will denote, with a slight abuse of notation, by $-\nabla_R E(\Gamma)$ (note that if R is related as previously to the inner product F, then $\nabla_F E = \nabla_R E$):

$$-\nabla_R E(\Gamma) = \arg\min_v [\ \delta E(\Gamma, v)\ +\ R(v)\ ] \tag{8.2}$$

Under some reasonable assumptions about $R(v)$, the new "gradient" $\nabla_R E(\Gamma)$ exists and has interesting properties.

First, the existence of $\nabla_R E(\Gamma)$, which is the solution of an infimum problem, is guaranteed if R is positive, superlinear and convex. The solution is then not necessarily unique; nevertheless, the set of solutions is convex, reduced to a single point in most cases. However, the question of the existence and unicity of ∇_R in general is not the main point here, it depends on the particular chosen application R. Here, R is supposed to stand for an application approximatively "looking like" the square of a norm; for reasonable choices of R from this point of view, the existence is guaranteed, and so is the uniqueness in most cases.

We now present the fundamental property of the extended gradient $\nabla_R E$ as the

Theorem 67. *If R is differentiable and reaches its global minimum at the zero field, then the flow $-\nabla_R E(\Gamma)$, if it exists, decreases the energy E.*

Proof. We prove that $\delta E(\Gamma, -\nabla_R E(\Gamma)) \leqslant 0$.

We have $-\nabla_R E(\Gamma) = \arg\min_v [\ \delta E(\Gamma, v)\ +\ R(v)\ ]$, so, in particular, considering the zero field $v = 0$:

$$\delta E(\Gamma, -\nabla_R E(\Gamma))\ +\ R(-\nabla_R E(\Gamma))\ \leqslant\ \delta E(\Gamma, 0)\ +\ R(0)$$

$$\delta E(\Gamma, -\nabla_R E(\Gamma)) \leqslant R(0) - R(-\nabla_R E(\Gamma))$$

As $v = 0$ is the global minimum of R, we have $R(-\nabla_R E(\Gamma)) \geqslant R(0)$, so:

$$\delta E(\Gamma, -\nabla_R E(\Gamma)) \leqslant 0$$

Moreover, this last inequality is strict if the usual gradient $\nabla_{L^2} E(\Gamma)$ is not zero. Indeed, in that case, as $R(v)$ reaches its global minimum at the zero field $v = 0$, its derivative is zero for $v = 0$. Consequently, the L^2 gradient of $\delta E(\Gamma, v) + R(v)$ with respect to v at the zero field equals $\nabla_{L^2} E(\Gamma)$, which is not zero by hypothesis. Therefore $\inf_v [\ \delta E(\Gamma, v)\ +\ R(v)\ ]$ is not reached at $v = 0$ and all inequalities in the proof are strict. □

Note that the application R is specific to the shape Γ (or, more exactly, to its tangent space) and there is no assumption about a "regularity" of the applications R_Γ with respect to Γ. However, as in the previous part where we had to associate to each shape an inner product and naturally chose the same general expression for all of them, we will restrict ourselves to the case where the application R_Γ has the same general expression $R(\Gamma)$ for all shapes Γ and consequently will commit a slight abuse of notation between R and R_Γ.

8.3. Remarks

8.3.1 Addition of an orthogonal term

Note that the method proposed in section 7.3.3, which consists in adding an orthogonal term to the gradient (see equation (7.7)), can be seen as a variation on the extended gradient theme, where the search for the infimum has been restricted to the affine hyperplane $\mathcal{H}$ containing the opposite of the gradient $-\nabla_F E$ and orthogonal to it. Indeed:

$$\begin{aligned}\arg\min_{v\in\mathcal{H}} \left[\delta E(\Gamma, v) + R(v)\right] &= \arg\min_{w; w\perp\nabla_F E} \left[\delta E(\Gamma, -\nabla_F E + w) + R(-\nabla_F E + w)\right] \\ &= \arg\min_{w; w\perp\nabla_F E} R(-\nabla_F E + w)\end{aligned}$$

8.3.2 Directional formulation

We have seen earlier (equation (8.1)) that the direction of the gradient could be defined as the field v of the unit sphere $\mathcal{U}_F = \{v \text{ s.t. } \|v\|_F = 1\}$ which most decreases the energy, and that changing the inner product F was precisely acting on the gradient by changing the unit sphere. One way to generalize the notion of gradient could have been to set any hypersurface $\mathcal{S}$ instead of the unit sphere $\mathcal{U}_F$ and to search for the best field v in $\mathcal{S}$. However, this would lead to some difficulties in practice (how to search for a minimum on an hypersurface of infinite dimension, how to represent this hypersurface?). A slightly better way to do this would be to focus on the hypersurfaces of the form $\mathcal{U}_R = \{v \text{ s.t. } R(v) = 1\}$, which is in the spirit of the level-set method. Note that this approach would be very close in practice to the one we described, the main difference being that we would only obtain a direction, without the magnitude.

8.3.3 Temporal coherence

The application $R(v)$ does not necessarily only deal with spatial coherence and can also be designed to favor temporally coherent motions. For example, at time step t of an evolution, one could force the new deformation field u_t to resemble the previous one u_{t-1}. If we transport u_{t-1} from the previous shape Γ_{t-1} to the new one Γ_t, we obtain a new field noted $T(u_{t-1})$ defined on the same space as u_t, and we can compare them, e.g., with $\|T(u_{t-1}) - u_t\|$. We are thus led to define $R(v) = \|T(u_{t-1}) - v\|$. This function however does not satisfy in general the condition $R(0) = 0$ which is necessary in theorem 67. Nevertheless this problem can be solved by defining $R(v)$ to be the norm of the projection of v orthogonally to $T(u_{t-1})$.

8.4. Computing the extended gradient

If R is simple enough so that the inverse application of $v \mapsto \nabla_{L^2} R(v)$ is easily computable, then the computation of the extended gradient $\nabla_R E$ is immediate from the knowledge of

the usual L^2-gradient. Indeed, the application $v \mapsto \delta E(\Gamma, v) + R(v)$ has a local minimum at $v = -\nabla_R E(\Gamma)$, so its derivative with respect to v is zero at this point:

$$D_v \left(\delta E(\Gamma, v) + R(v)\right)\big|_{v=-\nabla_R E} = 0$$

$$\nabla_{L^2} \left(\langle \nabla_{L^2} E(\Gamma) \,|v\rangle_{L^2} + R(v)\right)\big|_{v=-\nabla_R E} = 0$$

$$\nabla_{L^2} E(\Gamma) + \nabla_{L^2} R\big(-\nabla_R E(\Gamma)\big) = 0$$

$$-\nabla_R E(\Gamma) = (\nabla_{L^2} R)^{-1} \big(-\nabla_{L^2} E(\Gamma)\big)$$

This formula generalizes the one obtained previously in proposition 57 in section 7.3 concerning the relation between the gradient for an inner product P and the usual L^2 gradient. Now, for the extended gradient, the application $(\nabla_{L^2} R)^{-1}$ which stands in for L in this proposition is not supposed to be linear anymore.

In more general cases, if we cannot compute the application $(\nabla_{L^2} R)^{-1}$, we can still solve the infimum problem with a Partial Differential Equation (PDE) which is equivalent to a ... infinitesimal gradient descent! The definition in equation (8.2) can be seen indeed as a minimization problem which leads to the evolution:

$$\text{(8.3)} \qquad \begin{cases} v(0) = 0 \\ \dfrac{dv}{dt} = -\nabla_{L^2} E(\Gamma) - \nabla_{L^2} R(v) \end{cases}$$

This evolution leads to a local minimum of $\delta E(\Gamma, v) + R(v)$. Even if this local minimum is not the global one (if R has not been well-chosen) or if the evolution is stopped before the convergence, the final flow v computed will strictly decrease the energy $E(\Gamma)$ (same proof as in theorem 67). This point may be important in practice. Note also that there exist many other methods [9] than the usual gradient descent to solve that kind of problem, since the quantity to minimize is a sum of a linear term $\delta E(\Gamma, v)$ and another term $R(v)$ which "looks like" a quadratic term since it is supposed to play a role similar to the square of a norm.

8.5. Application: the semi-local rigidification

We now present an example for which the previous framework appears to be useful. We consider an energy $E(\Gamma)$ defined on plane curves. These curves are assumed to lie in an image Ω, in fact a bounded subset of $\mathbb{R}^2$. Instead of following a usual gradient descent in order to minimize $E(\Gamma)$ with respect to Γ, we would like to favor the deformation fields which preserve the rigidity of the shape as much as possible, or, more exactly, we would like to favor more rigid deformations, so that some kinds of local minima could be avoided. In section 7.4.1 we showed how to change the inner product so as to favor global rigid transformations. In case of articulated movement, this global method may not be sufficient, so we would like to favor fields containing parts close to rigid motions; this leads us to the notion of "semi-local rigidification". We use the expression "semi-local" in order to emphasize the contrast with usual smoothing methods such as Gaussian smoothing or H^1 smoothing, which we will qualify of "local".

Let us consider a shape Γ and any field v defined on it. We would like to find the parts, if any, of the field v which are well approximated by a translation or a rotation acting on the corresponding parts of Γ. In order to model this, we associate to each point $\mathbf{x}$ of Γ a rigid deformation $w_{\mathbf{x}}$ defined on the whole image Ω. In order to describe $w_{\mathbf{x}}$ we introduce three

functions defined on Γ: a translation $T(\mathbf{x})$, a center of rotation $C(\mathbf{x})$ and the magnitude $A(\mathbf{x})$ of the instantaneous rotation, so that:

$$\forall \mathbf{y} \in \Omega, w_{\mathbf{x}}(\mathbf{y}) = A(\mathbf{x})\,(\mathbf{y} - C(\mathbf{x}))^{\perp} + T(\mathbf{x})$$

where $a^{\perp}$ stands for the vector a rotated by $+\pi/2$. We suppose that this rigid deformation $w_{\mathbf{x}}$ varies slowly with respect to $\mathbf{x}$, that is to say we suppose the $L^2(\Omega, \mathbb{R}^2)$ norm of its derivative $\|D_{\mathbf{x}} w_{\mathbf{x}}(\cdot)\|_{L^2}$ to be small for each point $\mathbf{x}$ of the curve Γ. We consider the $L^2(\Gamma, \mathbb{R})$ norm of this application defined on Γ and obtain the criterion $\left\| \|D_{\mathbf{x}} w_{\mathbf{x}}(\cdot)\|_{L^2(\Omega,\mathbb{R}^2)} \right\|_{L^2(\Gamma,\mathbb{R})}$ to quantify the smoothness of the field $w_{\mathbf{x}}$ of rigid deformations on Γ.

It is always possible to express any field v on Γ as a member of the class rigid motions:

$$\forall \mathbf{x} \in \Gamma, v(\mathbf{x}) = w_{\mathbf{x}}(\mathbf{x}) = A(\mathbf{x})(\mathbf{x} - C(\mathbf{x}))^{\perp} + T(\mathbf{x}) \tag{8.4}$$

The field v is then completely defined by the knowledge of T, A and C. For a given field v, there exist of course many triplets (T, A, C) satisfying (8.4), the simplest one being $(v, 0, G)$, where G_Ω is the center of mass of the image Ω. In order to lift this ambiguity we define a deformation prior R which depends on T, A and C that should be seen as parameters of v:

$$R(T, A, C) = \|v\|_{L^2}^2 + \left\| \|D_{\mathbf{x}} w_{\mathbf{x}}(\cdot)\|_{L^2(\Omega,\mathbb{R}^2)} \right\|_{L^2}^2$$

which in fact can also be written simpler (by expanding and integrating the expression $\|D_{\mathbf{x}} w_{\mathbf{x}}(\mathbf{y})\|^2$) as:

$$R(T, A, C) = \|v\|_{L^2}^2 + \left\| DT + DA\,(G_\Omega - C)^{\perp} - A\,DC^{\perp} \right\|_{L^2}^2 + \sigma_\Omega^2 \|DA\|_{L^2}$$

where $\sigma_\Omega^2 = \int_\Omega (\mathbf{y} - G_\Omega)^2 d\mathbf{y}$ is a characteristic squared "length" of the image. The middle term represents the interaction between T, A and C; for example, changing the center of rotation $DC(\mathbf{x})$ has no consequence on the rigid motion $w_{\mathbf{x}}$ if it is compensated by the adequate added translation $DT = A\,DC^{\perp}$. Note that the quantities G_Ω and σ_Ω are the only ones where the influence of the image Ω appears.

In order to compute the generalized gradient ∇_R of an energy $E(\Gamma)$, we first compute the usual L^2 gradient $\nabla_{L^2} E$, initialize $(T, A, C) = (0, 0, G_\Omega)$ so that the corresponding field $v(T, A, C)$ is zero, as required in equation (8.3), and let (T, A, C) evolve to minimize $\delta E(\Gamma, v) + R(T, A, C)$. The corresponding PDEs are

$$\begin{cases} \partial_t T(\mathbf{x}) &= -\big(\nabla_{L^2} E(\Gamma)(\mathbf{x}) + v(\mathbf{x})\big) \;+\; \Delta\overline{w_{\mathbf{x}}} \\ \partial_t A(\mathbf{x}) &= -\big(\nabla_{L^2} E(\Gamma)(\mathbf{x}) + v(\mathbf{x})\big) \cdot \big(\mathbf{x} - C(\mathbf{x})\big)^{\perp} \\ &\qquad\qquad + \;\big(G_\Omega - C(\mathbf{x})\big)^{\perp} \cdot \Delta\overline{w_{\mathbf{x}}} \;+\; \sigma_\Omega^2 \Delta A(\mathbf{x}) \\ \partial_t C(\mathbf{x}) &= -A(\mathbf{x})\big(\nabla_{L^2} E(\Gamma)(\mathbf{x}) + v(\mathbf{x})\big)^{\perp} \;+\; A(\mathbf{x})\Delta\overline{w_{\mathbf{x}}}^{\perp} \end{cases}$$

where $\overline{w_{\mathbf{x}}} = w_{\mathbf{x}}(G_\Omega)$ is the mean of the linear application $\mathbf{y} \mapsto w_{\mathbf{x}}(\mathbf{y})$ on Ω. Note that if we had considered only translations T (and not rotations), we would have $\overline{w_{\mathbf{x}}} = T(\mathbf{x}) = v(\mathbf{x})$ and the algorithms would act as an H^1 smoothing.

8.6. Numerical Example

We now apply this method to a specific choice of the energy E to minimize. We would like to warp a given initial shape Γ_1 onto a given target shape Γ_2, that is to say, we would like

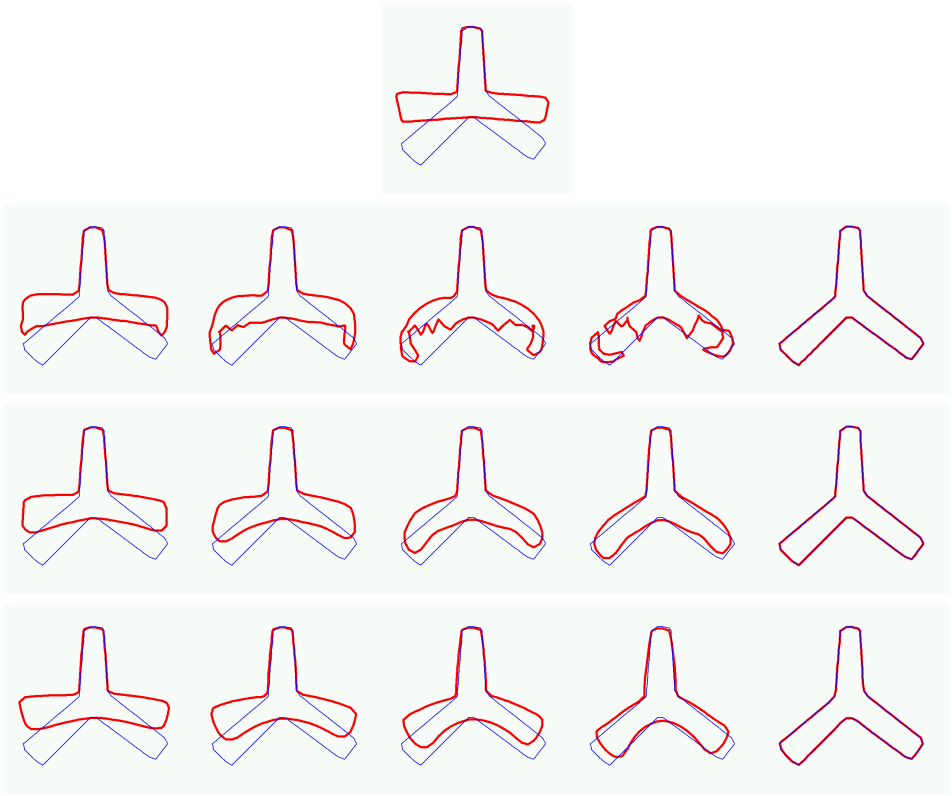


Figure 8.1: Warping the red shape onto the blue one *(top)* with the L^2 gradient descent *(first row)*, with a H^1 gradient descent *(second row)* and with a modified gradient descent favoring semi-local rigid motion *(third row)* for the same energy (Hausdorff distance). All evolutions converge to the same shape, but with different paths.

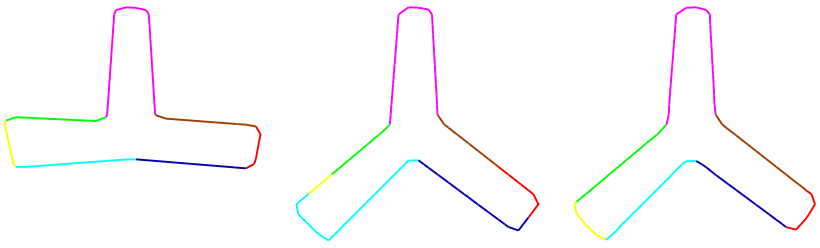


Figure 8.2: Comparison of the correspondences between the initial curve *(left)* and two final curves resulting from the H^1 gradient evolution *(middle)* and from the semi-local rigidification *(right)*. The different parts of the curves are shown with different colors, so that their respective evolutions can be followed. The correspondences for the semi-local rigidification case are more geometrically meaningful.

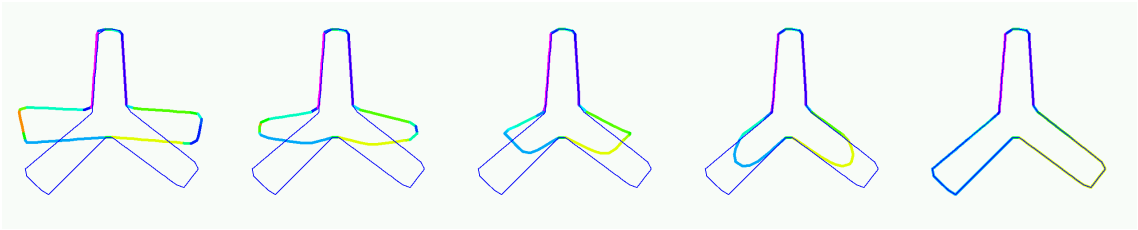


Figure 8.3: Comparison with a L^2 gradient descent on the L^2 norm of the signed distance functions associated to the curves. The gradient is naturally smooth but lacks geometric information.

to minimize the shape distance between Γ_1 and Γ_2 with respect to Γ_1. We choose for E the smooth approximation of the Hausdorff distance described in chapter 2 which we will denote here by $d_H(\Gamma_1, \Gamma_2)$.

This energy E achieves generally good warping between any two shapes which are relatively close one to the other, but, in case of large deformations, it can suffer from an important lack of spatial coherence if a part A of the moving shape Γ_1 has to cross a part B of the target one on its way to another parallel part C of the target shape (see figure 8.1 for an example), because the part A tries to minimize its distance to both parts B and C at the same time.

A global coherence can nonetheless be recovered by an adequate change of inner product which favors rigid transformations, as presented before. However, this is not sufficient for dealing with local deformations. The methods of Gaussian or H^1 smoothing studied in sections 7.4.3 and 7.4.2 could be helpful, since their action is local. But even if their influence is appreciable, these smoothing techniques do not favor semi-locally rigid deformations like the movements of an articulated object. We have noticed that, in practice, the quality of the matching between two shapes Γ_1 and Γ_2 generally depends on the quality of the path that has been followed during the evolution from one shape to the other, or, more precisely, on how natural a human observer would judge this path. This statement is very intuitive and qualitative but we believe that this quality precisely relies on notions such as articulated motion. There is clearly room here for further work. In any case this is the reason why we think that methods like the ones proposed in this chapter, which allow to set priors on the deformation fields, can have interesting practical applications.

We use the framework presented above and compare the evolutions resulting from three different approaches on a difficult example of shape warping in the case of the Hausdorff distance: the usual L^2 gradient method, the H^1 smoothing method of section 7.4.2 (for the best value of the smoothness parameter l in equation (7.10)) and the semi-local rigidification method (Figure 8.1). The last one achieves the best path and the best correspondences (Figure 8.2).

The gradient descent framework in the case of an extended gradient ∇_R could have needed some important additional time if we had to wait until the evolution of ∇_R converges at each time step of the global evolution of Γ_1. Fortunately, when necessary, thanks to the remark in section 8.4, we can choose to stop the evolution of ∇_R before convergence in order to keep the additional cost into reasonable limits. The result presented here was computed so that the total evolution time was multiplied by two, but the effect of the semi-local rigidification is already noticeable for an added cost time of 10%. The minimization

of the particular proposed criterion R appears to be difficult in practice due to some unexpected (and not understood) unstabilities of the rotation center $C(\mathbf{x})$. Other minimization techniques, like graph cuts, might be investigated to compute ∇_R. However other criteria could be proposed and the main point here is that such priors on the deformation flows can be taken into account.

For the particular example presented in figure 8.1, one could object that we should have considered other distances, such as the L^2 norm between the signed distance functions of the shapes, which always leads to very smooth evolutions. However, those smooth evolutions are not very sensible, in that they often miss entirely the similarity between the two shapes to match (see figure 8.3). As a consequence their gradient does not contain a lot of geometric information and cannot be very much improved by changes of inner products. This is why, despite the sometimes irregular behavior of the gradient of the Hausdorff distance, we prefer to use it in combination with new inner products, because this has both advantages of guaranteeing smoothness and making geometric sense.

In figure 8.4 we show an example with real contours from hand segmentation of pictures. As in the previous example, we show the evolution path obtained by minimization of the approximation of the Hausdorff distance between the two curves, with the semi-local rigidification approach. The evolution mainly consists in four local rotations (arms and legs), which fits well our intuition. We have colored, as previously, different parts of the initial curve in order to follow them through the evolution and notice how relevant the correspondences are. A usual gradient descent for this energy would have faced the same irregularity problems as in the first evolution presented in figure 8.1, and the choice of other usual energies, like the L^2 norm between the signed distance functions to the curves, would lack geometric sense, as in figure 8.3.

8.7. Conclusion

The impact of the inner product structure of the deformation space on the behavior of the active contours method had been overlooked so far in the computer vision community. We have explored several families of inner products, as well as some minimizing flows not deriving from any inner product by extending the notion of gradient. Given an energy, we now have several ways to minimize it, each of the proposed flows being a minimizing flow but leading to different kinds of evolutions. The inner products and the extended gradients should consequently be seen as priors on the deformation fields, that is to say priors on the evolution paths. They can be used to introduce different degrees of spatial coherence (local, semi-local or global) in the evolution of the contour.

We have shown, with some numerical experiments, that these evolutions better fit our intuitive notion of deformation cost and that they can mimic the behavior of the objects of interest. As a result, they are at the same time more meaningful and more robust to irrelevant local minima attraction.

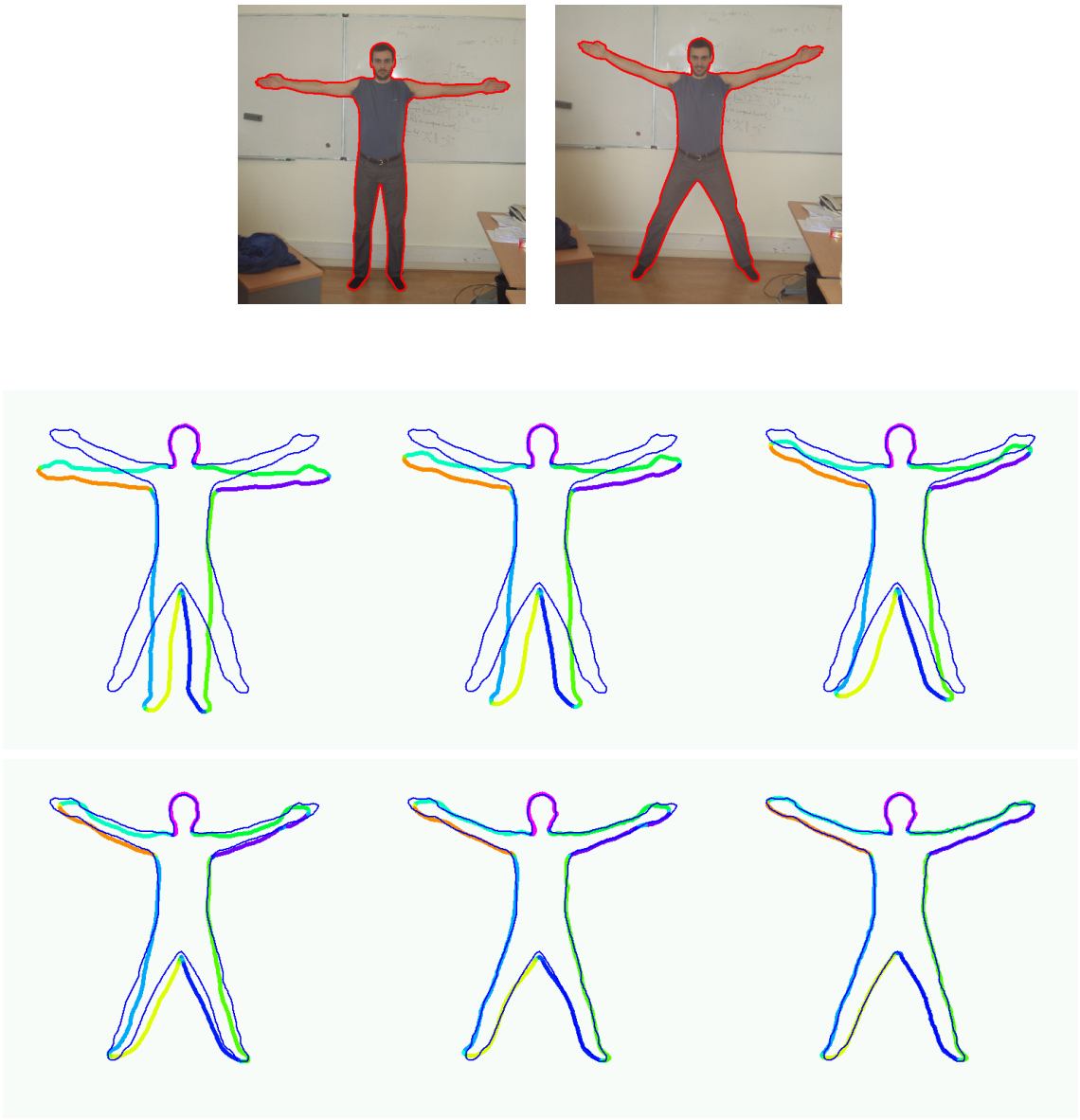


Figure 8.4: Warping real contours by minimization of the approximation of the Hausdorff distance with the semi-local rigidification approach. The colors show the correspondences between the moving curve and the initial one.

Part III

Shape Statistics and Priors

Chapter 9

Shape Statistics: Empirical Mean and Modes of Variation

Abstract

This chapter is dedicated to the computation of the empirical mean and covariance of a set of shape examples. We now use the previously developed tools to define several concepts relevant to a theory of stochastic shapes as well as to provide the means for their effective computation. They are based on the use of the function E defined by (5.12) which is the smooth approximation of a distance.

9.1 Empirical mean

The first task is to define and compute the mean of a set of shapes. Inspired by the work of Fréchet [46, 47], Karcher [63], Kendall [67], and Pennec [96], we provide the following (classical)

Definition 68. *Given $\Gamma_1, \cdots, \Gamma_N$, N shapes, we define their empirical mean as any shape $\hat{\Gamma}$ that achieves a local minimum of the function $\mu : \mathcal{S} \to \mathbb{R}^+$ defined by*

$$\Gamma \to \mu(\Gamma, \Gamma_1, \cdots, \Gamma_N) = \frac{1}{N} \sum_{i=1,\cdots,N} E^2(\Gamma, \Gamma_i)$$

Note that there may exist several means. We know from proposition 55 that there exists at least one. An algorithm for computing approximations to an empirical mean of N shapes readily follows from the previous section: start from an initial shape Γ_0 and solve the PDE

$$\begin{aligned} \Gamma_t &= -\nabla\mu(\Gamma, \Gamma_1, \cdots, \Gamma_N)\,\mathbf{n} \\ \Gamma(0, .) &= \Gamma_0(.) \end{aligned} \tag{9.1}$$

We show some examples of means computed by this algorithm in figure 9.1.

We have not explored the problem of the number of possible local minima in great detail but observed that the following heuristics led to "visually satisfying" results.

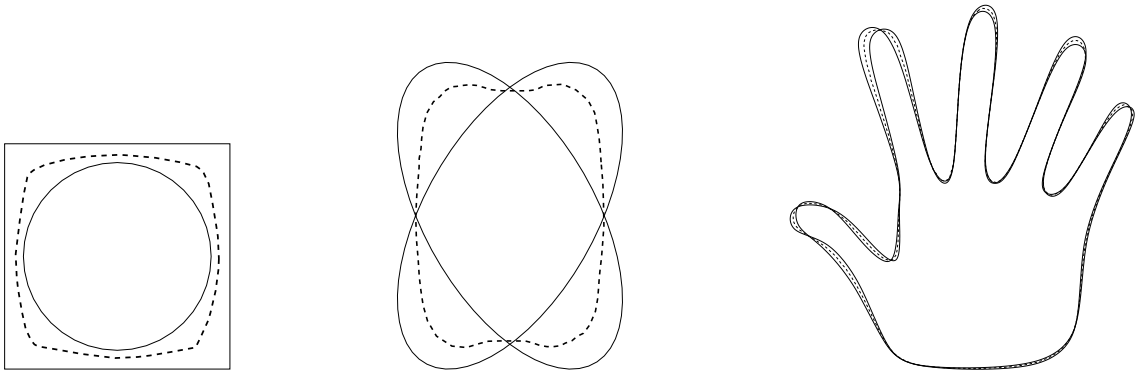


Figure 9.1: Examples of means of several curves: a square and a circle (left), two ellipses (middle), and two hands (right).

Suppose that the example shapes are given in some order, according to the way they are indexed from 1 to N. Initialize $\hat{\Gamma}^{(1)}$ to Γ_1, solve

$$\Gamma_t^{(i+1)} = -\nabla \left(E^2(\Gamma, \Gamma_{i+1}) + iE^2(\Gamma, \hat{\Gamma}^{(i)}) \right) \mathbf{n}$$
$$\Gamma^{(i+1)}(0, .) = \hat{\Gamma}^i(.),$$

and choose $\hat{\Gamma}^{(i+1)} = \Gamma^{(i+1)}$ at convergence, for $i = 1, \cdots, N-1$. Of course, there is not guarantee that either the result will be independent of the order of presentation (this may or may not be important, depending on the application) or that it will indeed be a local minimum of $\mu(\Gamma, \Gamma_1, \cdots, \Gamma_N)$. Another alternative is to solve (9.1) by choosing Γ_0 to be one of the given shapes. In all cases heuristically we have not found local minima.

We show the result of computing the mean of nine hands with this method in figure 9.2

Another example of mean is obtained from the previous fish silhouettes database: we have used eight silhouettes, normalized them so that their centers of gravity and principle axes were aligned, and computed their mean, as shown in figure 9.3. The initial curve, Γ_0 was chosen to be an enclosing circle.

9.2 Empirical covariance

We can go beyond the definition of the mean and in effect define something similar to the covariance matrix of a set of N shapes.

The function $\mathcal{S} \to \mathbb{R}^+$ defined by $\Gamma \to E^2(\Gamma, \Gamma_i)$ has a gradient which defines a normal velocity field, noted β_i, defined on Γ, such that if we consider the infinitesimal deformation $\Gamma - \beta_i \mathbf{n} d\tau$ of Γ, it decreases the value of $E^2(\Gamma, \Gamma_i)$. Each such β_i belongs to $L^2(\Gamma)$, the set of square integrable real functions defined on Γ. Each Γ_i defines such a normal velocity field β_i. We consider the mean velocity $\hat{\beta} = \frac{1}{N} \sum_{i=1}^N \beta_i$ and define the linear operator $\Lambda : L^2(\Gamma) \to L^2(\Gamma)$ such that $\beta \to \sum_{i=1,N} < \beta, \beta_i - \hat{\beta} > (\beta_i - \hat{\beta})$. We have the following

Definition 69. *Given N shapes of $\mathcal{S}$, the covariance operator of these N shapes relative to any shape Γ of $\mathcal{S}$ is the linear operator of $L^2(\Gamma)$ defined by*

$$\Lambda(\beta) = \sum_{i=1,N} < \beta, \beta_i - \hat{\beta} > (\beta_i - \hat{\beta}),$$

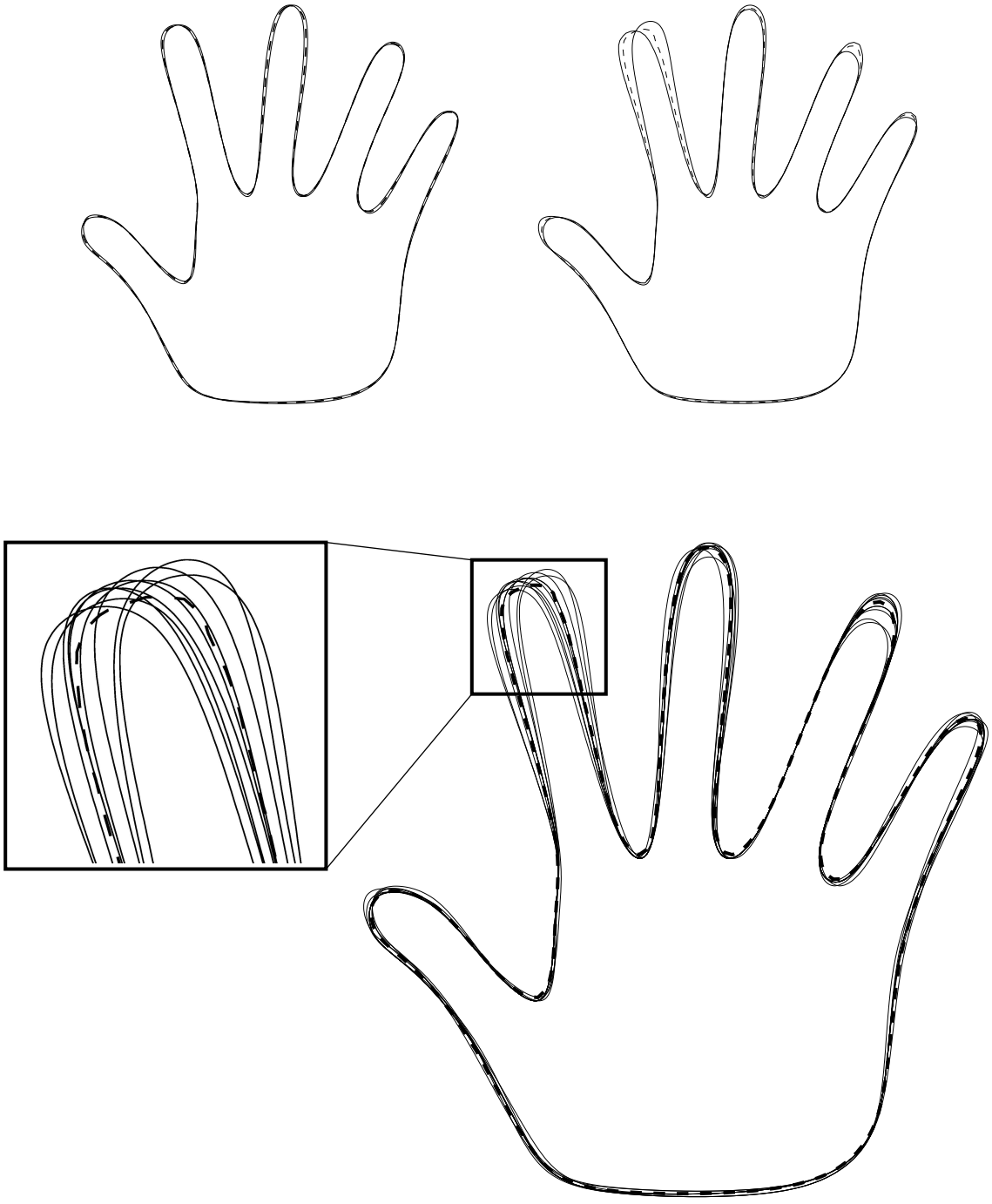


Figure 9.2: The mean (the dashed curve) of nine hand silhouettes (the continuous curves) obtained by the sequential suboptimal method described in the text: first step (mean of the two first curves), fifth step (weighted mean of the sixth curve and of the mean of the five first curves from the previous step), and final result.

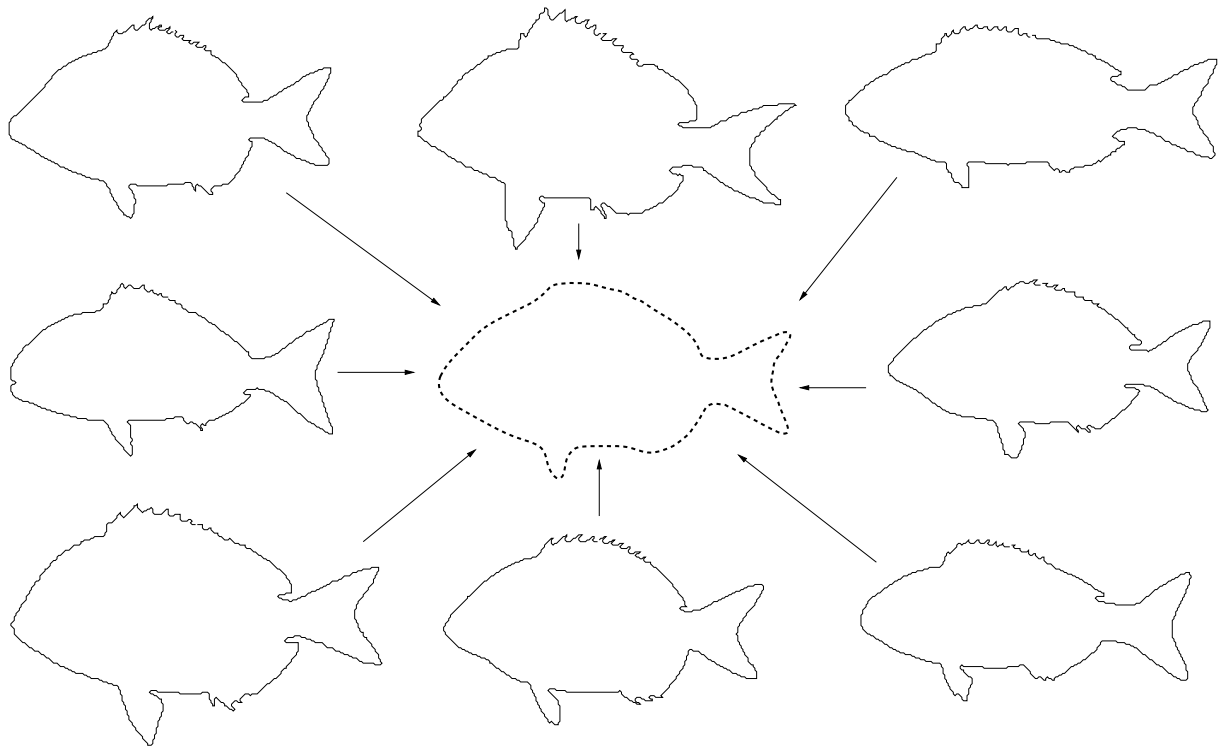


Figure 9.3: The mean of eight fishes.

where the β_i are defined as above, relatively to the shape Γ.

This operator has some interesting properties which we study next.

Proposition 70. *The operator Λ is a continuous mapping of $L^2(\Gamma)$ into $L^2(\Gamma)$.*

Proof. We have $\| \sum_{i=1,N} < \beta, \beta_i - \hat{\beta} > (\beta_i - \hat{\beta})\|_2 \leq \sum_{i=1,N} | < \beta, \beta_i - \hat{\beta} > | \|\beta_i - \hat{\beta}\|_2$ and, because of Schwarz inequality, $| < \beta, \beta_i - \hat{\beta} > | \leq \|\beta\|_2 \|\beta_i - \hat{\beta}\|_2$. This implies that $\| \sum_{i=1,N} < \beta, \beta_i - \hat{\beta} > (\beta_i - \hat{\beta})\|_2 \leq K\|\beta\|_2$ with $K = \sum_{i=1,N} \|\beta_i - \hat{\beta}\|_2^2$. □

Λ is in effect a mapping from $L^2(\Gamma)$ into its Hilbert subspace $A(\Gamma)$ generated by the N functions $\beta_i - \hat{\beta}$. Note that if Γ is one of the empirical means of the shapes Γ_i, by definition we have $\hat{\beta} = 0$.

This operator acts on what can be thought of as the tangent space to the manifold of all shapes at the point Γ. We then have the

Proposition 71. *The covariance operator is symmetric positive semi definite.*

Proof. This follows from the fact that $< \Lambda(\beta), \beta >=< \beta, \Lambda(\beta) >= \sum_{i=1,N} < \beta, \beta_i - \hat{\beta} >^2$. □

It is also instructive to look at the eigenvalues and eigenvectors of Λ. For this purpose we introduce the $N \times N$ matrix $\hat{\Lambda}$ defined by $\hat{\Lambda}_{ij} =< \beta_i - \hat{\beta}, \beta_j - \hat{\beta} >$. We have the

Proposition 72. *The $N \times N$ matrix $\hat{\Lambda}$ is symmetric positive semi definite. Let $p \leq N$ be its rank, $\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_p^2 > 0$ its positive eigenvalues, $\mathbf{u}_1, \cdots, \mathbf{u}_N$ the corresponding eigenvectors. They satisfy*

$$\begin{aligned} \mathbf{u}_i \cdot \mathbf{u}_j &= \delta_{ij} \quad i,\, j = 1, \cdots, N \\ \hat{\Lambda}\mathbf{u}_i &= \sigma_i^2 \mathbf{u}_i \quad i = 1, \cdots, p \\ \hat{\Lambda}\mathbf{u}_i &= 0 \quad p+1 \leq i \leq N \end{aligned}$$

Proof. The matrix $\hat{\Lambda}$ is clearly symmetric. Let now $\boldsymbol{\alpha} = [\alpha_1, \cdots, \alpha_N]^T$ be a vector of $\mathbb{R}^N$, $\boldsymbol{\alpha}^T \hat{\Lambda} \boldsymbol{\alpha} = \|\beta\|_2^2$, where $\beta = \sum_{i=1}^N \alpha_i(\beta_i - \hat{\beta})$. The remaining of the proposition is simply a statement of the existence of an orthonormal basis of eigenvectors for a symmetric matrix of $\mathbb{R}^N$. □

The N-dimensional vectors $\mathbf{u}_j$, $j = 1, \cdots, p$ and the p eigenvalues σ_k^2, $k = 1, \cdots, p$ define p modes of variation of the shape Γ. These modes of variation are normal deformation flows which are defined as follows. We note u_{ij}, $i,\, j = 1, \cdots, N$ the ith coordinate of the vector $\mathbf{u}_j$ and v_j the element of $A(\Gamma)$ defined by

$$v_j = \frac{1}{\sigma_j} \sum_{i=1}^N u_{ij}(\beta_i - \hat{\beta}) \tag{9.2}$$

In the case $\Gamma = \hat{\Gamma}$, $\hat{\beta} = 0$. We have the proposition

Proposition 73. *The functions v_j, $j = 1, \cdots, p$ are an orthonormal set of eigenvectors of the operator Λ and form a basis of $A(\Gamma)$.*

Proof. Let us form the product $< v_j,\, v_k >$:

$$< v_j, v_k > = \frac{1}{\sigma_j \sigma_k} < \sum_{l=1}^N u_{lj}(\beta_l - \hat{\beta}),\, \sum_{m=1}^N u_{mk}(\beta_m - \hat{\beta}) > =$$
$$\frac{1}{\sigma_j \sigma_k} \sum_{l=1}^N u_{lj} \sum_{m=1}^N < \beta_l - \hat{\beta}, \beta_m - \hat{\beta} > u_{mk} = \frac{1}{\sigma_j \sigma_k} \sum_{l=1}^N u_{lj} (\hat{\Lambda}\mathbf{u}_k)_l = \frac{1}{\sigma_j \sigma_k} \mathbf{u}_j \cdot (\hat{\Lambda}\mathbf{u}_k)$$

According to proposition 72, $\hat{\Lambda}\mathbf{u}_k = \sigma_k^2 \mathbf{u}_k$ and $\mathbf{u}_j \cdot \mathbf{u}_k = \delta_{jk}$, which proves the orthonormality and therefore the linear independence. There remains to show that they generate the whole of $A(\Gamma)$. In order to see this, we consider the element $\beta = \sum_{i=1}^N \alpha_i(\beta_i - \hat{\beta})$ of $A(\Gamma)$ and look for the coefficients μ_k, $k = 1, \cdots, p$ such that

$$\sum_{i=1}^N \alpha_i(\beta_i - \hat{\beta}) = \sum_{k=1}^p \mu_k v_k \tag{9.3}$$

We take the Hilbert product of both sides of this equation with $\beta_j - \hat{\beta}$ to obtain

$$(\hat{\Lambda}\boldsymbol{\alpha})_j = \sum_{k=1}^p \mu_k < v_k, \beta_j - \hat{\beta} > . \tag{9.4}$$

We then use (9.2), replace v_k with

$$v_k = \frac{1}{\sigma_k} \sum_{i=1}^{N} u_{ik}(\beta_i - \hat{\beta}),$$

and obtain

$$< v_k, \beta_j - \hat{\beta} >= \frac{1}{\sigma_k} \sum_{i=1}^{N} u_{ik} < \beta_i - \hat{\beta},\, \beta_j - \hat{\beta} >= \frac{1}{\sigma_k} (\hat{\Lambda}\mathbf{u}_k)_j = \sigma_k u_{kj}.$$

Replacing this value in (9.4) yields

$$(\hat{\Lambda}\boldsymbol{\alpha})_j = \sum_{k=1}^{p} \mu_k \sigma_k u_{kj},$$

or, in matrix form

$$\hat{\Lambda}\boldsymbol{\alpha} = \mathbf{U}\, diag(\sigma_1, \cdots, \sigma_p)\boldsymbol{\mu}$$

where the $N \times p$ matrix $\mathbf{U}$ is equal to $[\mathbf{u}_1, \cdots, \mathbf{u}_p]$. Because the matrix $\mathbf{U}$ satisfies $\mathbf{U}^T\mathbf{U} = I_p$, the $p \times p$ identity matrix, and $\hat{\Lambda}\mathbf{U} = \mathbf{U} diag(\sigma_1^2, \cdots, \sigma_p^2)$, we obtain the values of the μ_k:

$$diag(\sigma_1, \cdots, \sigma_p)\boldsymbol{\mu} = \mathbf{U}^T\hat{\Lambda}\boldsymbol{\alpha} = (\hat{\Lambda}\mathbf{U})^T\boldsymbol{\alpha} = (\mathbf{U} diag(\sigma_1^2, \cdots, \sigma_p^2))^T\boldsymbol{\alpha},$$

hence

$$\boldsymbol{\mu} = diag(\sigma_1, \cdots, \sigma_p)\mathbf{U}^T\boldsymbol{\alpha}.$$

Conversely, if we replace the μ_k by these values in the right handside of (9.3), we verify that we obtain the left handside.

It remains to verify that $\Lambda(v_j) = \sigma_j^2 v_j$, $j = 1, \cdots, p$. By definition

$$\Lambda(v_j) = \sum_{i=1}^{N} < v_j, \beta_i - \hat{\beta} > (\beta_i - \hat{\beta}).$$

We replace in the right handside of this equation v_j by its expression (9.2), use proposition 72, and obtain the desired result. □

The velocities v_k, $k = 1, \cdots, p$ can be interpreted as modes of variation of the shape and the σ_k^2's as variances for these modes. Looking at how the mean shape varies with respect to the kth mode is equivalent to solving the following PDEs:

$$\Gamma_t = \pm v_k\, \mathbf{n} \tag{9.5}$$

with initial conditions $\Gamma(0, .) = \hat{\Gamma}(.)$. Note that v_k is a function of Γ through Λ which has to be reevaluated at each time t. One usually solves these PDEs until the distance to $\hat{\Gamma}$ becomes equal to σ_k.

9.3 Examples of modes of variation

An example of this evolution for the case of the fingers is shown in figure 9.4. Another interesting case, drawn from the example of the eight fish of figure 9.3, is shown in figure 9.5 where the first four principal modes of the covariance operator corresponding to those eight sample shapes are displayed.

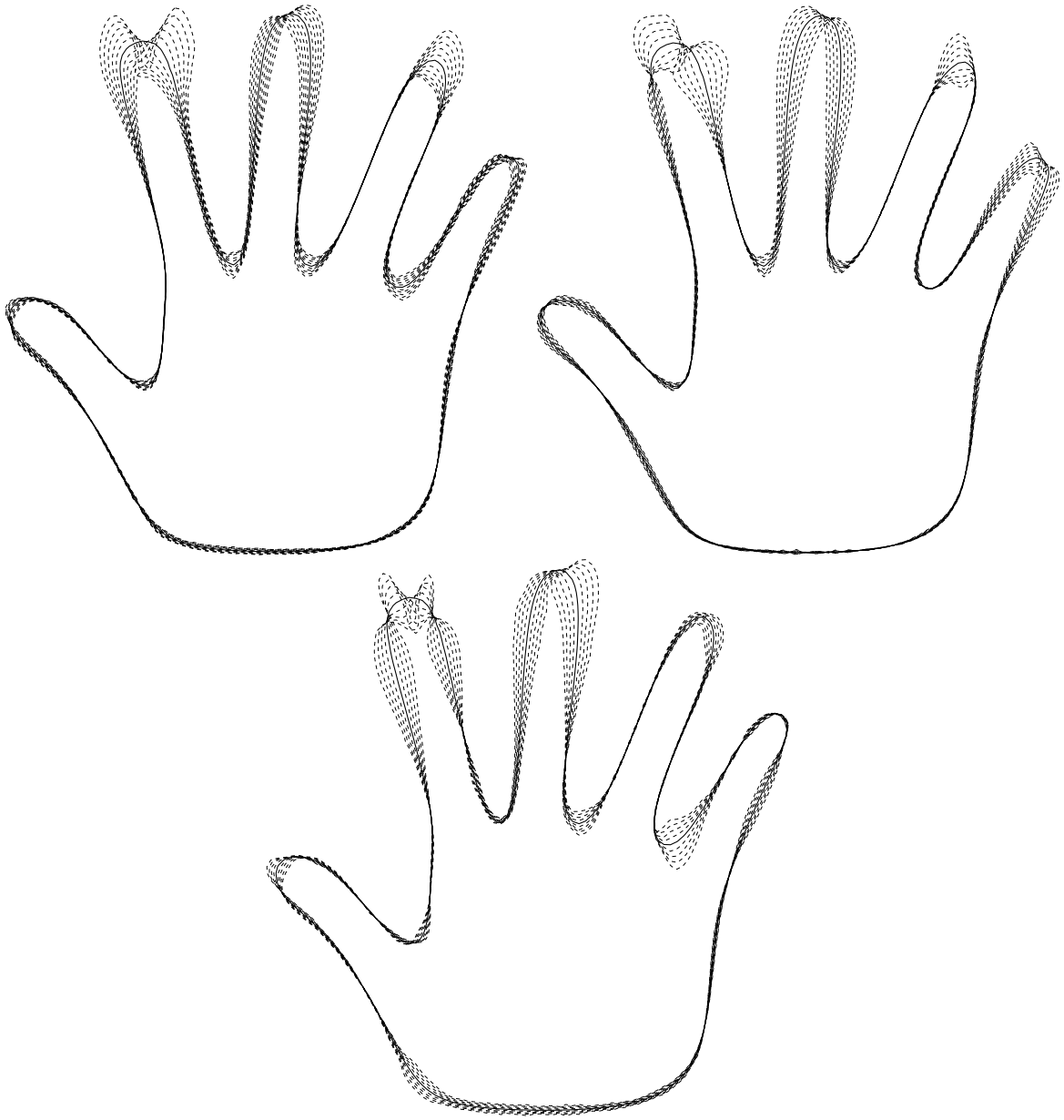


Figure 9.4: The first three modes of variation for the nine sample shapes and their mean shown in figure 9.2. The mean is shown in thick continuous line, the solutions of equation (9.5) for $k = 1, 2, 3$ are represented in dotted lines.

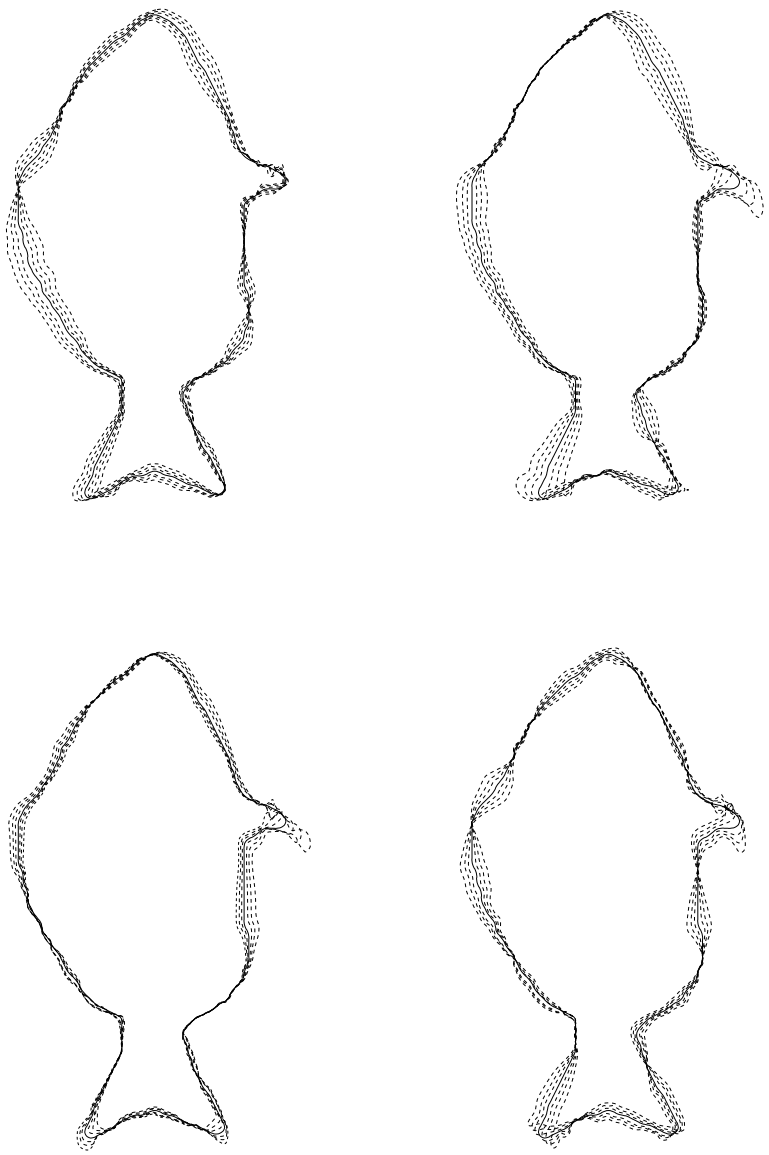


Figure 9.5: The first four modes of variation for the eight sample shapes and their mean shown in figure 9.3. The mean is shown in thick continuous line, the solutions of equation (9.5) for $k = 1, \cdots, 4$ are represented in dotted lines.

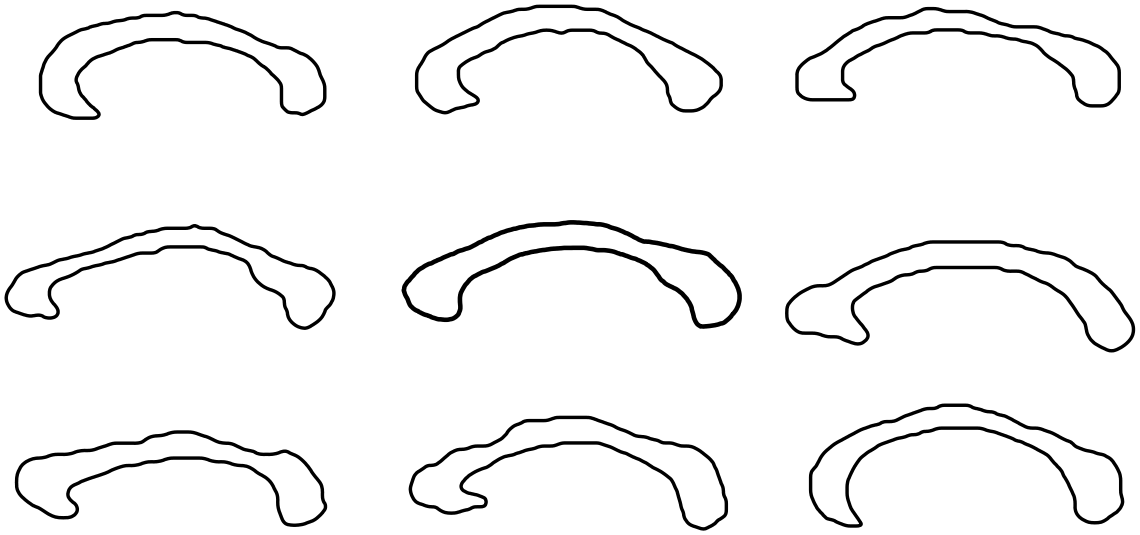


Figure 9.6: The mean of eight silhouettes of corpus callosum (middle, thick line).

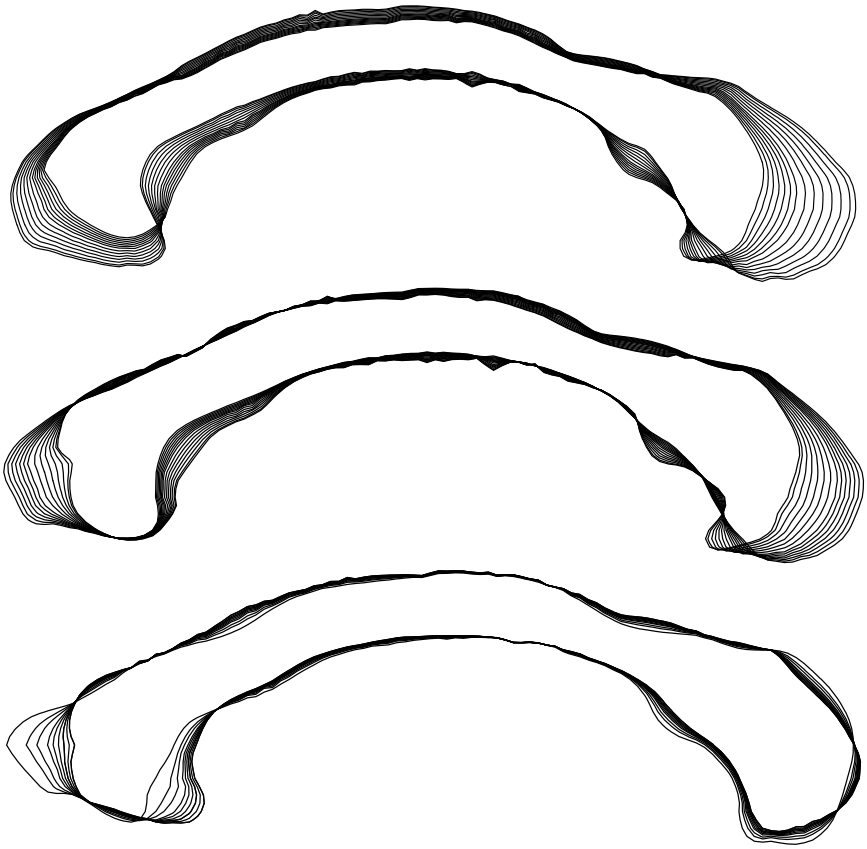


Figure 9.7: From top to bottom, the first three principal modes of variation for the eight sample shapes. They are the solutions of equation (9.5) for $k = 1, 2, 3$.

9.4 Comparison with other approaches

We have presented in section 1.1 the similarities and dissimilarities of our work with that of others. We would like to add to this presentation the fact that ours is an attempt to generalize to a nonlinear setting the work that has been done in a linear one by such scientists as Cootes, Taylor and their collaborators [26] and by Leventon et al. who, like us, proposed to use distance functions to represent shapes in a statistical framework but used a first-order approximation by assuming that the set of distance functions was a linear manifold [74, 73] which of course it is not. Our work shows that dropping the incorrect linearity assumption is possible at reasonable costs, both theoretical and computational.

In this respect we would also like to emphasize that in our framework the process of linear averaging shape representations has been more or less replaced by the linear averaging of the normal deformation fields which are tangent vectors to the manifold of all shapes (see the definition of the covariance operator in section 9.2) and by solving a PDE based on these normal deformation fields (see the definition of a mean in section 9.1 and of the deformation modes in section 9.2).

Another advantage of our viewpoint is that it apparently extends graciously to higher dimensions thanks to the fact that we do not rely on parameterizations of the shapes and work intrinsically with their distance functions (or approximations thereof).

Chapter 10

Image Statistics and Object Classification

Abstract

We would like now to apply to the case of images an approach similar to the one previously presented for curves and surfaces. We propose a way to deal with the first and second order statistics of a set of images. These statistics take into account the images characteristic deformations and their variations in intensity. The central algorithm is based on non-supervised diffeomorphic image matching (without landmarks or human intervention). As they convey the notion of the mean shape and colors of an object and the one of its common variations, such statistics of sets of images may be relevant in the context of object recognition. The proposed approach has been tested on a small database of face images to compute a mean face and second order statistics. The results are encouraging since, whereas the algorithm does not need any human intervention and is not specific to face image databases, the mean image looks like a real face and the characteristic modes of variation (deformation and intensity changes) are sensible. As a step further toward the evaluation of the approach, we present facial expression recognition experiments. We test the recognition of the facial expression of someone with and without the knowledge of his/her face with no expression.

10.1. Introduction

How to find or recognize an object in an image? This is one of the most outstanding open problems in computer vision. Its solution will require a better understanding of the various possible visual aspects of a given object or a class of objects. For example, in the case of faces the description should include variations due to viewpoint, illumination, expression (happiness, surprise, ...), or the identity of the person. Like [53, 56] we think that statistics on images are necessary in order to tackle this problem. What we propose in this article is in a sense an extension to the set of images of an object of the work done on the statistics of 2D or 3D shapes [81, 18, 68]: by computing, from a set of images of a class of objects, the various ways these images can be warped onto one another we define and compute a mean image for that class and its second order statistics. Note that unlike previous approaches, e.g., [56] our approach does not require any manual intervention to identify landmarks or regions of interest. We work directly on the deformation fields which establish the correspondences

between the whole images, since these fields are the fundamental elements of the problem. In order to do this we build upon previous work on non-supervised algorithms that build such correspondence fields between images, e.g., [81, 115, 60, 43]

In Section 10.2 we model the matching problem between two images and describe a variation of a matching algorithm proposed in [60] and analyzed in [43]. In Section 10.3 we use it to define and compute the mean image of a set of images and in Section 10.4 to define and compute its second order statistics. Then in Section 10.5 we show how to use the mean image in an expression recognition task.

10.2. Image matching

The main difficulty when defining the mean of several images is that this mean is supposed to *look like* each one of the images. This implies that the images have been registered and supposes the knowledge of a way to estimate the similarity of any couple of images. This is why we consider now the matching problem between only two images.

Let A and B be two images. We think of them as positive real functions functions defined in a rectangular subset Ω of the plane $\mathbb{R}^2$. We search for a deformation field $\mathbf{f}$ such that the warped image $A \circ \mathbf{f}$ resembles B. More precisely, we would like the field $\mathbf{f}$ to be smooth enough and invertible, i.e. it should be a diffeomorphism from the rectangular subset Ω to itself, which leads us to assume that the diffeomorphism $\mathbf{f}$ equals the identity on the image boundary $\partial\Omega$. Other possibilities are offered by extending the images to a larger subset Ω_1.

In order to keep $\mathbf{f}$ continuous, we have to consider a regularizing term $R(\mathbf{f})$ on $\mathbf{f}$, for example $R(\mathbf{f}) = \|\mathbf{f} - Id\|^{H^1}_{\Omega}$ where Id is the identity function on Ω and $\|a\|^{H^1}_{\Omega} = \int_{x\in\Omega} \|a(x)\|^2 + \|Da(x)\|^2 \, dx$. If we prefer to be sure $\mathbf{f}$ is a diffeomorphism and remains invertible, we can consider $\|\mathbf{f} - Id\|^{H^1}_{\Omega} + \|\mathbf{f}^{-1} - Id\|^{H^1}_{\Omega}$, where $\mathbf{f}^{-1}$ is the inverse of $\mathbf{f}$.

The choice of a similarity criterion between two images has already been studied in part 3.3.2 and we will consider here the local cross-correlation. Note that there exist many other local criteria, for instance the mutual information, to quantify the similarity between variations.

The two-image matching algorithm consists in minimizing with respect to the deformation field $\mathbf{f}$ (initialized to the identity) through a multi-scale gradient descent as proposed in paragraph 3.3.5 the following energy (see [43] for details)

$$E(A, B, \mathbf{f}) = LCC(A \circ \mathbf{f}, B) + R(\mathbf{f}).$$

Thus we obtain a field $\mathbf{f}$ which establishes the correspondences between the two images A and B.

10.3. The mean of a set of images

Now that we know how to compute a diffeomorphic matching between two images, we can try to infer from this a new algorithm for the computation of the mean of n images A_i indexed by $i \in \{1, \ldots, n\}$. This is not as easy as one could guess. We present here three different methods, from the simplest, naive one, to a less intuitive but far better one.

10.3.1. An intuitive algorithm: find the mean

We can first define the mean as the image M which looks the most like all the warped images, i.e., if we introduce n diffeomorphisms $\mathbf{f}_i$ in order to warp each image A_i on the

mean M, we could minimize

$$\sum_i E(A_i \circ \mathbf{f}_i, M, \mathbf{f}_i)$$

with respect to both M and the fields $\mathbf{f}_i$. But how do we choose the initial image M? Besides, here is the main problem: we should not minimize the energy E with respect to an image. Indeed, if we consider the case where $n = 2$ and the two images are the same one translated by a few pixels, the gradient term due to the diffeomorphisms should move them so as to find the translation, but this is prevented by the minimization with respect to the mean image M, which, by averaging the intensities, introduces new contours induced by those in the two images. Consequently, at the first step of the minimization the contours from any of the two images start appearing in M at the same location, and at the second step the diffeomorphisms will not evolve from the identity anymore.

10.3.2. Another intuitive algorithm

We can then try to substitute in E an expression for M as a function of the diffeomorphisms and images, thus effectively eliminating the unknown M, in order not to have to take the derivative of E with respect to an image. For example, we can choose $M = \frac{1}{n}\sum_i A_i \circ \mathbf{f}_i$ and minimize with respect to the $\mathbf{f}_i$ the following criterion:

$$\sum_i E(A_i \circ f_i, \frac{1}{n}\sum_k A_k \circ \mathbf{f}_k, \mathbf{f}_i)$$

We then encounter another problem: we do not take the derivative of E with respect to an image, but we try to match for each i the warped image $A_i \circ \mathbf{f}_i$ and $\frac{1}{n}\sum_i A_i \circ \mathbf{f}_i$. As $\frac{1}{n}\sum_i A_i \circ \mathbf{f}_i$ is the sum of the warped images, it contains in particular all the contours of $A_i \circ \mathbf{f}_i$, which means that we still have the same problem as before: the diffeomorphisms are immediately stuck in a local minimum.

10.3.3. The final word: eliminating the mean

The problem comes mostly from the fact that we are trying to work directly on the mean of the images, whereas we should work only with the fields $\mathbf{f}_i$, which carry all the information about the problem. Indeed, the mean M contains much less information than the diffeomorphisms $\mathbf{f}_i$: for example the mean of a white disk on a black background and a black disk on a white background is uniformly grey and consequently has not a large LCC-correlation with the initial images. Therefore we should rather deal with pairs of warped images than with pairs of a warped image and the mean. The mean then becomes an auxiliary quantity, just computed at the end when the diffeomorphisms are known.

The algorithm proceeds as follows: initialize all deformation fields $\mathbf{f}_i$ to the identity, and let them evolve in a multiscale framework in order to minimize

$$\frac{1}{n-1}\sum_{i\neq j} LCC(A_i \circ \mathbf{f}_i, A_j \circ \mathbf{f}_j) \quad + \quad \sum_k R(\mathbf{f}_k)$$

Thus, at the end of the evolution, each $A_i \circ \mathbf{f}_i$ is supposed to look like each of the others, and the mean is naturally computed as $M = \frac{1}{n}\sum_i A_i \circ \mathbf{f}_i$. The regularizing term $\sum_i R(\mathbf{f}_i)$ implies that if several sets of fields $\mathbf{f}_i$ conduct to approximatively the same energy $\sum_{i\neq j} LCC(A_i \circ \mathbf{f}_i, A_j \circ \mathbf{f}_j)$ (for example by adding a common diffeomorphism $\mathbf{f}_c$ to every field and replacing $\mathbf{f}_i$ with $\mathbf{f}_i \circ \mathbf{f}_c$), then the most intuitive one is chosen (the one of least regularizing cost). In order to accelerate the process in practice, we also impose the condition $\sum_i \mathbf{f}_i = 0$ at each time step by subtracting the mean of the fields $\frac{1}{n}\sum_i \mathbf{f}_i$ to each of them.

Figure 10.1: The ten face images.

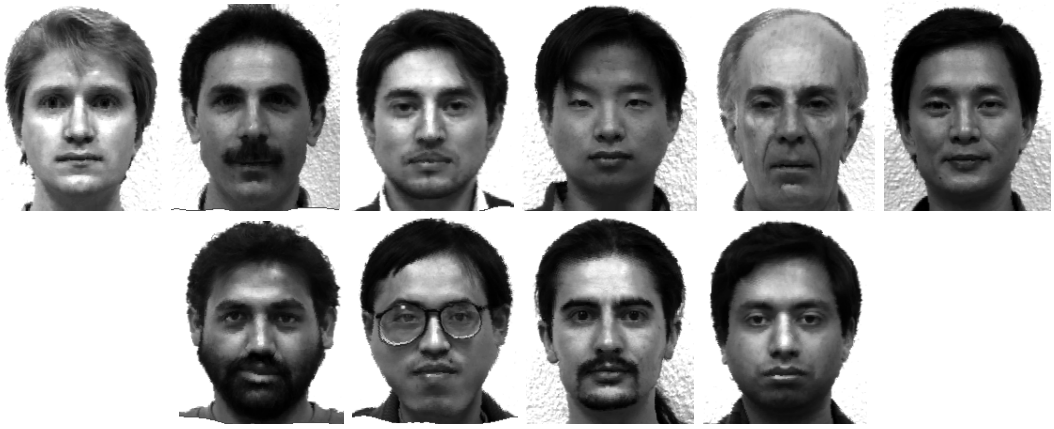


Figure 10.2: The ten warped images $A_i \circ \mathbf{f}_i$.

10.3.4. Example

We have tested this algorithm on a face database from Yale[1]. More precisely, we have computed the mean face out of photographs of ten different people with similar expressions, approximatively the same illumination and position conditions, and the same size (195 * 231 pixels). The ten image A_i are shown in figure 10.1, the ten warped images $A_i \circ \mathbf{f}_i$ in figure 10.2, and their mean in figure 10.3.

Note the accuracy of the mean: it looks like a real face, its features are very sharp, not blurred at all (except the ears), thanks to the simultaneous accurate matching of all images. If we had used one of the algorithm centered on the mean image instead of the diffeomorphisms themselves, we would have obtained a completely blurred image because of non-corresponding edges of different images (the fields being stuck in local minima before starting to evolve), not far better than a bad simple average of every pixel of all initial images without warping.

The strange white curved line below the eyes of the mean comes from the reflects of the light into the eighth man's glasses, which the algorithm probably confused (and matched) with the brightness of the top of the other cheeks.

[1] `http://cvc.yale.edu/projects/yalefaces/yalefaces.html`

Figure 10.3: The mean of the previous ten faces.

This computation took about 10 minutes on a standard workstation. Note once again the good job done by the diffeomorphisms $\mathbf{f}_i$, on figure 10.2, with in mind the fact that there is no human intervention to help the algorithm find the good correspondences, that the algorithm is absolutely not specific to face databases, and that there is no use of any kind of prior on the images.

10.4. Second order statistics of a set of images

Now that we are able to compute the mean image of a set of images, we would like to study its characteristic modes of variation. Indeed, the knowledge of only the mean may be not sufficient to have a good idea of the whole set of images. For example, there may exist some relevant typical kinds of changes in the shape or the intensity of an object, without the knowledge of which you may not be able to discuss the belonging of a new image to the class defined by the set of images you studied before.

As the information about the shape variations in the set of images A_i lies in the diffeomorphisms $\mathbf{h}_i$, we compute statistics on these warping fields. The same way, as the information about the intensity variations (changes of skin texture, of hair color...) lies in the intensity of the warped images $A_i \circ \mathbf{f}_i$, since when they are warped their pixels are corresponding, we also compute statistics on the intensity of the warped images. Finally, as there could be links between shape variations and intensity variations, we compute combined statistics.

10.4.1. Definition and computation

These deformation fields are functions from a subset Ω of the plane $\mathbb{R}^2$ to itself, therefore the natural way to express correlation between two fields $\mathbf{a}$ and $\mathbf{b}$ is to compute their inner product for the usual norm $L^2(\Omega \to \mathbb{R}^2)$:

$$\langle \mathbf{a} \,|\mathbf{b}\rangle_{L^2(\Omega\to\mathbb{R}^2)} = \frac{1}{|\Omega|}\int_\Omega \mathbf{a}(x)\cdot\mathbf{b}(x)\,dx$$

Since the mean $\overline{\mathbf{f}}$ of the fields $\mathbf{f}_i$ is 0 (see above), the (shape-)correlation matrix SCM defined by

$$SCM_{i,j} = \left\langle \mathbf{f}_i - \overline{\mathbf{f}} \,\middle|\, \mathbf{f}_j - \overline{\mathbf{f}} \right\rangle_{L^2(\Omega\to\mathbb{R}^2)}$$

can be simplified in

$$SCM_{i,j} = \langle \mathbf{f}_i \,|\mathbf{f}_j\rangle_{L^2(\Omega\to\mathbb{R}^2)}\,.$$

Then we diagonalize the correlation matrix SCM (its size $n \times n$ depends on the number of images, not the number of pixels), and extract its eigenvalues σ_k and normalized eigenvectors

$\mathbf{v}_k$. We obtain $n-1$ modes of deformation (one being null because of the linear constraint $\sum_i \mathbf{f}_i = 0$), and the k^{th} mode $\mathbf{m}_k$ is given by the coefficients of $\mathbf{v}_k$:

$$\mathbf{m}_k = \sum_i (\mathbf{v}_k)_i \, \mathbf{f}_i.$$

As statistics were made in the linear space $L^2(\mathbb{R}^2 \to \mathbb{R}^2)$, we can continuously apply a mode $\mathbf{m}_k$ to the mean image M with an amplitude α ($\in \mathbb{R}$) by computing the image $M \circ (Id + \alpha(\mathbf{m}_k - Id))$, and then produce animations of the deformations. Note however that a linear combination of such modes is not guaranteed anymore to be a diffeomorphism since it may happen not to be invertible, even if such a phenomenon was not observed in the studied example for reasonable values of the components α.

10.4.2. Example

These modes are illustrated in figure 10.4. Each column represents a mode, starting from the main one (leftmost column) to the one with the smallest eigenvalue, which is actually 0 since one mode is null (rightmost column). Each column is divided in five images: in the central image of each column, we represent the mean we computed before; in the images just above and underneath the mean, we represent the mode applied to the mean with amplitude $+\sigma_k$ and σ_k; and then with amplitude $+2\sigma_k$ and $-2\sigma_k$ in first and last image of each column, in order to exaggerate and better visualize the deformations.

Note that the images on the second and fourth lines still look like normal faces of various people; it is a very good point since they are supposed to be characteristic examples of what shape variations the mean face can undergo without getting out of the class of face images.

On the contrary, images on the first and last lines are stranger: even if we still recognize they look human a bit, we see immediately that there are not real; which is not the case of the other lines. This is also a good point, since these images have been obtained by applying the characteristic modes twice too far (with amplitude $2\sigma_k$ instead of σ_k), which shows that the amplitudes of the deformations (the values of σ_k) are right, and shows that a set of images is well described by its characteristic shape variations.

10.4.3. Intensity variations

In order to take all the face variations into account, we should not only consider the shape variations (i.e. the diffeomorphisms) but also the intensity variations. As before, we can define an intensity-correlation matrix ICM on the intensity variations I_i:

$$I_i = A_i \circ \mathbf{f}_i - M$$

for the $L^2(\mathbb{R}^2 \to \mathbb{R})$ inner product. Thus, we can compute the principal modes of intensity variations, which correspond to skin or hair changes for a shape-fixed head (see figure 10.5).

We can also combine shape and intensity variations. If we note $\sigma_S^2 = \frac{1}{n}\sum_i \|\mathbf{f}_i\|^2$ and $\sigma_I^2 = \frac{1}{n}\sum_i \|I_i\|^2$ the standard deviations of shapes and intensities, we can define a combined correlation matrix CCM by

$$CCM = \frac{1}{\sigma_S^2} SCM + \frac{1}{\sigma_I^2} ICM$$

and proceed as before, compute and display principal modes of variations. This matrix can be considered as resulting from a inner product on the set of variations (shape and intensity):

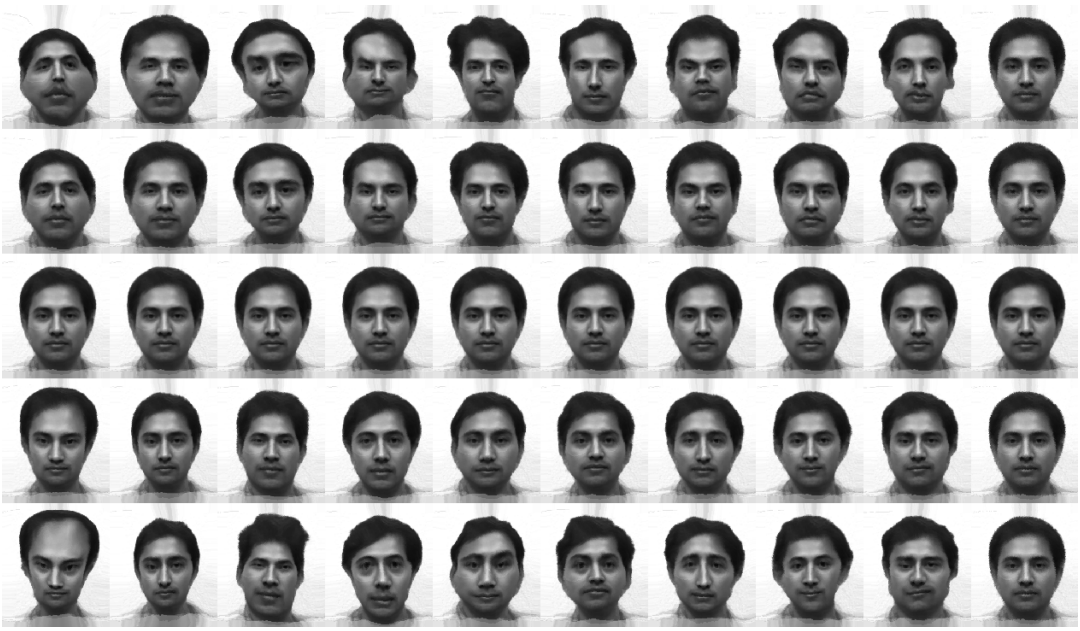


Figure 10.4: The shape modes of deformation of the previous set of images. Each column represents a mode, applied to the mean image with amplitude $\alpha = \{2\sigma_k, \sigma_k, 0, -\sigma_k, -2\sigma_k\}$.The (relative) values of the eigenvalues are, from left to right, 1, 0.5, 0.3, 0.1,..., 0.05, 0.

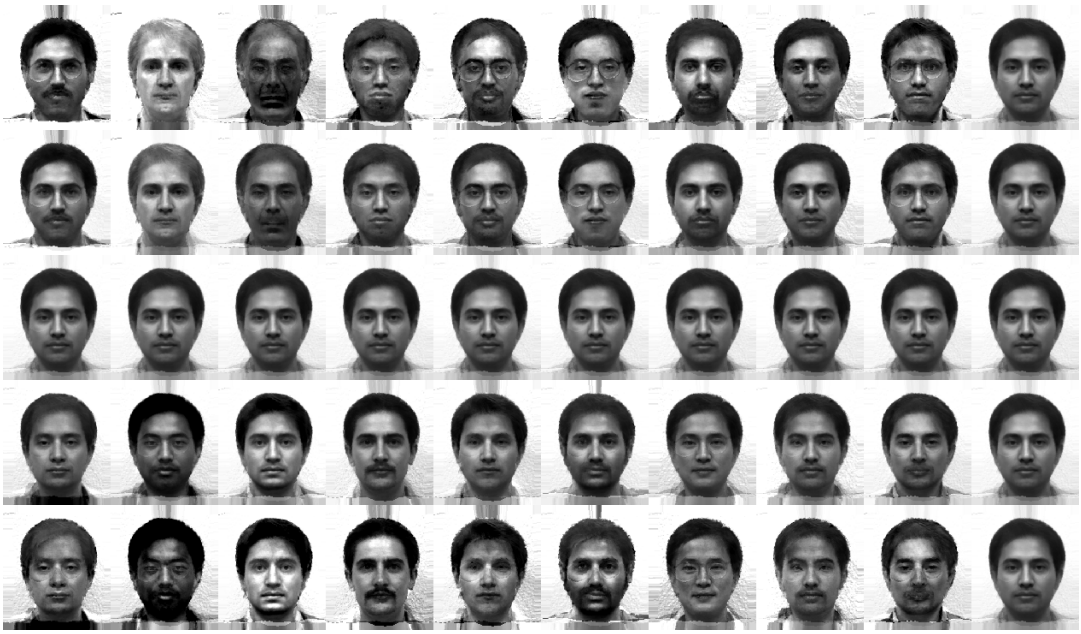


Figure 10.5: The characteristic modes of variation of the intensity (same set of images). Each column represents a mode, applied to the mean image with amplitude $\alpha = \{2\sigma_k, \sigma_k, 0, -\sigma_k, -2\sigma_k\}$.

from two elements $(\mathbf{f}_i, I_i)$ and $(\mathbf{f}_j, I_j)$ in this set, we can compute their correlation:

$$\langle \mathbf{f}_i, I_i \,|\mathbf{f}_j, I_j \rangle = \frac{1}{\sigma_S^2} \langle \mathbf{f}_i \,|\mathbf{f}_j \rangle + \frac{1}{\sigma_I^2} \langle I_i \,|I_j \rangle$$

where the two coefficients stand for the relative variability of each component. The results are shown on figure 10.6. Note again how these faces are realistic and diversified (hair, skin, illumination, mustache). We can see more various attitude than before, when we considered only shape variations. This is partly due to the fact that illumination and shadow carry information on the 3D shape (for example, the shape of the cheeks) which is not directly retrievable from the sharpest edges of 2D images.

Figure 10.6: The eight non-zero combined modes of deformation of the same set of images without the subject with glasses. Each column represents a mode, applied to their mean image with amplitude $\alpha = \{\sigma_k, -\sigma_k\}$. The (relative) values of the eigenvalues are, from left to right, 1, 0.555, 0.505, 0.424, 0.286, 0.232, 0.162, 0.135.

10.5. Classification: Expression Recognition

Let us now consider the facial expression recognition task. The goal is to associate with any new face its expression. We still use the Yale database. We remove from this database the 2 subjects wearing glasses and we consider the 5 following facial expressions: happy, sad, sleepy, surprised and winking, beside the "normal" one.

10.5.1. From the mean image

The following simulations show that deformations from a mean face can be used to classify facial expressions. More precisely we choose as a reference face the mean "normal" face of the first 9 subjects of the database. Our first classification procedure uses a Support Vector Machine with Gaussian kernel[2] on the deformations from this face to expressive faces. To measure the efficiency of the method, we cross-validate the errors by taking out one subject among the 13 subjects in the database and consequently using 60 faces labeled by their expression to deduce the expression of the five remaining faces[3].

[2]The bandwidth of the kernel is equal to the median of the norms of the difference between 2 deformations of the training set.

[3]Thus we have no prior information on the subject to classify his facial expressions.

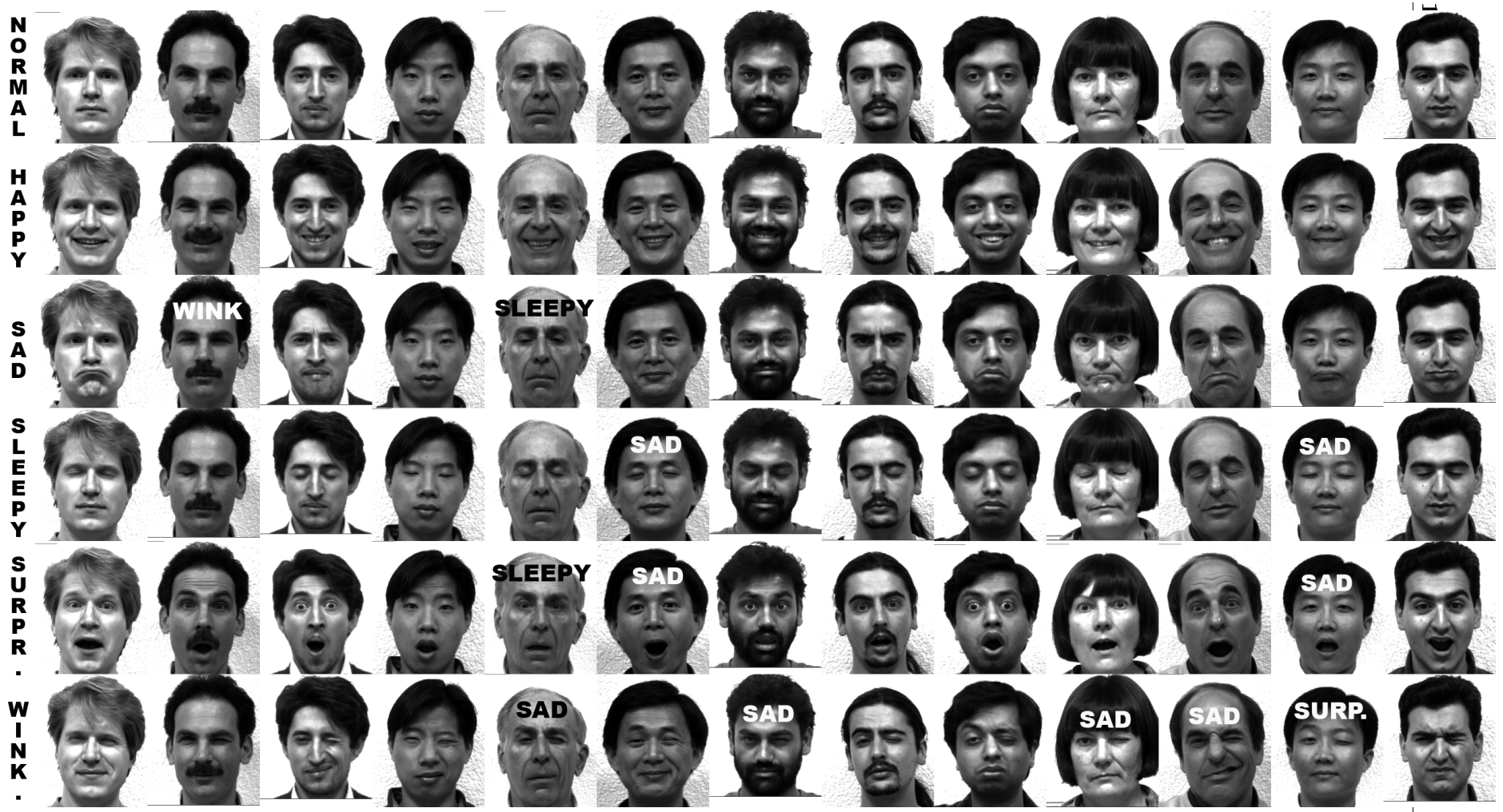


Figure 10.7: Expression recognition using deformations of the normal face, aligned with the mean face. From top to bottom: normal, happy, sad, sleepy, surprised and winking faces. 53 of the 65 expressive faces are correctly classified, 12 are not. For these we show the incorrect label.

The cross-validation error is 24 upon 65 faces. For comparison purposes, we trained a Support Vector Machine with Gaussian kernel[4] using only the gray level intensity information. In this case we obtained a larger cross-validation error of 27 upon 65 faces which shows the interest of using the diffeomorphisms.

10.5.2. With knowledge of the face without expression

The advantage of using the deformations instead of the gray level intensities is even larger when we know whose face it is that we want to process. More precisely, if we use the subject's "normal" face to compute the deformation between the expressive one and classify this "expression" deformation after alignment on the mean face (using the "subject" deformation between the mean face and the face without expression), the cross validation error goes down to 12 (upon 65), whereas the classification using the difference of gray level intensities between the expressive face and the normal one leads to 17 errors.

These results, although preliminary, indicate that the mean image can be very useful in a classifying task, considering that the database is small, that the procedures were not specialized to faces, and that even a human classifier may have problem with some of the considered faces (see figure (10.7) again)!

They also stress the importance of the notion of deformation: from the raw gray level images we extracted spatial deformations, and from the only knowledge of these deformations (without even the knowledge of the intensity variations !) we obtained better results with SVM than the direct application of SVM to the gray level images, whereas there is *a priori* more information in the whole initial images. The design of the quantities of interest when trying to build statistics via such tools as kernel methods is consequently fundamental, even if in the case of these particular experiments SVM on raw images appears to be surprisingly efficient.

10.6. Summary and Conclusions

We have defined and computed first and second order statistics of a set of images with a diffeomorphic matching approach (without landmarks or human intervention), and shown how to use them in a classification task. We have tested this general approach on a face database, and the results are encouraging: the mean face looks like that of a real human being, with sharp contours, the modes of variations (shape and intensity) are very sensible, and the expression recognition results are good, especially if we are also given a "normal" image of the face to classify. We insist on the fact that our methods are not specific to faces and do not use any prior on the kind of images.

Acknowledgments

The SVM were run using the Spider software [5] provided by the Department of Empirical Inference for Machine Learning of the Max Planck Institute for biological cybernetics.

[4]The bandwidth of the kernel is the median of the norms of the intensity difference between 2 images of the training set.

[5]`http://www.kyb.tuebingen.mpg.de/bs/people/spider`

Chapter 11

Image Segmentation with Shape Priors

Abstract

The definition of shape priors based on the previously computed shape statistics and to their application to an image segmentation task is now addressed. Several kinds of priors are considered, some necessary heavy calculi are computed, and some results are shown.

11.1 Image Segmentation

Finding the contour of an object in an image is a very difficult and ill-posed task. We are here interested in the precise task of finding an object in an image with the knowledge of a set of examples, i.e. a set of images such that in each of them the object appears and has already been segmented. Usually, the only information retrieved from the set of examples comes from statistics on the intensity of the two regions (the inside of the object and the background), which does not carry any information about the shape of the object. Consequently, during an active-contour based evolution in order to segment a new image, the only restriction concerning shape will be brought by the regularity term which imposes the smoothness of the contour. Nevertheless, priors about the shape of the object that is looked for in an image can greatly improve these established techniques.

We present here some ways to take into account the shape statistics computed in chapter 9 into the standard active-contour algorithms for segmentation.

Given a set of n shapes $\mathcal{D} = (\Gamma_i)_{1 \leqslant i \leqslant n}$ and their mean M, let us denote by α_i the deformation field $-\nabla_M E^2(M, \Gamma_i)$ defined on M. As studied before, each α_i is a best deformation field whose application to M moves it closer to the shape Γ_i. As all these fields α_i belong to the same space, the tangent space of the mean curve M, we can easily compare them and compute statistics, as done in chapter 9. Besides, drawing directly an histogram of these warping fields α_i is not easy for the simple reason that these fields are infinite dimensioned. The next section is devoted to the definition of several shape priors based on different ways to introduce a distance to the shape distribution $\mathcal{D}$.

11.2 Shape Priors

11.2.1 Context

Let C be the evolving curve which we would like to fit the contour of the object in the new given image I. We can express any energy minimization as a probability maximization, the two approaches being essentially the same. We would like to maximize the probability:

$$P(C|I)$$

with respect to C. By the way:

$$P(C|I) \;=\; \frac{P(C,I)}{P(I)} \;=\; P(I|C)\,\frac{P(C)}{P(I)} \;\simeq\; P(I|C)\,P(C).$$

Then $P(I|C)$ is given by the standard approach (based on intensity gradients, statistics on textures and so on), while $P(C)$ expresses the probability that C has such shape. In the sequel $P(C)$ will stand for any positive, upper-bounded energy depending on C, and its total mass will not necessarily be equal to one.

11.2.2 Shape Probability

A simplistic method

Given a distance or energy E between shapes, for instance the approximation of the Hausdorff distance, a new distance between C and the whole shape distribution $\mathcal{D} = (\Gamma_i)$ could be defined as:

$$\sum_i E^2(C, \Gamma_i).$$

But the global minimum of this energy has already been studied and it is the mean shape M, so that this energy is not really more interesting than $E(C, M)$. In particular the shape variations in $\mathcal{D}$ are not taken into account.

Parzen method

Another way to express the probability of belonging to the set $\mathcal{D}$ is to estimate the density of sample shapes around C. The Parzen method sets:

$$P(C) = \sum_i \exp^{-\frac{E^2(C,\Gamma_i)}{2\sigma^2}}$$

for a given scale parameter σ. At scale σ, if the sample density is low around C, then C is trapped into a low potential area around its nearest neighbor and the maximization of $P(C)$ by gradient ascent will lead to this nearest neighbor. On the contrary if the sample density is high with respect to the scale σ then C will be attracted towards a local mean of its nearest neighbors. Some variations on this kind of prior have been proposed by Cremers and Schnörr and colleagues [28, 29, 27].

Kernel method based on the gradient

Another way to estimate the probability of the shape C is to compare the deformation field $\alpha = -\nabla_M E^2(M, C)$ which "links" the mean M to the curve C with the observed fields $\alpha_i = -\nabla_M E^2(M, \Gamma_i)$ which "link" M to each one of the sample curves Γ_i. There is a natural norm in this tangent space, the one that comes from the inner product with respect to which the gradients are computed, say $\|\cdot\|_{L^2}$ for the sake of readability. With Parzen-like considerations, this probability can then be given by

$$P(C) \;=\; \frac{1}{N}\sum_i K(\alpha - \alpha_i)$$

where $K(\alpha - \alpha_i)$ expresses the difference between α and α_i, for example one can choose the Gaussian kernel $K(\beta) = e^{-\frac{1}{2\sigma^2}\|\beta\|^2_{L_2}}$. Note that the maximization of this probability (as well as the following ones) requires the computation of the derivative of the field $\alpha = -\nabla_M E^2(M, C)$ with respect to C, i.e. the second order cross-derivative of E.

Gaussian Eigenmodes (PCA on gradients)

It is also possible to include second order statistics (from section 9.2) into the design of the shape probability. The eigenmodes β_k of deformation of the set of shapes are obtained by PCA applied to the set of observed fields α_i. The most significant modes β_k are the ones with highest associated standard deviation σ_k; therefore you can take only the very first modes into account. However the issue to determine the number of modes of interest is not fundamental since the importance of each mode will be related to its standard deviation and consequently the consideration of some extra modes with low standard deviation will not change significantly the distribution.

The PCA decomposition supposes implicitly that linear combinations of instantaneous deformation fields α_i make sense (which is precisely the expectation of such an approach) and that their distribution is Gaussian, that is to say, that for any mode β_k, the distribution of the k-th principal component $\langle \beta_k \,|\alpha_i\rangle_{L^2}$ of the observed fields α_i is Gaussian, which implies that the shape probability should be of the form

$$P(C) = \prod_k e^{-\frac{\langle \beta_k|\alpha\rangle^2_{L^2}}{2\sigma_k^2}} \times e^{-\frac{\|\mathrm{Rem.}(\alpha)\|_2^2}{2\sigma^2_{\mathrm{noise}}}}$$

where Rem.(α) stands for the remaining part $\alpha - \sum_k \langle \alpha \,|\beta_k\rangle_{L^2}$ that cannot be described by the eigenmodes. Note that the corresponding distance between a shape and the shape distribution is:

$$\sqrt{\sum_k \frac{1}{\sigma_k^2}\,\langle \alpha\,|\beta_k\rangle^2 \;+\; \frac{1}{\sigma^2_{\mathrm{noise}}}\,\|\mathrm{Rem.}(\alpha)\|_2^2}.$$

From a certain point of view, the level sets of this distance are ellipsoidal. This distance is a variation on the Mahalanobis distance.

Eigenmode Histograms

However the distribution is generally not Gaussian. Then the modes can be computed with ICA (independent component analysis) instead of PCA, and for each mode k the

histogram h_k of the observed components $\langle \beta_k \, | \alpha_i \rangle$ can be drawn. The histograms may need some smoothing if the density of the distribution is low. Then the empirical probability is given by:

$$P(C) = P(\alpha) = \prod_k h_k(\langle \beta_k \, | \alpha \rangle).$$

11.2.3 Invariance with respect to Rigid Motion

In most cases one may prefer the shape probability $P(C)$ to be invariant with respect to rotations, translations and scalings of the shape C. Let us denote by $\mathbf{t}$ a translation in the plane $\mathbb{R}^2$, by θ the angle of a rotation and by s the factor of a scaling centered on the inertial center of C. To be concise the rigid motion with these parameters is denoted by $R_{\mathbf{t},\theta,s} = R$. Thus we introduce:

$$P(C, \mathbf{t}, \theta, s) \; = \; P(R_{\mathbf{t},\theta,s} C) \; = \; P\,(RC)$$

which leads to a new probability distribution:

$$P_2(C) \; = \; \max_{\mathbf{t},\theta,s} P(C, \mathbf{t}, \theta, s) \; = \; \max_R P(RC)$$

However, in practice, the estimation of the best parameters $\mathbf{t}$, θ and s for the curve C could be progressively computed at the general level at the same time as the curve C itself, rather than completely at each step of the maximization algorithm (for each successive value of C). Hence we have the choice between two slightly different possibilities, induced by these two ways of writing the expression to minimize:

$$\log \left[\max_C P(C|I) \right] \; = \; \max_C \log \left[P(I|C) \right] \; + \; \max_C \log \left[\max_R P(RC) \right]$$

or

$$\log \left[\max_{C,\mathbf{t},\theta,s} P(C|I) \right] \; = \; \max_C \log \left[P(I|C) \right] \; + \; \max_{C,\mathbf{t},\theta,s} \log \left[P(RC) \right].$$

11.2.4 Pre-Computing

In the three last cases, α stands in the expression of the probability to maximize, which implies that the derivative of α with respect to C will have to be computed. For example in the kernel method based on the gradient, we need the expression of the gradient of $P(C)$ with respect to the curve C (more exactly, of $P(C, \mathbf{t}, \theta, s)$ with respect to each of its variables, which essentially leads to the same calculus):

$$\nabla_C P(C) \; = \; \frac{1}{N} \sum_i \; \nabla_\beta K(\beta)|_{\beta = \nabla_M E^2(M,C) - \alpha_i} \; \times \; \nabla_C \nabla_M E^2(M, C).$$

Hence we need to compute part of the second order derivative of E, which requires some bravery in the case of the approximation of the Hausdorff distance... Many authors have considered signed distance functions as a mean to express shape priors [94, 98]. Here, the calculations happen to be sometimes much simpler if the energy is based not on a shape but on its distance function, that is to say if you consider that the real object of interest is now a function defined on the whole space $\mathbb{R}^n$ and not only its zero level, regardless of whether it is a distance function or not. In the case of the $W^{1,2}$ norm of the distance functions of the shapes, if the derivatives are directly computed with respect to the function itself

(which does not guarantee that the evolving function should remain a distance function), the obtained probability gradient is:

$$\nabla_{d_C} P(d_C) = \sum_i (Id - \Delta) \left(\nabla_\beta \, K(\beta)|_{\beta=(Id-\Delta)(d_C - d_{\Gamma_i})} \right)$$

and for the L^2 norm only the gradient is:

$$\nabla_{d_C} P(d_C) = \sum_i \nabla_\beta \, K(\beta)|_{\beta=d_C - d_{\Gamma_i}} .$$

This last criterion is very similar to the one developed by Leventon, Grimson and Faugeras in [73, 74], which consists in the application of PCA to signed distance functions. However such an approach is questionable since a linear combination of signed distance functions is generally not a signed distance function and has sometimes a really unexpected zero level.

11.2.5 Influence of the inner product

The choice of an inner product in the tangent space to a shape defines the notion of gradient. As shape priors involve the gradient α as well as statistics about the observed gradients in the sample, it could be interesting to study more precisely the influence of this choice on the prior. In the spirit of the section 7.3.1, if the chosen inner product P is related to the L^2 inner product through a symmetric definite operator S by $\langle u \,|v\rangle_P = \left\langle S^{-1}u \,|v\right\rangle_{L^2}$, then the proposition 57 states that the gradient of any energy $E(\Gamma)$ is linked to the L^2 gradient through the expression: $\nabla^P E(\Gamma) = S\left(\nabla^{L^2} E(\Gamma)\right)$.

The quantities of interest here are derivatives with respect to the average shape M: $\alpha = -\nabla^{L^2}_M E^2(M, C)$ and $\alpha_i = -\nabla^{L^2}_M E^2(M, \Gamma_i)$. For another inner product P, they are respectively:

$$\alpha^P = -\nabla^P_M E^2(M, C) = S\alpha \quad \text{and} \quad \alpha_i^P = -\nabla^P_M E^2(M, \Gamma_i) = S\alpha_i \, .$$

Then, for any field β, $\left\langle \alpha^P \,|\beta\right\rangle_P = \langle \alpha \,|\beta\rangle_{L^2}$ and, if $S^{1/2}$ is such that ${}^{\mathbf{t}}S^{1/2}S^{1/2} = S$:

$$\left\langle \alpha_i^P \,|\alpha_j^P\right\rangle_P = \left\langle \alpha_i \,|\alpha_j^P\right\rangle_{L^2} = \left\langle \alpha_i^P | \, \alpha_j\right\rangle_{L^2} = \left\langle S^{1/2}\alpha_i \,\middle|\, S^{1/2}\alpha_j \right\rangle_{L^2} .$$

If S stands for a smoothing operator (or rigidifying, etc.) then the correlation between the smoothed fields $\left\langle \alpha_i^P \,\middle|\, \alpha_j^P \right\rangle_P$ is in fact the L^2 correlation between the half-smoothed fields $\left\langle S^{1/2}\alpha_i \,\middle|\, S^{1/2}\alpha_j \right\rangle_{L^2}$.

In the case of the L^2 inner product, the eigenmodes β_k are obtained by diagonalization of the correlation matrix $A = \left(\langle \alpha_i \,|\alpha_j\rangle_{L^2}\right)_{1 \leqslant i,j \leqslant n}$. More exactly, A is a $n \times n$ symmetric positive semi definite matrix and admits n eigenvectors $\mathbf{v}_k$, which are in $\mathbb{R}^n$ and satisfy $\langle \mathbf{v}_k \,|\mathbf{v}_{k'}\rangle_{L^2(\mathbb{R}^n)} = \delta_{k=k'}$. Then the modes β_k are defined by $\beta_k = \sum_i (\mathbf{v}_k)_i \, \alpha_i$ and consequently satisfy $\langle \beta_k \,|\beta_{k'}\rangle_{L^2} = \sum_{i,j} (\mathbf{v}_k)_i \, \langle \alpha_i \,|\alpha_j\rangle_{L^2} \, (\mathbf{v}_{k'})_j = {}^{\mathbf{t}}\mathbf{v}_k A \mathbf{v}_{k'} = \delta_{k=k'}\sigma_k^2$.

With the inner product P, the new eigenmodes are related to the matrix

$$A^P = \left(\left\langle \alpha_i^P \,|\alpha_j^P\right\rangle_P\right)_{1 \leqslant i,j \leqslant n} = \left(\left\langle S^{1/2}\alpha_i \,\middle|\, S^{1/2}\alpha_j \right\rangle_{L^2}\right)_{1 \leqslant i,j \leqslant n} .$$

The new eigenvectors $\mathbf{v}_k$ (different from the previous ones) are still orthogonal for the L^2 inner product in $\mathbb{R}^n$. Then new eigenmodes β_k can be defined by $\beta_k = \sum_i (\mathbf{v}_k)_i \, \alpha_i$, or

equivalently, they can be defined by $\beta_k^P = \sum_i (\mathbf{v}_k)_i \, \alpha_i^P = S\beta_k$. They satisfy $\left\langle \beta_k^P \middle| \beta_{k'}^P \right\rangle_P = \left\langle S^{1/2}\beta_k \middle| S^{1/2}\beta_{k'} \right\rangle_{L^2} = \left\langle \beta_k \middle| \beta_{k'}^P \right\rangle_{L^2} = \sum_{i,j} (\mathbf{v}_k)_i \left\langle \alpha_i^P \middle| \alpha_j^P \right\rangle_P (\mathbf{v}_{k'})_j = {}^{\mathbf{t}}\mathbf{v}_k A^P \mathbf{v}_{k'} = \delta_{k=k'}\sigma_k^2$.

Thus, given a set of shapes (Γ_i) and their associated modes (β_k^P), the component of a new field α^P on the kth mode can be computed in several manners:

$$\left\langle \alpha^P \middle| \beta_k^P \right\rangle_P = \left\langle \alpha \middle| \beta_k^P \right\rangle_{L^2} = \left\langle \alpha^P \middle| \beta_k \right\rangle_{L^2} = \left\langle S^{1/2}\alpha \middle| S^{1/2}\beta_k \right\rangle_{L^2}.$$

11.3 Hausdorff second order derivative

We need the literal expression of the second order derivative of the approximation of the Hausdorff distance with respect to the curves in order to include into the usual image segmentation framework an a priori term involving second-order statistics of shapes. More exactly, we need the expression of $\partial_B \partial_A d_H(A, B)$. This is the subject of the four following pages.

11.3.1 Notations

A, B two planar curves

$d(x,y)$ usual Euclidean distance between points x and y

ϕ a strictly decreasing application $\mathbb{R}^+ \mapsto \mathbb{R}^+$

ψ, Ψ strictly increasing applications $\mathbb{R}^+ \mapsto \mathbb{R}^+$

Φ $\dfrac{\psi'}{\phi'}$

Θ $\dfrac{\Psi'}{\psi'}$

$\langle f(x)\rangle_{x\in A}$, $\langle f(\cdot)\rangle_A$ $\dfrac{1}{|A|}\displaystyle\int_A f(x)\,dx$ (mean of application f on curve A)

$\sigma_A(f(\cdot),y)$ $\langle f\rangle_A - f(y)$ (deviation of application f at point y of curve A)

$\phi\circ d(\cdot)$ $\phi\left(d(\cdot)\right)$

$\langle d\rangle_A^{\phi}$ $\phi^{-1}\left(\langle \phi\circ d(\cdot)\rangle_B\right)$ mean on A of application $d(\cdot)$ in the sense of function ϕ

$\langle a,b\rangle^{\Psi}$ $\Psi^{-1}\left(\dfrac{1}{2}\Psi(a)+\dfrac{1}{2}\Psi(b)\right)$ mean of two real values in the sense of function Ψ

$\Phi_{B\phi}(y)$ $\Phi\left(\langle d(\cdot,y)\rangle_B^{\phi}\right)$

$\Theta_{B\phi A\psi}$ $\Theta\left(\left\langle\langle d(\cdot,\cdot)\rangle_B^{\phi}\right\rangle_A^{\psi}\right) = \Theta\left(\left\langle\langle d(x,y)\rangle_{x\in B}^{\phi}\right\rangle_{y\in A}^{\psi}\right)$

$\overrightarrow{D}(y,x)$ $\dfrac{y-x}{d(x,y)}\,\phi'(d(x,y))$

$d_H(A,B)$ $\left\langle\left\langle\langle d(\cdot,\cdot)\rangle_A^{\phi}\right\rangle_B^{\psi}, \left\langle\langle d(\cdot,\cdot)\rangle_B^{\phi}\right\rangle_A^{\psi}\right\rangle^{\Psi}$ smooth approximation of the Hausdorff distance

$\overrightarrow{n}(y)$ (unit vector) normal to the current curve at point y

$\kappa(y)$ curvature of the current curve at point y

$\xi(a)$ $1 - a\,\dfrac{\Psi''(a)}{\Psi'(a)}$

$$
\begin{aligned}
\mathbf{U}(y,x)(\delta y)(\delta x) &= D_x\left[\overrightarrow{\mathcal{D}}(y,x)\cdot\overrightarrow{n}(y)\ \ \overrightarrow{n}(y)\cdot\delta y\right]\cdot\delta x \quad \text{(symmetric)}\\
&= \phi''(d(x,y))\left(\frac{x-y}{d(x,y)}\cdot\delta x\right)\left(\frac{y-x}{d(x,y)}\cdot\delta y\right)\\
&\quad + \frac{\phi'(d(x,y))}{d(x,y)}\left[-\delta x\cdot\delta y \ + \left(\frac{x-y}{d(x,y)}\cdot\delta x\right)\left(\frac{y-x}{d(x,y)}\cdot\delta y\right)\right]
\end{aligned}
$$

11.3.2 First order derivative

The first order derivative of a functional with respect to a curve A is a vector field defined on this curve, that is, an application that associates to each point y of curve A a vector. Note we consider here the derivative of the square of d_H. We recall its expression:

$$\begin{aligned}\nabla_A d_H^2(A,B) \;=\; & \frac{1}{2|A|}\frac{d_H(A,B)}{\Psi'(d_H(A,B))}\;\Big[\;\Theta_{B\phi A\psi}\;\;\sigma_A\left(\psi_{B\phi}(\cdot),\;y\right)\;\kappa(y)\overrightarrow{n}(y)\\ & +\;\Big\langle\;\left(\Phi_{B\phi}(y)\;\Theta_{B\phi A\psi}\;+\;\Phi_{A\phi}(x)\;\Theta_{A\phi B\psi}\right)\;\overrightarrow{D}(y,x)\cdot\overrightarrow{n}(y)\Big\rangle_{x\in B}\;\overrightarrow{n}(y)\\ & +\;\Theta_{A\phi B\psi}\;\left\langle\;\Phi_{A\phi}(x)\;\;\sigma_A\left(\phi\circ d(x,\cdot),\;y\right)\right\rangle_{x\in B}\;\kappa(y)\overrightarrow{n}(y)\;\;\Big]\end{aligned}$$

11.3.3 Second order derivative

The second order cross-derivative $\partial_B\left[\partial_A\, d_H^2(A,B)\right](\delta A)(\delta B)$ is an application that associates a real value to any couple of two fields δB and δA defined respectively on B and A. This application is supposed to be symmetric ($\partial_A\partial_B=\partial_B\partial_A$ and $d_H(A,B)$ is symmetric).

It can be rewritten as:

$$\partial_B\left[\left\langle\,\nabla_A\, d_H^2(A,B)(y)\right|\,\delta A(y)\rangle_{y\in A}\right](\delta B)$$

hence

$$\partial_B\left[\nabla_A\, d_H^2(A,B)\right](y)$$

is an application that associates to any field δB a field defined on A.

With a slight abuse of notations:

$$\nabla_B\left[\nabla_A\, d_H^2(A,B)\right](y)(z)$$

is a kind of super-matrix which for any two points y on A and z on B is a 2×2 matrix. For two fields δA and δB,

$$\nabla_B\left[\nabla_A d_H^2(A,B)\right](y)(z)\,(\delta A(y))\,(\delta B(z))$$

is a real value for each (y,z). In the sequel, $\nabla_B\left[\nabla_A\, d_H^2(A,B)\right](y)(z)$ will be proportional to the tensor product $\overrightarrow{n}(y)\;\overrightarrow{n}(z)$, where $\overrightarrow{n}(y)$ is to be thought as $\overrightarrow{n}(y)\cdot\delta A(y)$, and $\overrightarrow{n}(z)$ as $\overrightarrow{n}(z)\cdot\delta B(z)$.

Theorem 74. *With the previous notations, the second order cross-derivative of the approximation of the Hausdorff distance is:*

$$
\begin{aligned}
&\nabla_B \left[\nabla_A \, d_H^2(A,B)\right](y)(z) \\
&\quad = \\
&\xi(d_H(A,B)) \quad \nabla_B \left[d_H(A,B)\right](z) \quad \nabla_A \left[d_H(A,B)\right](y) \\
&+ \quad \frac{1}{2|A| \; |B|} \; \frac{d_H(A,B)}{\Psi'(d_H(A,B))} \times \Bigg[\\
&\qquad \Theta_{A\phi B\psi} \; \Phi_{A\phi}(z) \;\; \sigma_A\left(\overrightarrow{D}(z,\cdot),\; y\right) \cdot \overrightarrow{n}(z) \; \overrightarrow{n}(z) \quad \kappa(y)\overrightarrow{n}(y) \\
&\qquad + \; \Theta_{A\phi B\psi} \;\; \sigma_B\left(\Phi_{A\phi}(\cdot) \; \sigma_A\left(\phi \circ d(\cdot,\cdot),\; y\right),\; z\right) \; \kappa(z)\overrightarrow{n}(z) \quad \kappa(y)\overrightarrow{n}(y) \\
&\qquad + \; \Theta_{A\phi B\psi} \left(\frac{\Phi'}{\phi'}\right)_{A\phi}(z) \quad \sigma_A\left(\phi \circ d(z,\cdot),\; y\right) \left\langle \overrightarrow{\mathcal{D}}(z,\cdot) \right\rangle_A \cdot \overrightarrow{n}(z) \; \overrightarrow{n}(z) \quad \kappa(y)\overrightarrow{n}(y) \\
&\qquad + \; \left(\frac{\Theta'}{\psi'}\right)_{A\phi B\psi} \Phi_{A\phi}(z) \; \left\langle \Phi_{A\phi}(\cdot) \;\; \sigma_A\left(\phi \circ d(\cdot,\cdot),\; y\right)\right\rangle_B \left\langle \overrightarrow{\mathcal{D}}(z,\cdot) \right\rangle_A \cdot \overrightarrow{n}(z) \; \overrightarrow{n}(z) \quad \kappa(y)\overrightarrow{n}(y) \\
&\qquad + \; \left(\frac{\Theta'}{\psi'}\right)_{A\phi B\psi} \sigma_B\left(\psi_{A\phi}(\cdot),\; z\right) \; \left\langle \Phi_{A\phi}(\cdot) \;\; \sigma_A\left(\phi \circ d(\cdot,\cdot),\; y\right)\right\rangle_B \quad \kappa(z)\overrightarrow{n}(z) \quad \kappa(y)\overrightarrow{n}(y) \\
&\qquad + \; \Theta_{A\phi B\psi} \left(\frac{\Phi'}{\phi'}\right)_{A\phi}(z) \quad \left\langle \overrightarrow{\mathcal{D}}(z,\cdot) \right\rangle_A \cdot \overrightarrow{n}(z) \; \overrightarrow{n}(z) \quad \overrightarrow{\mathcal{D}}(y,z) \cdot \overrightarrow{n}(y) \; \overrightarrow{n}(y) \\
&\qquad + \; \Theta_{A\phi B\psi} \;\; \sigma_B\left(\Phi_{A\phi}(\cdot) \; \overrightarrow{\mathcal{D}}(y,\cdot),\; z\right) \cdot \overrightarrow{n}(y) \; \overrightarrow{n}(y) \;\; \kappa(z)\overrightarrow{n}(z) \\
&\qquad + \; \left(\frac{\Theta'}{\psi'}\right)_{A\phi B\psi} \Phi_{A\phi}(z) \; \left\langle \Phi_{A\phi}(\cdot) \;\; \overrightarrow{\mathcal{D}}(y,\cdot) \right\rangle_B \cdot \overrightarrow{n}(y) \; \overrightarrow{n}(y) \;\; \left\langle \overrightarrow{\mathcal{D}}(z,\cdot) \right\rangle_A \cdot \overrightarrow{n}(z) \; \overrightarrow{n}(z) \\
&\qquad + \; \left(\frac{\Theta'}{\psi'}\right)_{A\phi B\psi} \sigma_B\left(\psi_{A\phi}(\cdot),\; z\right) \; \left\langle \Phi_{A\phi}(\cdot) \;\; \overrightarrow{\mathcal{D}}(y,\cdot) \right\rangle_B \cdot \overrightarrow{n}(y) \; \overrightarrow{n}(y) \;\; \kappa(z)\overrightarrow{n}(z) \\
&\qquad + \; \Theta_{A\phi B\psi} \; \Phi_{A\phi}(z) \; \mathbf{U}(y,z) \left(\overrightarrow{n}(y)\right) \left(\overrightarrow{n}(z)\right) \\
&\quad + \; \textbf{symmetric term } (A \mapsto B, \quad y \mapsto z) \\
&\Bigg].
\end{aligned}
$$

11.3.4 Calculi

In order to compute $\nabla_B \left(\nabla_A d_H^2(A,B)(y)\right)(z)$, where z is a point of the curve B, the following derivatives have been used:

$$\nabla_B \left(\frac{1}{2|A|} \frac{d_H(A,H)}{\Psi'(d_H(A,B))} \right) = \frac{1}{2|A|} \frac{\Psi'(d_H) - \Psi''(d_H)}{\Psi'(d_H)^2} \ \nabla_B d_H(A,B)$$

$$\nabla_B \left(\langle f \rangle_B \right)(y) = \frac{1}{|B|} \left[\sigma_B(f,y)\kappa(y) \ + \ f'(y) \cdot \mathbf{n}(y) \right] \ \mathbf{n}(y)$$

$$\nabla_B \left(\langle f(B,\cdot) \rangle_B \right)(y) = \langle \nabla_1 f(B,x)(y) \rangle_{x \in B} \ + \ \frac{1}{|B|} \left[\sigma_B(f(B,\cdot),y)\kappa(y) \ + \ \nabla_2 f(B,y) \cdot \mathbf{n}(y) \right] \ \mathbf{n}(y)$$

$$\nabla_B \left(\langle d(\cdot,y) \rangle_{B\phi} \right)(z) = \frac{1}{|B|\phi'(\langle d(\cdot,y) \rangle_{B\phi})} \left[\sigma_B \left(\phi \circ d(\cdot,y), z \right) \kappa(z) + \mathbf{U}(z,y) \cdot \mathbf{n}(z) \right] \ \mathbf{n}(z)$$

$$\nabla_B \left(\Phi_{B\phi}(y) \right)(z) = \Phi' \left(\langle d(\cdot,y) \rangle_{B\phi} \right) \ \nabla_B \left(\langle d(\cdot,y) \rangle_{B\phi} \right)(z)$$

$$\begin{aligned} \nabla_B \left(\Theta_{B\phi A\psi} \right)(z) \ &= \ \frac{\Theta'}{\psi'} \left(\left\langle \langle d \rangle_{B\phi} \right\rangle_{A\psi} \right) \ \times \\ & \left\langle \frac{1}{|B|} \Phi_{B\phi}(y) \ \left[\ \sigma_B \left(\phi \circ d(\cdot,y), z \right) \kappa(z) \ + \ \mathbf{U}(z,y) \cdot \mathbf{n}(z) \ \right] \right\rangle_{y \in A} \mathbf{n}(z) \end{aligned}$$

$$\begin{aligned} \nabla_B \left(\Theta_{A\phi B\psi} \right)(z) \ &= \ \frac{\Theta'}{\psi'} \left(\left\langle \langle d \rangle_{A\phi} \right\rangle_{B\psi} \right) \ \times \\ & \frac{1}{|B|} \left[\sigma_B \left(\psi \left(\langle d(\cdot,\cdot) \rangle_{A\phi} \right), z \right) \kappa(z) \ + \ \Phi_{A\phi}(z) \left\langle \mathbf{U}(z,y) \cdot \mathbf{n}(z) \right\rangle_{y \in A} \right] \mathbf{n}(z) \end{aligned}$$

$\mathbf{U}$ is a derivative: $\mathbf{U}(y,x) \cdot (\delta y) = \nabla_y \left(\phi(d(y,x)) \right)(\delta y)$ so $\nabla_x \left(\mathbf{U}(y,x) \right)$ is a second order (symmetric) derivative:

$$\begin{aligned} \nabla_x \left(\mathbf{U}(y,x) \right)(\delta x)(\delta y) \ &= \ \phi''(d(x,y)) \left(\frac{x-y}{d(x,y)} \cdot \delta x \right) \left(\frac{y-x}{d(x,y)} \cdot \delta y \right) \\ & + \frac{\phi'(d(x,y))}{d(x,y)} \left[-\delta x \cdot \delta y \ + \ \left(\frac{x-y}{d(x,y)} \cdot \delta x \right) \left(\frac{y-x}{d(x,y)} \cdot \delta y \right) \right] \end{aligned}$$

$$\nabla_x \left(\sigma_A \left(\phi(d(x,\cdot)), y \right) \right) = \langle \mathbf{U}(x,z) \rangle_{z \in A} - \mathbf{U}(x,y) = \sigma_A \left(\mathbf{U}(x,\cdot), y \right)$$

11.4 Examples of Segmentation with Shape Priors

A few experiments are now computed in order to validate some of the shape priors introduced in section 11.2.2.

11.4.1 Rigid registration of a fixed shape

A set of 12 images of sea-stars[1] (see figure 11.1) has been segmented by hand. This could be automatized in the general case if the examples of the sample set are chosen so that they are easy images to segment with usual algorithms. The sea-stars have then be registered so that their center of mass and average radius coincide, that is to say that they have been scaled and translated automatically to predefined values of these two quantities.

The mean curve of the set of sea-stars has been computed by estimating the best rigid deformation for each sea-star (translation, scaling, but rotation also) and moving it accordingly (as in section 11.2.3) while computing the mean shape of the thus-rigidly-warped sea-stars (as in section 9.1). More quantitatively, to each sample S_i a rigid motion R_i with parameters $(\mathbf{t}_i, \theta_i, s_i)$ is associated. Each R_i is initialized so that the center of mass and the mean radius of the warped shape $R_i(S_i)$ are predefined values, and the angle parameter θ_i is initialized to 0. The mean shape M is arbitrary initialized (experience seems to confirm that the result does not depend on this initialization), for instance as being the first example S_1. The criterion, where E stands for the approximation of the Hausdorff distance,

$$\sum_i E^2(M, R_i(S_i))$$

is minimized (by gradient descent) at the same time with respect to M and to all parameters $(\mathbf{t}_i, \theta_i, s_i)_{i \in I}$.

At convergence of the mean shape M, the computed rigid motions R_i associated to each sample are deliberately left out, while all these translation, rotation and scaling parameters could have been useful if one had wished to express statistics about the location, orientation and size of sea-stars in images. Here however the algorithm is designed to be invariant with respect to rigid motion.

The shape prior in the image segmentation task is chosen to be equal to $E^2(M, R(C))$, where C is the current evolving shape and R a rigid motion. The shape C and the parameters of R are estimated simultaneously within a framework similar to the previous one.

There exist many approaches to image segmentation in the computer vision literature, for instance the Mumford-Shah functional [83, 82], geodesic active contours [14, 16, 15], or region histograms [92, 12, 61, 89, 90, 93, 88].

A new image of a sea-star is segmented using a region intensity histogram criterion with and without this shape prior (see figure 11.2). More precisely, as the image has "naturally" 256 grey levels, the intensity histograms of the inside h_I and outside h_O of the contour C are real-valued functions of an integer which can have 256 possible values. They associate to any grey level the number of such colored pixels in the corresponding region. The two histograms are supposed to be relatively homogeneous and as much different one from the other as possible. The segmentation criterion to minimize which is considered here is a weighted sum of the length $|C|$ of the contour C plus the correlation between a slightly

[1] found via Google Images

Figure 11.1: The learning set of sea-stars.

smoothed version of the intensity probability distributions of the two regions $p_I = h_I/|I|$ and $p_O = h_O/|O|$ (where $|I|$ is the area of the region I):

$$\frac{1}{256} \sum_{a \in [\![0,255]\!]} \left(G_\sigma \star p_I\right)(a) \ \left(G_\sigma \star p_O\right)(a) \quad + \quad |C|$$

where $G_\sigma\star$ is the Gaussian smoothing with parameter σ.

In figure 11.2, the location of the sea-star is found, but of course the shape variability of the sample set has not been taken into account. Consequently, either the result of the segmentation is exactly the mean without any non-rigid deformation (if the weight of the shape prior is huge), or it is the mean plus any small deformation (without any priors on this deformation).

This algorithm is simplistic since the prior is a fixed shape (up to rigid motion), but it helps finding a not-too-varying object as well as an occluded, shape-fixed object. Note also that the rigid registration process could be partly handled by the use of the rigidified gradient.

11.4.2 Parzen method with the Hausdorff distance

To illustrate the Parzen method, three pictures (figure 11.4) of a same object under slightly different points of view have been segmented by hand and the obtained contours have been considered as a (tiny) sample set of examples. A new picture of the same object under a new point of view is then segmented (see figure 11.5) using a criterion based on region intensity histograms, with or without the shape prior previously introduced as inspired from the "Parzen method".

The segmentation of the new image without the shape prior lacks the mouse's contour because the lightest gray levels inside the mouse are indeed lighter than the darkest gray levels in the remaining part of the image. An approach based on the intensity gradient may have led to better results in this case. The segmentation with the same initialization and intensity criterion but with the Parzen-like shape prior is quite good. The method works well here because the contour to find is not very far from one of the given examples. However if a higher shape variability has to be taken into account, this method requires to be previously given a dense enough sample of examples. Obviously this is problematic when the intrinsic dimension of the shape variability gets higher. This is due to the fact that, in the best case, this method is only able to build a local mean between examples (if the parameter σ and the weight between the two terms of the segmentation criterion are well suited to). Consequently in the case of sparse data it cannot infer from the observed deformations new combinations of them. Depending on the aim of the user, this can be seen as a good point or as a limitation.

In comparison, the eigenmode method (computed next) has the advantage that it can deal with sparser data. If the purpose is to depict accurately the observed sample, then the ICA method with component histograms may show itself useful; but if the aim is to infer as most as possible from a non-exhaustive sample, then the eigenmode decomposition also provides an appropriate framework by potentially allowing combinations of observed deformations. For example this is the case with the Gaussian distribution assumption.

Another drawback of the Parzen method is that is the data sampling is dense, then the algorithm will be slow since it requires at each time step the successive consideration of each sample, whereas in the Gaussian eigenmodes approach, the number of considered modes can be easily limited if necessary.

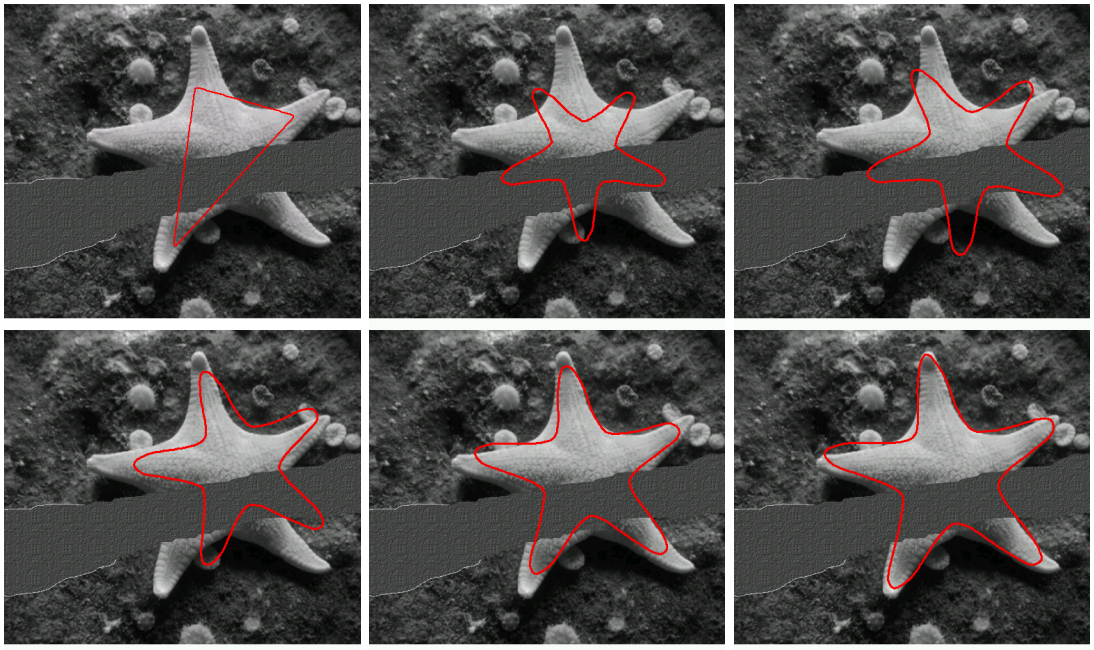


Figure 11.2: Top left: any reasonable initialization for the region intensity histogram criterion. A part of the image has been erased in order to increase the difficulty. Top middle: automatic change for the mean at the same location with similar size. Rest: some steps of the segmentation process with knowledge of the mean shape. This rigid criterion finds the location of sea-star but lacks information about how to adapt the final shape. Indeed for a smaller weight of the prior, the result will include the small white balls connected to the sea-star, as in figure 11.3.

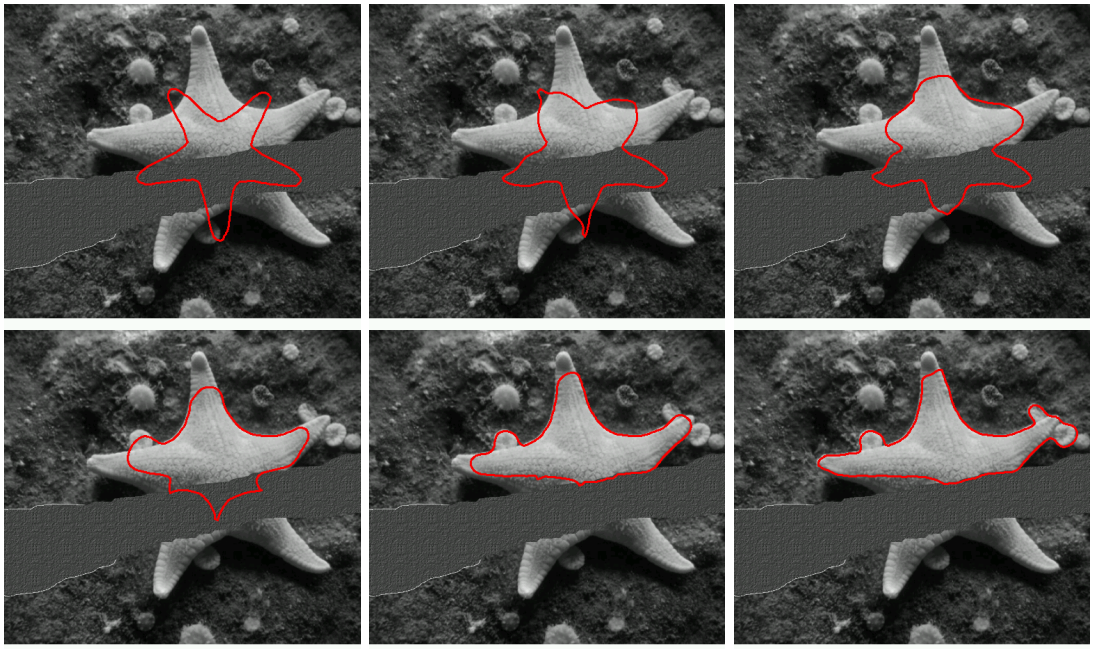


Figure 11.3: Top left: same initialization as before. Rest: some steps of segmentation process without any shape prior, for the same region criterion. The result lacks the global shape of the sea-star and include the small white balls.

Figure 11.4: The sample set of pictures whose extracted contours are given to the algorithm.

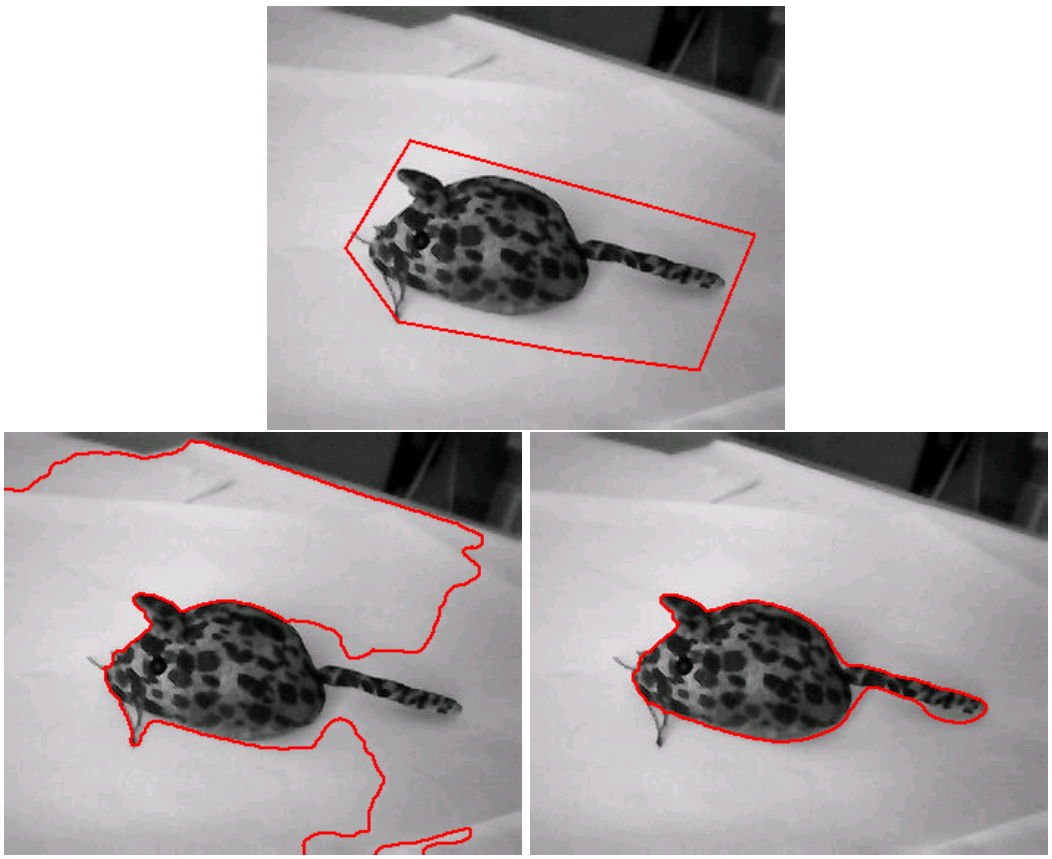


Figure 11.5: Top: new picture to be segmented and initial contour. Left: result of the image segmentation for a criterion based on region histograms. Right: result of the segmentation for the same initialization and the same criterion based on histograms, but also with the Parzen-like shape prior.

11.4.3 Gaussian Eigenmodes

The mean shape computed in section 11.4.1 is considered again. But this time, at convergence, the mean shape M as well as the instantaneous deformation fields $\partial_M E^2(M, R_i(S_i))$ from M to each of the finally-rigidly-warped examples are constituents of the shape prior, while possible statistics on the rigid motions R_i are still ignored. The Gaussian eigenmodes β_k related to the instantaneous deformation fields are computed. The shape prior that is used for these experiments is the square of the Mahalanobis distance.

Starting with a toy example to show the strength of this approach, a small set of four similar rectangles with two kinds of outgrowths is considered. For this particular example, the mean shape was computed without simultaneous registration of the samples and there was no optimization concerning rigid motion between the mean shape and the evolving shape during the segmentation process, that is to say that the mean had a fixed location. Indeed this toy example is not designed to show how the algorithm could be invariant to rigid motion, but to show how the shape variability can be taken into account. The segmentation of a new rectangle that combines the two outgrowths as well as a third new one leads to a shape which can be described as a new combination of the two already observed outgrowths, ignoring the third one (figure 11.6).

One interesting point of this method is that one does not need to choose the number of relevant eigenmodes since the importance of each of them is naturally described by its associated standard deviation. In order to show the interest of these characteristic modes of deformation, the standard deviation associated to the noise has been chosen to be equal to 5.10^{-3} times the standard deviation associated to the highest eigenmode, that is to say that a "noisy" deformation field with null component on each mode will cost 200 times more than a field of same norm but collinear to the first mode. In order to be coherent, if there exist eigenmodes with eigenvalues smaller than the one associated to the noise, then they have to be forgotten and considered as noise. In particular in the case of figure 11.6, the first two eigenmodes were found to have nearly the same eigenvalue and the two others were about a hundred times smaller, and indeed the segmentations with all modes or only the first two modes were the same.

Independently on the value of the weight of the shape prior providing this one is high enough, the qualitative behavior of the shape prior on the evolution is to "project" the evolving shape onto a linear combination of the eigenmodes, in the sense that the gradient of E from the mean M to the evolving shape C will progressively reduce its components on eigenmodes (and remaining noisy part) according to their standard deviation. As the distribution is here supposed to be Gaussian, any increasing of the weight of the shape prior will make the result a little nearer to the mean shape. However in the case of component histograms as proposed in section 11.2.2 this phenomenon would not appear since the shape with highest *a priori* probability is not the mean anymore.

We go back to the sea-star example presented in section 11.4.1. The mean and the first eigenmodes of this set are displayed on figure 11.7. The results of the segmentation of the original image for different shape priors are shown in figure 11.8. On the top row are shown typical results for the segmentation criterion without any shape prior, for various initializations. As the region based criterion has many local minima, the segmentation result depends strongly on the initialization. Not that the small white balls around the sea-star are difficult to distinguish from the sea-star, and the shaded regions of the sea-star have colors similar to the background. Then, on the middle row (left handside) is shown the result obtained with a rigid registration of the mean shape (as in figure 11.2). This result is much

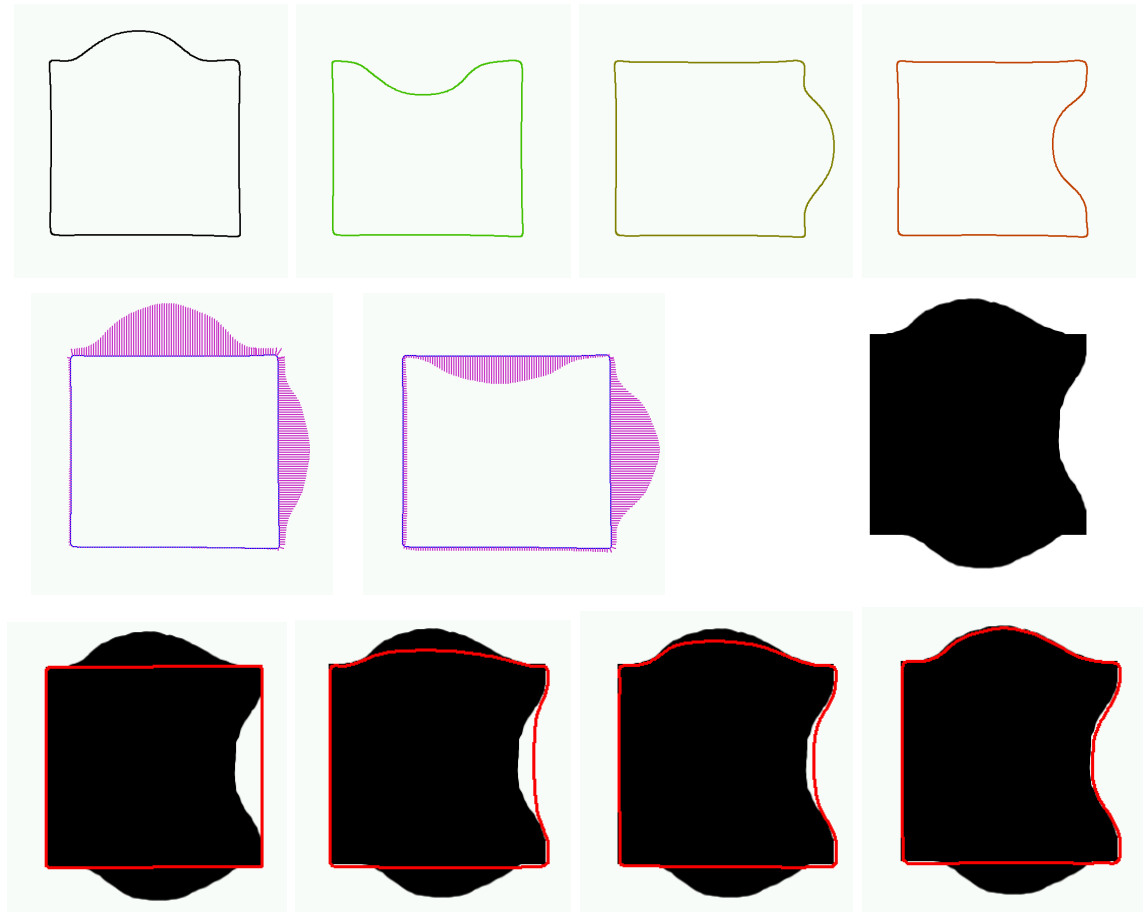


Figure 11.6: Top row: the four shapes that compose the sample set. Middle, left: the mean of these shapes (in blue), with the two first eigenmodes (in purple), successively. Middle, right: the new image to be segmented, built approximately as a new combination of the previously observed deformations added to a new non-observed deformation. Bottom: segmentation with the only knowledge of the mean, eigenmodes and eigenvalues, under the Gaussian distribution assumption which is obviously not satisfied (left: initialization; middle: some steps of the evolution; right: result at convergence).

more "stable" in the sense that it can be obtained from any reasonable initialization. The reason for this is that the minimization is processed with respect to only few parameters (translation, orientation, scaling) instead of a whole shape (which is infinite-dimensioned). Therefore in practice the space to be explored in order to find the solution is much smaller in the case of rigid registration. In order to allow some deformations around the registered mean shape, we start from the result of the rigid registration, and minimize the sum of the square of the distance to the mean shape and of the region histogram criterion. This minimization is computed with respect to both the evolving shape and the location parameters of the mean. For high values of the weight of the shape prior, the result is of course close to the one obtained by only rigid registration. For low values, the result is close to the one obtained without shape prior, that is to say to one similar to the ones on the top row. A typical example is shown (middle row, right handside) for a middle value of the weight:

as there is no prior on the deformations applied to the mean shape, the algorithm leads to outgrowths that are non-sensible for a sea-star. For instance it includes the white balls within the sea-star and let a deformation grow far inside the sea-star in order to get rid of the shadow regions.

Finally, still starting from the result of the rigid registration, we minimize the sum of the Gaussian eigenmode shape prior and of the region histogram criterion with respect to both the evolving shape (in red) and the location parameters of the mean shape (in blue). The deformations that are required for the inclusion of the white balls within the sea-star have a heavy cost for the shape prior since they are not characteristic deformations of the mean shape. Consequently the algorithm finds globally the shape of the sea-star, except a part of its shadowed regions which the region intensity criterion considers as included into the background (see top row). Note however that the deformations due to the shadow have been described as best as possible as resulting from a combination of eigenmodes. Therefore they have been reduced to a reasonable deformation that a sea-star can undergo, and they look far better than the observed ones in the other segmentation results.

For all presented examples, the added computational cost due to the shape prior is reasonable. In these experiments, the total time cost with the prior was found to be about three to four times the total time cost without prior. Most of the time cost is due to the computation of a double integration needed by the approximation of the Hausdorff distance. As these integrations approximate the infimum or the supremum of a function defined on the shape that depends very strongly on the Euclidean distance to particular point of interest, the time cost could be reduced by estimating which small part of the shape has a non-negligible weight in the integration.

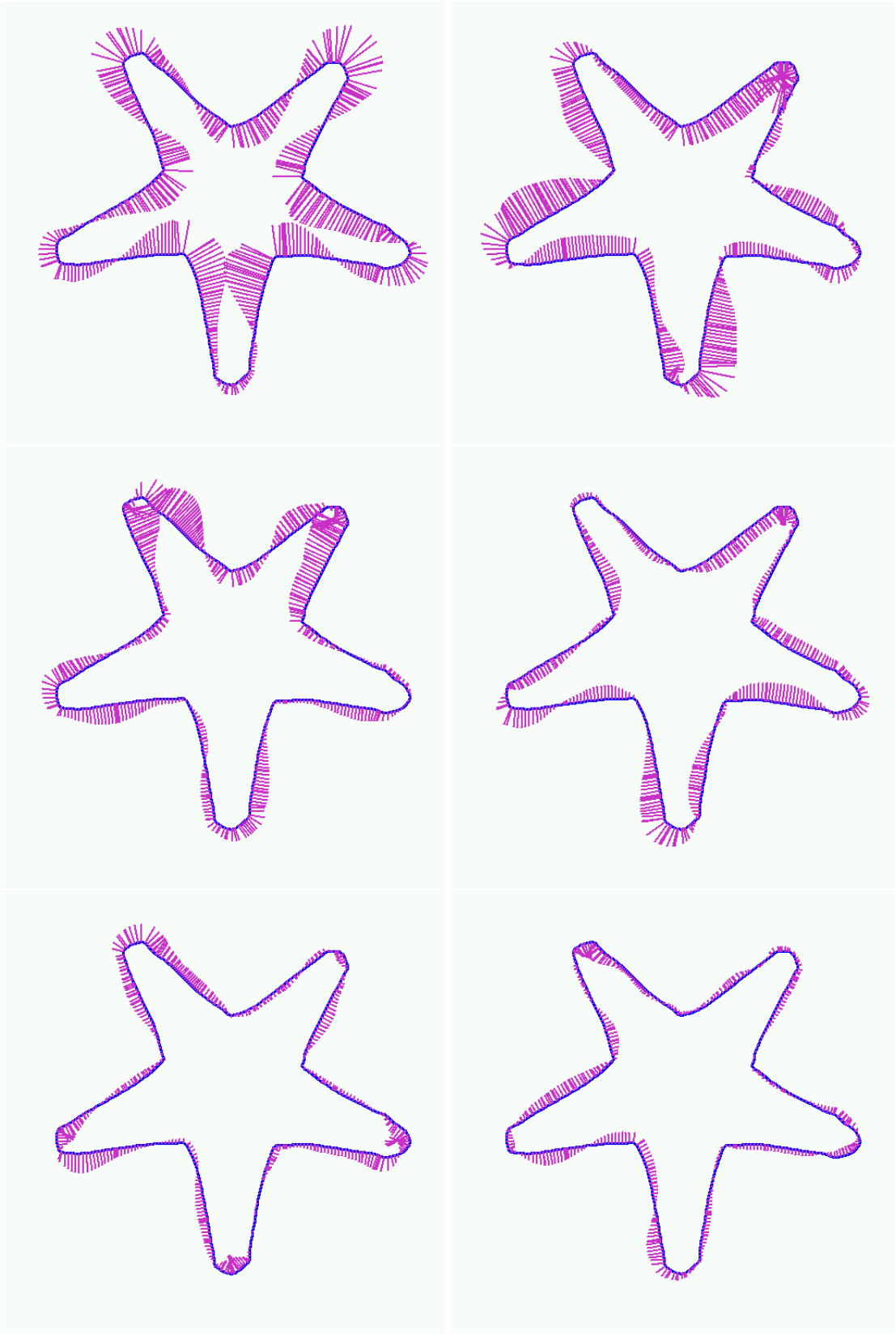


Figure 11.7: The mean of the set of sea-stars with its first six eigenmodes.

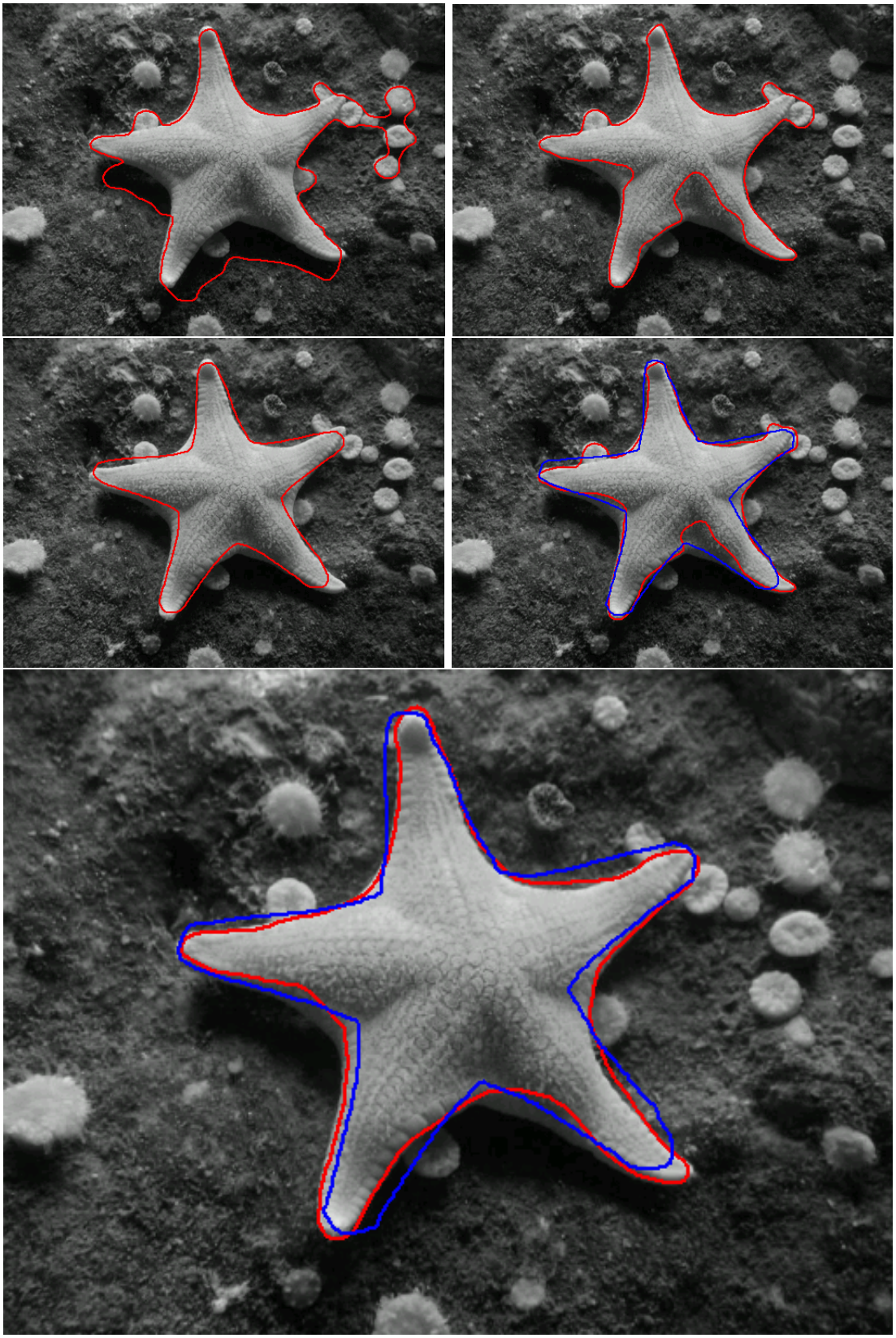


Figure 11.8: Segmentation results for different priors. See text for details. Top row: two examples of results obtained without any shape prior, for two different initializations. Middle row, left: result of rigid registration. Middle row, right: result of the segmentation (in red) with the (non-rigid) mean shape prior (whose estimated location is also shown, in blue). Bottom: result (in red) for the Gaussian eigenmode prior (with estimated location of the mean in blue).

Discussion

We have shown that it is possible to build shape statistics based on instantaneous deformation fields and to include them into a rigid-motion-invariant shape prior for image segmentation at a reasonable cost.

In the first part *Shapes and Distances* the topologies related to three usual shape metrics were shown to be equivalent on the set $\mathcal{S}$ of smooth shapes with lower-bounded Federer reach. In practice during shape warping the evolving shape generally remains in this set $\mathcal{S}$, except when a topological change occurs, that is to say when the number of connected components of the shape changes. The question of the equivalence of the three topologies at these precise moments is still addressed since, even if the topologies are generally not equivalent outside $\mathcal{S}$, there might exist a large subset including part of these moments of interest for which the equivalence would be still satisfied.

In the next chapter a family of differentiable approximations of the Hausdorff distance was proposed and shown to converge towards the Hausdorff distance when its parameters get closer to the infinite. The gradient of this approximation is an infinitesimal deformation field and is the fundamental tool in most of the thesis, except in chapter 4 which is an attempt to classify shapes with the only knowledge of the mutual distances. Several similar techniques such as the graph Laplacian lead to a low-dimensional representation of a set of shapes, but the quality of the result strongly depends on the sample density. The application of such "static" approaches in a segmentation task is still addressed. One notable drawback is that the addition of a new shape into an already drawn map requires the whole re-computation of the map. Another drawback is that the intuitive segmentation criterion related to a graph-based approach is nothing but one very close to the "Parzen method" which does not require the computation of the map. The design of a suitable, non-trivial criterion based on such maps and its differentiation with respect to an evolving curve still need to be studied.

The warping and statistical framework introduced thereafter is not specific to the Hausdorff distance. Other distances (or approximations of them) can be considered, although the most often used ones in the computer vision literature were shown to have equivalent topologies. In particular an extension of the Hausdorff distance which takes into account local shape descriptors has been proposed. However the search for the best such descriptors, as well as the implementation and validation of this extension, have still to be done.

In the second part *Shape Warping*, the notion of differentiation in $\mathcal{S}$ was detailed, as well as the ones of shape gradient and intrinsic gradient. The differentiation tools allowed to warp a shape onto another one by minimizing the chosen distance by gradient descent. The approximation of the Hausdorff distance was found to lead to sensible warpings, except in the presence of non-registered large narrow oscillations (like moving fingers on a hand) on the target shape. The usual L^2 inner product in the tangent space of each shape was changed in order to favor spatially coherent motions, such as rigid or smooth motions, which helped much the matching process while keeping the mathematical framework of the gradient descent. The choice of the inner product was seen as a prior on the deformation flow. As a change of inner product can be expressed as a linear transform of the usual gradient, an extension of the definition of the gradient was given in order to take non-linear priors into account. An example with locally rigid motions was given. Although the results were satisfying, the particular related prior happened to be difficult to minimize in practice by variational method. The design of more stable similar priors or the search for other minimization methods such as graph cuts for the proposed criterion should be investigated. More generally, the incorporation of linear or non-linear flow priors into the gradient descent framework can probably lead to interesting further developments.

The third part *Shape Statistics and Priors* introduced a definition of the mean of a sample set of shapes and its characteristic modes of variations via PCA of the observed instantaneous deformation fields. The mean thus introduced is not necessarily unique; its uniqueness is an open question. The modes of variations were then included in an image segmentation prior, and the shape variability that they convey showed itself useful to improve classical segmentation tools. A way to build statistics on images was also presented. Based on diffeomorphic matching, the definition of the mean image and of the eigenmodes is similar to the previous ones, except that the image intensity has to be taken into account. Linear combinations of these modes are however not guaranteed to be diffeomorphisms. A way to deal with that would be to build statistics on the associated infinitesimal deformation fields (the "logarithm" of the diffeomorphisms) instead of the diffeomorphisms themselves. However the point is that, in an image classification task, SVM gave better results on diffeomorphisms (without even intensity variations) than on the raw images. Consequently the relevant information (or at least a relevant part of it) stands in these computed diffeomorphisms.

Statistics on images and statistics on shapes could be incorporated together in a same framework. For example it might be interesting to study how a set of images could be segmented simultaneously *while* the images are themselves warped in order to compute their mean. The thus obtained statistics could then be used in both segmenting a new image and warping it in order to classify it at the same time, with the hope that each of the two processes would help the other.

Bibliography

[1] D. Adalsteinsson and J.A. Sethian, *The fast construction of extension velocities in level set methods*, Journal of Computational Physics **148** (1999), no. 1, 2–22.

[2] N.I. Akhiezer and I.M. Glazman, *Theory of linear operators in hilbert space*, Pitman, 1981.

[3] L. Ambrosio and H.M. Soner, *Level set approach to mean curvature flow in arbitrary codimension*, Journal of Differential Geometry **43** (1996), 693–737.

[4] M. Belkin and P. Niyogi, *Semi-supervised learning on manifolds*, Journal of Machine Learning Research (2003).

[5] Mikhail Belkin and Partha Niyogi, *Laplacian Eigenmaps for Dimensionality Reduction and Data Representation*, Neural Computation **15** (2003), 1373–1396.

[6] M. Bertalmío, L.T. Cheng, S. Osher, and G. Sapiro, *Variational problems and partial differential equations on implicit surfaces*, Journal of Computational Physics **174** (2001), no. 2, 759–780.

[7] M. Bertalmio, G. Sapiro, L.T Cheng, and S. Osher, *Variational problems and PDE's on implicit surfaces*, IEEE Workshop on Variational and Level Set Methods (Vancouver, Canada) (IEEE, ed.), July 2001, pp. 186–193.

[8] A. Blake and M. Isard, *Active Contours*, Springer–Verlag, 1998.

[9] J.F. Bonnans, J.C. Gilbert, C. Lemarechal, and C.A. Sagastizabal, *Numerical optimization: Theoretical and practical aspects*, Springer-Verlag, 2002.

[10] F.L. Bookstein, *Size and shape spaces for landmark data in two dimensions*, Statistical Science **1** (1986), 181–242.

[11] Y. Boykov and V. Kolmogorov, *Computing geodesics and minimal surfaces via graph cuts*, International Conference on Computer Vision, vol. 1, 2003, pp. 26–33.

[12] Thomas Brox, Mikael Rousson, Rachid Deriche, and Joachim Weickert, *Unsupervised segmentation incorporating colour, texture, and motion*, 10th International Computer Analysis of Images and Patterns, Lecture Notes on Computer Science., no. 2756, Springer Verlag, August 2003, pp. 353–360.

[13] T.K. Carne, *The geometry of shape spaces*, Proc. of the London Math. Soc. **3** (1990), no. 61, 407–432.

[14] V. Caselles, R. Kimmel, and G. Sapiro, *Geodesic active contours*, The International Journal of Computer Vision **22** (1997), no. 1, 61–79.

[15] V. Caselles, R. Kimmel, G. Sapiro, and C. Sbert, *3d active contours*, Images,Wavelets and PDEs (M.O. Berger, R. Deriche, I. Herlin, J. Jaffre, and J.M. Morel, eds.), Lecture Notes in Control and Information Sciences, vol. 219, Springer, June 1996, pp. 43–49.

[16] ———, *Minimal surfaces based object segmentation*, IEEE Transactions on Pattern Analysis and Machine Intelligence **9** (1997), no. 4, 394–398.

[17] G. Charpiat, O. Faugeras, and R. Keriven, *Shape metrics, warping and statistics*, Proceedings of the International Conference on Image Processing, IEEE Signal Processing Society, 2003.

[18] ———, *Approximations of shape metrics and application to shape warping and empirical shape statistics*, Foundations of Computational Mathematics **5** (2005), no. 1, 1–58.

[19] ———, *Image statistics based on diffeomorphic matching*, 10th International Conference on Computer Vision (Beijing, China), 2005.

[20] G. Charpiat, O. Faugeras, R. Keriven, and P. Maurel, *Approximations of shape metrics and application to shape warping and empirical shape statistics*, Statistics and Analysis of Shapes, pp. 363–395, H. Krim and A. Yezzi, Jr., Birkhäuser, 2006.

[21] G. Charpiat, R. Keriven, J.P. Pons, and O. Faugeras, *Designing spatially coherent minimizing flows for variational problems based on active contours*, 10th International Conference on Computer Vision (Beijing, China), 2005.

[22] G. Charpiat, P. Maurel, R. Keriven, and O. Faugeras, *Distance-based shape statistics*, International Conference on Acoustics, Speech, and Signal Processing, Special Session: Statistical Inferences on Nonlinear Manifolds with Applications in Signal and Image Processing, 2006.

[23] G. Charpiat, P. Maurel, J.-P. Pons, R. Keriven, and O. Faugeras, *Generalized gradients: Priors on minimization flows*, The International Journal of Computer Vision (2007).

[24] Hyeong In Choi, Sung Woo Choi, and Hwan Pyo Moon, *Mathematical theory of the medial axis transform*, Pacific J. Math. **181** (1997), 57–88.

[25] G. Choquet, *Cours d'analyse*, vol. II, Masson, 1969.

[26] T. Cootes, C. Taylor, D. Cooper, and J. Graham, *Active shape models-their training and application*, Computer Vision and Image Understanding **61** (1995), no. 1, 38–59.

[27] D. Cremers, T. Kohlberger, and C. Schnörr, *Shape statistics in kernel space for variational image segmentation*, Pattern Recognition **36** (2003), no. 9, 1929–1943, Special Issue on Kernel and Subspace Methods in Computer Vision.

[28] D. Cremers, C. Schnoerr, and J. Weickert, *Diffusion-snakes: Combining statistical shape knowledge and image information in a variational framework*, 1st IEEE Workshop on Variational and Level Set Methods in Computer Vision, 2001.

[29] Daniel Cremers, Timo Kohlberger, and Christoph Schn orr, *Nonlinear shape statistics in mumford shah based segmentation*, ECCV, 2002, pp. 93–108.

[30] V. de Silva and J. Tenenbaum, *Global versus local methods in nonlinear dimensionality reduction*, 2003.

[31] M. C. Delfour and J.-P. Zolésio, *Shape analysis via oriented distance functions*, Journal of Functional Analysis **123** (1994), no. 1, 129–201.

[32] M.C. Delfour and J.-P. Zolésio, *Shape analysis via distance functions: Local theory*, Boundaries, interfaces and transitions, CRM Proc. Lecture Notes, vol. 13, AMS, Providence, RI, 1998, pp. 91–123.

[33] ———, *Shapes and geometries*, Advances in Design and Control, Siam, 2001.

[34] A. Dervieux and F. Thomasset, *A finite element method for the simulation of Rayleigh-Taylor instability*, Lecture Notes in Mathematics **771** (1979), 145–159.

[35] Jean Dieudonné, *Éléments d'analyse*, vol. I, Gauthier-Villars, Paris, 1969.

[36] M. P. DoCarmo, *Differential Geometry of Curves and Surfaces*, Prentice-Hall, 1976.

[37] David L. Donoho and Carrie Grimes, *When does isomap recover the natural parametrization of families of articulated images?*, TR 2002-27, Stanford University, 2002.

[38] ———, *Hessian eigenmaps: new locally linear embedding techniques for high-dimensional data*, TR 2003-08, Stantford University, 2003.

[39] I.L. Dryden and K.V. Mardia, *Statistical shape analysis*, John Wiley & Son, 1998.

[40] Y. Duan, L. Yang, H. Qin, and D. Samaras, *Shape reconstruction from 3D and 2D data using PDE-based deformable surfaces*, European Conference on Computer Vision, vol. 3, 2004, pp. 238–251.

[41] P. Dupuis, U. Grenander, and M. Miller, *Variational problems on flows of diffeomorphisms for image matching*, Quarterly of Applied Math. **56** (1998), 587–600.

[42] L.C. Evans, *Partial Differential Equations*, Graduate Studies in Mathematics, vol. 19, Proceedings of the American Mathematical Society, 1998.

[43] O. Faugeras and G. Hermosillo, *Well-posedness of two non-rigid multimodal image registration methods*, Siam Journal of Applied Mathematics **64** (2004), no. 5, 1550–1587.

[44] O. Faugeras and R. Keriven, *Variational principles, surface evolution, PDE's, level set methods and the stereo problem*, IEEE Transactions on Image Processing **7** (1998), no. 3, 336–344.

[45] H. Federer, *Hausdorff measure and Lebesgue area*, Proc. Nat. Acad. Sci. USA **37** (1951), 90–94.

[46] M. Fréchet, *L'intégrale abstraite d'une fonction abstraite d'une variable abstraite et son application à la moyenne d'un élément aléatoire de nature quelconque*, Revue Scientifique (1944), 483–512, 82ème année.

[47] ———, *Les éléments aléatoires de nature quelconque dans un espace distancié*, Ann. Inst. H. Poincaré **X** (1948), no. IV, 215–310.

[48] ———, *Les courbes aléatoires*, Bull. Inst. Internat. Statist. **38** (1961), 499–504.

[49] J. Glaunes, L. Younes, and A. Trouvé, *Diffeomorphic matching of distributions: A new approach for unlabelled point-sets and sub-manifolds matching*, International Conference on Pattern Recognition (Cambridge), 2004.

[50] B. Goldlücke and M. Magnor, *Space-time isosurface evolution for temporally coherent 3D reconstruction*, International Conference on Computer Vision and Pattern Recognition, vol. 1, 2004, pp. 350–355.

[51] ———, *Weighted minimal hypersurfaces and their applications in computer vision*, European Conference on Computer Vision, vol. 2, 2004, pp. 366–378.

[52] J. Gomes and O. Faugeras, *Reconciling distance functions and level sets*, Journal of Visual Communication and Image Representation **11** (2000), no. 2, 209–223.

[53] U. Grenander, *General pattern theory*, Oxford University Press, 1993.

[54] U. Grenander, Y. Chow, and D. Keenan, *HANDS: A pattern theoretic study of biological shapes*, Springer-Verlag, 1990.

[55] U. Grenander and M. Miller, *Computational anatomy: an emerging discipline*, Quart. Appl. Math. **56** (1998), no. 4, 617–694.

[56] P.L. Hallinan, G.G. Gordon, A.L. Yuille, P. Giblin, and D. Mumford, *Two- and Three-Dimensional Patterns of the Face*, A K Peters, 1999.

[57] E. G. Harding and D. G. Kendall (eds.), *Stochastic geometry*, ch. Foundation of a theory of random sets, pp. 322–376, John Wiley Sons, New-York, 1973.

[58] Xiaofei He, *Locality preserving projections*, Ph.D. thesis, Chicago, IL, USA, 2005, Adviser-Partha Niyogi.

[59] Xiaofei He, Deng Cai, and Wanli Min, *Statistical and computational analysis of locality preserving projection*, ICML '05: Proceedings of the 22nd international conference on Machine learning (New York, NY, USA), ACM Press, 2005, pp. 281–288.

[60] Gerardo Hermosillo, *Variational methods for multimodal image matching*, Ph.D. thesis, Université de Nice-Sophia Antipolis, May 2002.

[61] Stéphanie Jehan-Besson, Michel Barlaud, Gilles Aubert, and Olivier Faugeras, *Shape gradients for histogram segmentation using active contours*, Proceedings of the 9th International Conference on Computer Vision (Nice, France), IEEE Computer Society, IEEE Computer Society Press, 2003, pp. 408–415.

[62] H. Jin, S. Soatto, and A.J. Yezzi, *Multi-view stereo beyond Lambert*, International Conference on Computer Vision and Pattern Recognition, vol. 1, 2003, pp. 171–178.

[63] H. Karcher, *Riemannian centre of mass and mollifier smoothing*, Comm. Pure Appl. Math **30** (1977), 509–541.

[64] M. Kass, A. Witkin, and D. Terzopoulos, *Snakes: Active contour models*, The International Journal of Computer Vision **1** (1987), no. 4, 321–331.

[65] D.G. Kendall, *Shape manifolds, procrustean metrics and complex projective spaces*, Bulletin of London Mathematical Society **16** (1984), 81–121.

[66] ———, *A survey of the statistical theory of shape*, Statist. Sci. **4** (1989), no. 2, 87–120.

[67] W. Kendall, *Probability, convexity, and harmonic maps with small image i: uniqueness and fine existence*, Proc. London Math. Soc. **61** (1990), no. 2, 371–406.

[68] E. Klassen, A. Srivastava, W. Mio, and S.H. Joshi, *Analysis of planar shapes using geodesic paths on shape spaces*, IEEE Transactions on Pattern Analysis and Machine Intelligence **26** (2004), no. 3, 372–383.

[69] V. Kolmogorov and R. Zabih, *Multi-camera scene reconstruction via graph cuts*, European Conference on Computer Vision, vol. 3, 2002, pp. 82–96.

[70] ———, *What energy functions can be minimized via graph cuts?*, IEEE Transactions on Pattern Analysis and Machine Intelligence **26** (2004), no. 2, 147–159.

[71] S. Lafon, Y. Keller, and R.R. Coifman, *Data fusion and multi-cue data matching by diffusion maps*, IEEE transactions on Pattern Analysis and Machine Intelligence (2006).

[72] S. Lafon and A.B. Lee, *Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning and data set parameterization*, IEEE transactions on Pattern Analysis and Machine Intelligence **28** (2006), no. 9, 1393–1403.

[73] M. Leventon, E. Grimson, and O. Faugeras, *Statistical Shape Influence in Geodesic Active Contours*, Proceedings of the International Conference on Computer Vision and Pattern Recognition (Hilton Head Island, South Carolina), IEEE Computer Society, June 2000, pp. 316–323.

[74] Michael Leventon, *Anatomical Shape Models for Medical Image Analysis*, Ph.D. thesis, MIT, 2000.

[75] William E. Lorensen and Harvey E. Cline, *Marching cubes: A high resolution 3d surface construction algorithm*, Proceedings of the SIGGRAPH (Anaheim, CA) (M.C. Stone, ed.), July 1987, in Computer Graphics, Volume 21, Number 4, pp. 163–169.

[76] G. Matheron, *Random Sets and Integral Geometry*, John Wiley & Sons, 1975.

[77] P. Maurel, R. Keriven, and O. Faugeras, *Reconciling landmarks and level sets*, International Conference on Pattern Recognition, 2006.

[78] F. Memoli and G. Sapiro, *Distance functions and geodesics on submanifolds of $\mathbb{R}^d$ and point clouds*, SIAM J. Appl. Math. 65 (2005), no. 4, 1227–1260.

[79] _____, *A theoretical and computational framework for isometry invariant recognition of point cloud data*, Found. Comput. Math. 5 (2005), no. 3, 313–347.

[80] P.W. Michor and D. Mumford, *Riemannian geometries of space of plane curves*, Preprint, 2005.

[81] M. Miller and L. Younes, *Group actions, homeomorphisms, and matching : A general framework*, International Journal of Computer Vision **41** (2001), no. 1/2, 61–84.

[82] D. Mumford and J. Shah, *Boundary detection by minimizing functionals*, Proceedings of the International Conference on Computer Vision and Pattern Recognition (San Francisco, CA), IEEE, June 1985, pp. 22–26.

[83] D. Mumford and J. Shah, *Optimal approximations by piecewise smooth functions and associated variational problems*, Communications on Pure and Applied Mathematics **42** (1989), 577–684.

[84] S. Osher and R. Fedkiw, *The level set method and dynamic implicit surfaces*, Springer-Verlag, 2002.

[85] S. Osher and N. Paragios (eds.), *Geometric level set methods in imaging, vision and graphics*, Springer Verlag, 2003.

[86] S. Osher and J.A. Sethian, *Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton–Jacobi formulations*, Journal of Computational Physics **79** (1988), no. 1, 12–49.

[87] N.C. Overgaard and J.E. Solem, *An analysis of variational alignment of curves in images*, International Conference on Scale Space and PDE Methods in Computer Vision, 2005, pp. 480–491.

[88] N. Paragios and R. Deriche, *Unifying boundary and region-based information for geodesic active tracking*, Proceedings of the International Conference on Computer Vision and Pattern Recognition (Fort Collins, Colorado), IEEE Computer Society, June 1999.

[89] _____, *Geodesic active regions: a new paradigm to deal with frame partition problems in computer vision*, Journal of Visual Communication and Image Representation, Special Issue on Partial Differential Equations in Image Processing, Computer Vision and Computer Graphics (2001), Appeared in 2002 - Vol 13 No 1/2 - Pages 249-268 - USE INSTEAD paragios-deriche:02b.

[90] _____, *Geodesic active regions and level set methods for supervised texture segmentation*, The International Journal of Computer Vision **46** (2002), no. 3, 223–247.

[91] _____, *Geodesic active regions and level set methods for motion estimation and tracking*, Computer Vision and Image Understanding **97** (2005), no. 3, 259–282.

[92] Nikos Paragios, *Geodesic active regions and level set methods: Contributions and applications on artificial vision*, Ph.D. thesis, University of Nice Sophia-Antipolis, January 2000.

[93] Nikos Paragios and Rachid Deriche, *Coupled geodesic active regions for image segmentation : a level set approach*, ECCV 2000, vol. II, June 2000, pp. 224–240.

[94] Nikos Paragios, Mikaël Rousson, and Visvanathan Ramesh, *Non-rigid registration using distance functions*, cviu **89** (2003), no. 2-3, 142–165.

[95] D. Peng, B. Merriman, S. Osher, H.-K. Zhao, and M. Kang, *A PDE-based fast local level set method*, Journal of Computational Physics **155** (1999), no. 2, 410–438.

[96] Xavier Pennec, *L'incertitude dans les problèmes de reconnaissance et de recalage – applications en imagerie médicale et biologie moléculaire*, Ph.D. thesis, Ecole Polytechnique, Palaiseau (France), December 1996.

[97] J.-P. Pons, G. Hermosillo, R. Keriven, and O. Faugeras, *How to deal with point correspondences and tangential velocities in the level set framework*, International Conference on Computer Vision, vol. 2, 2003, pp. 894–899.

[98] Mikael Rousson and Nikos Paragios, *Shape priors for level set representations*, ECCV 2002, vol. 2, May 2002, pp. 78–92.

[99] Sam. T. Roweis and Lawrence K. Saul, *Nonlinear Dimensionality Reduction by locally Linear Embedding*, Science **290** (2000), 2323–2326.

[100] W. Rudin, *Real and Complex Analysis*, McGraw-Hill, 1966.

[101] J. Serra, *Hausdorff distances and interpolations*, Mathematical Morphology and its Applications to Image and Signal Processing (1998), 107–114.

[102] Jean Serra, *Image Analysis and Mathematical Morphology*, Academic Press, London, 1982.

[103] J.A. Sethian, *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Sciences*, Cambridge Monograph on Applied and Computational Mathematics, Cambridge University Press, 1999.

[104] C.G. Small, *The statistical theory of shapes*, Springer-Verlag, 1996.

[105] S. Soatto and A.J. Yezzi, *DEFORMOTION, deforming motion, shape average and the joint registration and segmentation of images*, ECCV 2002, pp. 32–47.

[106] J.E. Solem and N.C. Overgaard, *A geometric formulation of gradient descent for variational problems with moving surfaces*, International Conference on Scale Space and PDE Methods in Computer Vision, 2005, pp. 419–430.

[107] G. Sundaramoorthi, A.J. Yezzi, and A. Mennucci, *Sobolev active contours*, IEEE Workshop on Variational and Level Set Methods (Beijing, China), 2005, pp. 109–120.

[108] M. Sussman, P. Smereka, and S. Osher, *A level set approach for computing solutions to incompressible two-phase flow*, Journal of Computational Physics **114** (1994), no. 1, 146–159.

[109] Joshua Tenenbaum, Vin De Silva, and C. Langford, John, *A Global Geometric Framework for Nonlinear Dimensionality Reduction*, Science **290** (2000), 2319–2323.

[110] A. Toga and P. Thompson, *The role of image registration in brain mapping*, Image and Vision Computing **19** (2001), no. 1-2, 3–24.

[111] Arthur Toga (ed.), *Brain warping*, Academic Press, 1998.

[112] A. Trouvé, *Diffeomorphisms groups and pattern matching in image analysis*, The International Journal of Computer Vision **28** (1998), no. 3, 213–21.

[113] A. Trouvé and L. Younes, *Diffeomorphic matching problems in one dimension: designing and minimizing matching functionals*, ECCV 2000, pp. 573–587.

[114] ———, *Mise en correspondance par difféomorphismes en une dimension: définition et maximisation de fonctionnelles*, 12ème Congrès RFIA'00 (Paris), February 2000.

[115] A. Trouvé and L. Younes, *Metamorphoses through Lie group action*, Foundation of Computational Mathematics (2005), To appear.

[116] A.J. Yezzi and A.C.G. Mennucci, *Metrics in the space of curves*, Preprint, 2005.

[117] A.J. Yezzi and S. Soatto, *Deformotion: Deforming motion, shape average and the joint registration and approximation of structures in images*, The International Journal of Computer Vision **53** (2003), no. 2, 153–167.

[118] L. Younes, *Computable elastic distances between shapes*, SIAM Journal of Applied Mathematics **58** (1998), no. 2, 565–586.

[119] ———, *Optimal matching between shapes via elastic deformations*, Image and Vision Computing **17** (1999), no. 5/6, 381–389.

[120] ———, *Invariance, déformations et reconnaissance de formes*, Mathématiques et Applications, Springer-Verlag, 2003.

[121] H.-K. Zhao, S. Osher, B. Merriman, and M. Kang, *Implicit and non-parametric shape reconstruction from unorganized points using a variational level set method*, Computer Vision and Image Understanding **80** (2000), no. 3, 295–314.