Image segmentation using MRFs and statistical shape modeling

Ahmed Besbes

ECOLE CENTRALE PARIS

P H D T H E S I S

presented by

Ahmed BESBES

to obtain the title of

Doctor of Ecole Centrale Paris

Specialty : Applied Mathematics

Image Segmentation using MRFs and Statistical Shape Modeling

prepared at Ecole Centrale Paris, MAS Laboratory

defended on September 13, 2010

Committee:

Chairman:	Christian Saguez	-	Académie des Technologies
Reviewers:	Horst Bischof	-	Graz University
	Patrick Pérez	-	Thomson Corporate Research
Examiners:	Leo Grady	-	Siemens Corporate Research
	Grégoire Malandain	-	INRIA Sophia-Antipolis
	Mads Nielsen	-	University of Copenhagen
	Ramin Zabih	-	Cornell University
Advisor:	Nikos Paragios	-	ECP - INRIA Saclay

2010ECAP0024

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

INRIA

centre de recherche SACLAY - ÎLE-DE-FRANCE

Image Segmentation using MRFs and Statistical Shape Modeling

Abstract: In this thesis, we introduce a new statistical shape model and use it for knowledge-based image segmentation. The model is represented by a Markov Random Field (MRF). The vertices of the graph correspond to landmarks lying on the shape boundary, whereas the edges of the graph encode the dependencies between the landmarks. The MRF structure is determined from a training set of shapes using manifold learning and unsupervised clustering techniques. The inter-point constraints are enforced using the learned probability distribution function of the normalized chord lengths.

This model is used as a basis for knowledge-based segmentation. We adopt two approaches to incorporate the data support: one is based on landmark correspondences and the other one uses image region information. In the first case, correspondences between the model and the image are obtained through detectors and the optimal configuration is achieved through combination of detector responses and prior knowledge. The second approach consists of minimizing an energy that discriminates the object from the background while accounting for the shape prior. A Voronoi decomposition is used to express this objective function in a distributed manner using the landmarks of the model. Both algorithms are optimized using state-of-the art efficient optimization methods.

We validate our approach on various 2D and 3D datasets of images, for computer vision applications as well as medical image analysis.

Keywords: Shape Modeling, Segmentation, Markov Random Fields, Discrete Optimization

Segmentation d'Images avec des Champs de Markov et Modélisation Statistique de Formes

Résumé : Nous présentons dans cette thèse un nouveau modèle statistique de forme et l'utilisons pour la segmentation d'images avec a priori. Ce modèle est représenté par un champ de Markov. Les nœuds du graphe correspondent aux points de contrôle situés sur le contour de la forme géométrique, et les arêtes du graphe représentent les dépendances entre les points de contrôle. La structure du champ de Markov est déterminée à partir d'un ensemble de formes, en utilisant des techniques d'apprentissage de variétés et de groupement non-supervisé. Les contraintes entre les points sont assurées par l'estimation des fonctions de densité de probabilité des longueurs de cordes normalisées.

Dans une deuxième étape, nous construisons un algorithme de segmentation qui intègre le modèle statistique de forme, et qui le relie à l'image grâce à un terme région, à travers l'utilisation de diagrammes de Voronoi. Dans cette approche, un contour de forme déformable évolue vers l'objet à segmenter. Nous formulons aussi un algorithme de segmentation basé sur des détecteurs de points d'intérêt, où le terme de régularisation est lié à l'a priori de forme. Dans ce cas, on cherche à faire correspondre le modèle aux meilleurs points candidats extraits de l'image par le détecteur. L'optimisation pour les deux algorithmes est faite en utilisant des méthodes récentes et efficaces.

Nous validons notre approche à travers plusieurs jeux de données en 2D et en 3D, pour des applications de vision par ordinateur ainsi que l'analyse d'images médicales.

Mots clés : Modélisation de formes, Segmentation, Champs de Markov, Optimisation discrète

Acknowledgements

I would like to express my sincere gratitude to my advisor Prof. Nikos Paragios for his guidance and support during the four years of my PhD. He gave me the opportunity to explore a challenging research topic, and constantly offered to me his help and encouragements. He always had an open door for discussions whenever I needed support or advice. This PhD was a great experience, and it would not have been possible without his insightful input and continual trust.

I was honored to have Prof. Horst Bischof, Dr. Leo Grady, Dr. Grégoire Malandain, Prof. Mads Nielsen, Dr. Patrick Pérez, Prof. Christian Saguez and Prof. Ramin Zabih as members of my thesis committee. I would like to thank them for traveling from far away to attend my defense, despite their busy schedule, and for providing me with timely feedback. Prof. Horst Bischof and Dr. Patrick Pérez reviewed this manuscript, and their pertinent comments contributed to improve its quality. I owe my deep gratitude to all of them.

I also would like to thank Prof. James Duncan for having welcomed me at the Biomedical Engineering laboratory at Yale University, during two months in the summer of 2007. Under the supervision of Prof. Nikos Paragios, the work that was conducted during this stay was very important for the remainder of my PhD studies.

This thesis would not also have been possible without the collaboration and help of my colleagues. I particularly acknowledge Dr. Nikos Komodakis and Dr. Georg Langs for their discussions and for kindly sharing their code. My thanks also go to Yangming Ou and Yun Zeng for their contribution to this work.

I am grateful as well to the members of the Applied Mathematics laboratory of Ecole Centrale Paris, and especially to the members of the computer vision and medical imaging team. I am indebted to my office

mates Charlotte, Martin, Régis, Noura, Olivier J., Daniel, Fabrice, Pacale and Bo for creating a friendly atmosphere and an inspiring environment. I was also very lucky to be surrounded by other outstanding people including Salma, Aris, Olivier T., Loïc, Panos, Katerina, Maxime, Mickaël, Chaohui, Iasonas, Dimitris, etc. We shared very nice moments and I enjoyed traveling with them to attend conferences, after some long nights of work spent at the lab to meet the submission deadlines. We also had the opportunity to go sightseeing and to visit wonderful places together, and the 2010 Easter in Greece will remain among my best memories. Thank you for your team spirit and for helping me out in the difficult times. It was a great pleasure to work with you. Special thanks go to Fabrice for reading my thesis manuscript and for his very helpful comments and suggestions. I would also like to thank Sylvie and Annie for their help and kindness, for so nicely taking care of us in the lab, and for making things so easier.

Thanks to my wonderful friends Zied, Mathieu, Benoit, Mourad, Aymen, Imen, Atef, Tarek, Asma, Leila, Hamza, Majed, Dorra, Sami and Heyfa, my life in Paris was filled with joy and happiness. I am grateful to them for their support, understanding and company throughout the years. My thanks also go to my dear friends back home in Tunisia, Jihen, Sofiane and Jihen, who always wished me all the best for my studies. To Soufiene my longtime great friend, thank you for your kindness and your faithfulness.

I would like to express my extreme gratitude to my dear friend and colleague Radhouène. We made this long journey together, from the very beginning to the very end, and he was always there for me. His rigor as a scientist and his uprightness as a man always inspired me respect and admiration. On top of the numerous discussions where his bright intuition always proved to be true, I gratefully acknowledge him for being my unofficial thesis reader. His precise and constructive input and constant moral support during the time pressure of thesis dissertation was invaluable. Coming back to Paris from abroad to attend my defense truly warmed my heart. Thank you Radhouène for being this pillar that never shivered throughout the years.

This thesis would have remained a dream had it not been for my beloved girlfriend Memia. She shared with me this wonderful experience and backed me at every moment, with patience and passion. Thank you Memia for everything.

Last but not least, I would like to pay a tribute to my family: my bothers Slim and Raouf, my father Abdelwaheb and my mother Salsabil. Thanks to their unconditional love, their eternal care and and their endless support, I coped with the most difficult and tough situations. Without the high ideals of Knowledge and Science that my parents instilled in me since my little childhood, without the enormous sacrifices they made for my well-being and my education, this work would not have been possible.

To you, Dad and Mom, is dedicated this thesis.

إلى أبي و أمّي
عبد الوهاب و سلسبيل

Contents

List of Figures

Chapter 1

Context and Motivation

Image segmentation is a fundamental task in computer vision and medical image analysis. It can be defined as the extraction of the boundaries of an object of interest from the image. Alternatively, it can be viewed as a labeling problem, where image pixels are assigned either to the object or to the background. Segmentation is of a great importance because it provides qualitative and quantitative information about the image. For instance, semi-automatic photo-editing is a popular application that benefits from segmentation techniques. Behavior analysis is another domain where segmentation represents a data pre-processing step. Note that the accuracy of the extracted boundaries would highly influence the quality of the resulting model. Image/video-based automatic translation of hand signs is a concrete example of gesture recognition that could benefit from efficient segmentation as well. Such an application can hardly be achieved without precisely outlining the hand and detecting the relative finger poses. In medical image analysis, clinical routines have integrated segmentation algorithms, to assist diagnosis. These tools allow physicians to gain precious time, alleviate the need for tedious manual processing, while providing satisfactory accuracy. For instance, assessing the left ventricular function of the heart is done through the estimation of the left ventricle ejection fraction. The latter is a quantitative indicator that can be computed from segmented magnetic resonance cardiac images.

A common methodology to address computer vision tasks defines a mathematical model that represents the problem, and then extracts the model parameters from image observations using optimization methods. Low-level segmentation approaches exclusively rely on neighboring image features and grouping methods. These techniques have limited success when compared to high-level methods that integrate shape priors in their modeling. In fields where a prior knowledge is available (like medical imaging), such methods carry on great potentials since the domain knowledge can be used to intro-

duce constraints and improve the reliability and accuracy of the segmentation result. In this challenging task, one first has to determine a model representing these constraints and then an inference process which aims to combine the visual support with the prior knowledge.

Different classes of shape models have been used for segmentation purposes. These models are linear or non-linear, global or local, represent the shape in an explicit or implicit manner, and may benefit from statistical learning. Although global and linear statistical shape modeling approaches are very popular in the computer vision community, and have been applied to a wide variety of segmentation problems, we believe that non-linear local shape priors can be very powerful. They can be further enhanced if they are combined in a structured manner, which can be derived from shape populations using manifold learning techniques. Such local object parts interactions are also very suitable for the use of Markov Random Fields (MRF) optimization techniques.

In this work, we made the choice to use discrete optimization methods. Despite more modeling constraints with respect to the continuous optimization setting, discrete methods have very appealing properties in general. Variational methods are known to converge to local minima in the general case, without giving any quality evaluation of the obtained solution. They require the computation of the gradient of the energy functional, which might not be differentiable in some cases, and also depend on a careful choice of an approximate numerical scheme, which very often leads to slow algorithms. On the contrary, the recent developments in the discrete optimization of MRFs provide tools that guarantee convergence in certain cases, provide suboptimality bounds beforehand, and per-instance tighter bounds in practice. Although they converge in general to a local solution as well, these methods could yield a better result as they permit to "jump" over local minima. They do not require the differentiation of the objective function, exhibit numerical robustness and are computationally efficient. If such a method gives an unsatisfactory result, then it is probably due to an erroneous modeling rather than to numerical instabilities.

We introduce in this thesis a new statistical shape model, in order to use it for segmentation tasks. This goal adds to the natural requirement of proper

shape variation representation the necessity of a tractable inference procedure. Our statistical shape model (SSM) relies on three building blocks that encapsulate geometric, manifold learning and graph theoretic aspects. Essentially, we adopt a landmark-based representation of shapes, that is embedded in a MRF, with an appropriate structure, namely a k-fan structure. Probability density distributions of the normalized inter-point distances are associated to the MRF to form the SSM, leading to a representation that is invariant with respect to translation, rotation and scale changes. In our development, we ensure the coherence of the chosen MRF structure with respect to the geometric aspect, namely the normalized distance representation. We also propose a framework that learns a particular instance of this structure for a particular class of objects, given a representative population of shapes. Hence, the final structure is directly related to the information encoded by the data. Furthermore, we show that the k-fan has desirable graph properties that permit exact inference only in some limited cases, due to the curse of dimensionality. In practice, we resort to state-of-the art efficient discrete MRF optimization techniques.

The second step of our work attempts to prove the relevance of our model for object segmentation. Our SSM enables the use of general image cues, and is suitable for segmentation algorithms in 2D and in 3D. We propose a method that relies on image region intensities to extract the object boundaries, and we show its robustness with respect to noise and occlusions. We also suggest another framework that benefits from multi-scale image features and trained classifiers in combination with our SSM. Experiments are performed in 2D and 3D with medical and natural images.

The remainder of this thesis is organized as follows. In chapter 2, we first review the shape modeling techniques related to segmentation applications. Then, we develop our statistical shape model by emphasizing our threefold contribution. We discuss the Euclidean distance geometry issues related to our k-fan structure, by making the link with a more general class of graphs called k-trees. We explain then how we use the shape maps technique to embed the shape population in a Euclidean space, where we apply an unsupervised linear programming-based clustering method to derive the particular graph structure that represents the data. The obtained graph being chordal, exact

inference is possible using dynamic programming, if the search space is small enough. After demonstrating the adequacy of the shape modeling technique on some practical examples, we proceed to the development of the segmentation algorithm in chapter 3. We first give an overview of the discrete MRF optimization methods. Then, we describe the segmentation techniques in computer vision, and especially the graph-based methods, discussing the relation of our method to the previous work. We particularly notice that incorporating statistical shape priors in MRFs is a difficult task. Then, we develop our aforementioned segmentation algorithms. Furthermore, extensions of our segmentation approach to other application settings are presented in appendix A. More precisely, we propose a method to incorporate dynamic shape priors in the MRF formulation in order to perform object tracking. In a different application, we use graph matching techniques to extract discriminatively-matched pairs of landmarks towards image registration. We conclude this work by discussing the limitations of our approach, and we propose several ways of improvements and directions for future research.

Chapter 2

Shape Model

Contents

2.1 Introduction

Bottom-up computer vision methods rely on comparing and grouping pixel features in local image neighborhoods with little prior knowledge. Top-down approaches is an alternative that gained a significant popularity in computer vision. These techniques allow the combination low-level cues from the image in a principled manner, based on a priori constraints. The learning of such prior information is central to a number of computer vision applications, such as recognition, segmentation and tracking of objects. In particular, shape models proved to be very useful in the context where one can determine a priori knowledge of the object of interest. Such models incorporate information about the geometry of the object, and can also describe its visual image features. This method applies to a wide range of applications in general computer vision (for instance faces, eyes, hands, human bodies...) and medical image processing as well (modeling of the heart, liver, corpus callosum...).

Building a shape model involves in general two components:

- A mathematical shape representation is defined.
- Constraints are added to this representation such that an admissible set of shapes is constructed. These constrains can be for instance learned as a statistical model that represents the variations of the shape.

The literature of shape modeling is very rich, and different representations were used in the context of computer vision. In particular, continuous efforts have been made to incorporate shape priors in segmentation algorithms, since the early 1990's. In [Heimann & Meinzer 2009], a review of statistical shape models (SSM) for three dimensional medical image segmentation is provided. Another review of statistical shape priors for level set segmentation is available in [Cremers *et al.* 2007]. We now review the major shape modeling techniques previously proposed in the literature.

2.1.1 A Review of Shape Models

Point-Distribution Models - Linear Models

One of the most widely used algorithms in the area of shape modeling is the active shape model (ASM), also called point distribution model (PDM).

Point-based or landmark-based representations have been studied, developed and successfully applied in vision applications as well as medical image analysis following the pioneering work of [Cootes *et al.* 1995]. In this approach, the shape is represented as a vector of concatenated point coordinates in a space of dimension $n \times d$, where n is the number of points, and d is the dimension of the points space. These points are in general boundary points, and the shape is easily recovered by connecting them, to form a meshed surface or a 2D curve. Then, a (linear) statistical model is learned from a training set of t shape instances, where correspondences between the landmarks are known. Very often, the correspondences are obtained through the manual labeling of the training shapes. However, such data are not always available (especially for 3D medical applications). Moreover, the manual precision of the correspondences is questionable in general, and can vary from one person to another. For these reasons, different methods for automatic correspondence finding were developed. For instance, [Davies *et al.* 2002] propose a population-based formulation to the problem by minimizing an objective function that features the minimum description length (MDL) of the statistical model.

The learning phase of ASM is based on principal component analysis (PCA). First, an alignment of the training shapes is performed, to filter out translation, rotation and scale change effects. This can be done using various rigid alignment methods like Procrustes analysis. Next, a mean shape $\bar{\mathbf{s}}$ and a covariance matrix $\mathbf{C}_s$ are computed. An eigendecomposition of $\mathbf{C}_s$ gives the modes of variations of the shape with respect to the mean shape. These modes are represented by the eigenvectors $\mathbf{e}_i$ of the covariance matrix. By projecting the residual of an instance shape $\mathbf{s}$ from the mean on the obtained basis of eigenvectors, it can be expressed as a linear combination of the eigenmodes. More explicitly, we have $\mathbf{s} = \bar{\mathbf{s}} + \sum_{i=1}^{n \times d} b_i \mathbf{e}_i$. In the last step, the dimensionality of the problem is reduced and only the eigenvectors corresponding to the largest m eigenvalues are kept, where in general m is chosen such that a large part of the trace of $\mathbf{C}_s$ is retained. This yields the compressed expression of a shape $\mathbf{s} = \bar{\mathbf{s}} + \sum_{i=1}^{m} b_i \mathbf{e}_i$ with a small residual error. The admissible set of shapes is generally defined by constraining the coefficients b_i in the range $[-3\sqrt{\lambda_i}, 3\sqrt{\lambda_i}]$, where λ_i is the eigenvalue corresponding to the eigenvector $\mathbf{e}_i$. Point distribution models are hence linear models that make the Gaussianity

assumption of the data, and rotate the coordinate axes to the directions of maximum variance. They are global shape models of local variations (with respect to the mean).

Because of their global effects, eigenmode variations of the PDM result in a change of all the landmarks. Each single mode controls the variations of the whole shape. In some applications, shape model sparsity is a desired property. For instance, PDM can handle the variations of a healthy organ, since the global shape assumption is reasonable in this case. However, for pathologies that affect the organ (and hence its shape) locally, the success of a PDM might be limited. This observation motivated the work of [Stegmann *et al.* 2006] who looked for variation modes that affect the shape locally. This is achieved through the Orthomax method that rotates the eigenmodes such that a sparsity criterion is optimized. This leads to a similar linear orthogonal decomposition of the shape as in PDM, yet with local influence properties. Independent component analysis (ICA) [Hyvärinen *et al.* 2001] is a different approach that tackles sparse modeling. It supposes that the measured signal is a linear mix of non-Gaussian source signals that are statistically mutually independent (without orthogonality constraints). It was for example used by [Üzümcü *et al.* 2003] to segment the left ventricle (LV) from magnetic resonance (MR) images in an active appearance model (AAM) [Cootes *et al.* 2001] framework. A variation on the PCA eigenmodes rotation is also introduced by [Alcantara *et al.* 2009]. Following the ideas presented in [Chennubhotla & Jepson 2001], they define an energy function that is a trade-off between a variance term and a locality term. The optimization is performed by a PCA-based initialization, followed by the rotation of pairs of eigenmodes in their plane, such that orthonormality is preserved. These methods however suffer from the lack of natural order in the variation modes, in contrast with PCA, where the eigenmodes order is specified by variance. Hence, the authors used different methods and criteria in order to cope with this limitation.

Another problem arises in practice when dealing with PDM. The number of available shapes in the training set is often insufficient, especially in the 3D case, where a manual labeling of medical data can be very tedious while establishing correspondences can also be very challenging. The

maximum number of degrees of freedom that can be described by the training set is $t-1$, and so the number of possible eigenmodes. In this case, the learned PDM model would result in incorrect fitting during the search step. [Cootes & Taylor 1996] attempt to solve this problem by directly acting on the covariance matrix, and adding (synthetic) variance and covariance values for neighboring curve points, to correlate their displacements. [Wang & Staib 1998] use the same idea to build a flexible deformable model. They keep the mean shape, learned from the training data, and discard the covariance matrix, replacing it by the so-called smoothness covariance matrix. The latter contains only variance values, and covariance term for neighboring landmarks on the curve. This model does not depend on statistical shape variations, and allows the deformation of the shape under smooth boundary constraints.

Linear shape models have also been used in the variational framework. [Cremers *et al.* 2001] introduced shape priors into the Mumford Shah functional. They used a spline parametrization of the contours and learned a multivariate Gaussian distribution over the control point positions. They introduced in [Cremers *et al.* 2002b] invariance to translation rotation and scale changes in the variational framework. To overcome the non-invertibility of the covariance matrix in the case of a small training set, the latter was regularized by replacing the zero eigenvalues by the smallest eigenvalue. Hence, a shape variation that is not represented by the training set has a smaller probability than any shape variation encoded by the training set.

A common criticism that is formulated to PDM regards their limited ability to model shapes with relative rotations, or bending variations, which makes the assumption of linearity (or locality on the Riemannian manifold of shapes) not valid any more. This limitation was indeed described by [Cootes *et al.* 1995], and motivated the development of non-linear models.

Non-Linear Models

Different approaches were considered to model non-linearity of shape variations. Among others, [Heap & Hogg 1997] build a hierarchical shape model by clustering the data into (overlapping) clusters, the number of clusters being specified by the user. They define afterwards a PDM for each cluster,

with bounds on the shape parameters, such that the defined shape region is valid. [Cootes & Taylor 1999] also use a supervised method to deal with non-linearities. After projecting the training data on the PCA eigenmodes, they estimate a probability density function (PDF) of the shape as a mixture model of Gaussians, where the number of Gaussians is specified beforehand. The estimation is done using the expectation maximization (EM) algorithm.

In an alternative approach, a non-linear mapping of the data to a feature space is introduced, and PCA is performed in the feature space, supposed to be linear. This is the principle of kernel principal component analysis (KPCA) [Schölkopf *et al.* 1998] that was applied by [Romdhani *et al.* 1999] to the shape modeling problem in the case of multiple 2D views of a 3D object, and by [Twining & Taylor 2001] for the modeling of large-amplitude bending objects (nematode worms). The central idea of this approach lies in Mercer's theorem, where the non-linear mapping does not have to be explicitly defined, but can be specified in terms of inner products, expressed using a Mercer kernel. The feature space is called a Reproducing Kernel Hilbert Space (RKHS). A related method is [Cremers *et al.* 2002a] where the authors extend their previous Gaussian prior Mumford-Shah formulation [Cremers *et al.* 2001] to the non-linear case. They estimate the distribution of the mapped training data by a Gaussian PDF in the RKHS, and discuss the close relation of their method to the Parzen windows estimator and KPCA. Gaussian kernels were often chosen in the above methods.

[Charpiat *et al.* 2007] develop a method to compute statistics of a training set of contours and to use them as a non-linear shape prior in a variational framework. Considering a differentiable approximation of the Hausdorff distance between shapes, they define the mean shape as the one that minimizes the sum of distances to all the training samples up to rigid motions, and propose a gradient descent framework to compute it. Then, statistics on the deformation fields that drive this mean shape to each sample are learned using PCA. They are plugged into a variational framework for image segmentation, and define priors that are invariant to rigid motions.

Implicit Representations

Implicit representation is another typical approach to model shapes, where the contour of the shape is embedded in a high dimensional space and considered as the zero level of a representing surface. The signed distance transform or signed distance map (SDM) of the contour is commonly considered as the embedding function in this case [Osher & Sethian 1988]. [Leventon *et al.* 2000] augment this representation with a statistical model to capture shape variations. After a rigid alignment of the training contours, they also use PCA, and apply it on the distance transform image collection obtained from the training contours. After dimensionality reduction, eigen distance maps represent the variations of the shapes. This same representation is also used by [Tsai *et al.* 2001], in an efficient level-set segmentation algorithm. [Chan & Zhu 2005] define a distance between two SDM using the Heaviside distribution.

Similarly to the PDM case, criticism was addressed to the above approaches regarding the linearity assumption. Hence, the underlying approximation of the manifold of signed distance maps as a linear vector space is questionable. Linear combinations of signed distance maps do not necessarily give a signed distance map, which can compromise the use of PCA in this case. Moreover, as pointed out in [Cremers *et al.* 2007], the benefits of PCA dimensionality reduction can have a downside. The principal eigen distance maps encode most of the variations of the training set on the space of embedding functions, which does not necessarily correspond to most of the variations of the embedded contours in the learning population. This may result in selecting a large number of eigenmodes to build a satisfying shape model. These considerations motivated the use of the following alternative statistical models.

Implicit representations were also used in [Rousson & Paragios 2002] and [Rousson & Paragios 2008]. The prior information was however learned pixel-wise (grid-wise) as a Gaussian probability density function. [Taron *et al.* 2005] adopt a similar approach and represent shapes as signed distance functions. However, they deal with the correspondence problem by registering the training shape implicit representations using a topology-preserving free-from deformation algorithm. Then, the registration uncertainty is estimated by comput-

ing a covariance matrix at the surface zero-level. The following estimation of the probability density function (PDF) of the shape, using variable bandwidth kernels, takes into account the registration uncertainties. [Pohl *et al.* 2006] and [Pohl *et al.* 2007] dealt differently with the problem as they tackled the representation part. They introduced the logarithm of the odds ratio (LogOdds) representation to the area of computer vision. The LogOdds function maps a probability $p \in \mathbb{P}$ ($\mathbb{P}$ being the probability space) to a real value through the logarithm of the ratio between p and its complement $1-p$ (the inverse map is the logistic function). The LogOdds maps form a vector space ($\mathbb{L}^n = \mathbb{R}^n$), and the authors deduce a vector space structure on $\mathbb{P}^n$. This allows probabilistic interpretation for addition and scalar multiplication in the LogOdds space $\mathbb{L}^n$. A shape silhouette (binary map) can be hence transformed to a LogOdds map, where the zero-crossing corresponds to the object boundary. This framework provides an embedding of the signed distance maps manifold into the LogOdds vector space. Hence, SDMs can be considered as LogOdds maps (while the reverse is in general not true) and interpreted in terms of space conditioned probabilities. Determining the statistics of SDMs, as LogOdds maps, using PCA is hence possible, and the authors incorporate it in a Bayesian classification algorithm.

[Etyngier *et al.* 2007] addressed the problem differently by using manifold learning techniques. They first compute the distance between two training shapes as the Sobolev $W^{1,2}$-norm of the difference between their SDM. These pairwise distances are plugged in the diffusion maps algorithm [Coifman & Lafon 2006] to build an approximation to the shape manifold through an isometric mapping of the data. A Delaunay triangulation is then computed in the reduced manifold space, and used to find the nearest neighbors of a new shape instance, in order to estimate its projection on the manifold. This projection is introduced in a shape prior term in the variational framework, to attract the segmentation result to the defined manifold.

Parametric Geometric Representations

Parametric geometric representations like superquadrics were used as deformable models for object segmentation and reconstruction. [Metaxas & Terzopoulos 1993] combined superquadrics with splines to

couple global and local shape parameters, and applied them to the tracking of human limbs. The global parameters enable to fit approximately the target shape in the image, and the local degrees of freedom allow the segmentation to be refined. [Vemuri *et al.* 1993] represent the deformable superquadric model in an orthonormal wavelet basis to inherit multi-resolution capability, and apply it to the reconstruction of human brain structure from MRI images. Then [Bardinet *et al.* 1995] and [Bardinet *et al.* 1998] also used a deformable superquadric model to segment the left ventricle (LV). The model was combined with free-from deformations (FFD), which allows one to refine the initial guess of the object and to improve the reconstruction.

Medial Models

The medial axis model is a shape representation introduced by [Blum 1973]. It is defined as the set of centers of all internal bi-tangent spheres to the object boundary that are not inscribed in any other such sphere. The medial transform of an object describes it by its medial axis as well as the corresponding radii of the bi-tangent spheres. This approach is used by [Pizer *et al.* 1999], who represent the interior of an object by sheets of medial atoms from which the boundary can be synthesized. This model is called m-rep, and besides this discrete 2D version, a three-dimensional extension is proposed in [Pizer *et al.* 2003]. A continuous formulation of m-reps is presented in [Yushkevich *et al.* 2006]. M-reps are very suitable for representing shape variations such as thickness change, bending, and widening. However, their non-Euclidean parameters prohibit the use of linear statistical approaches. [Fletcher *et al.* 2004] circumvent this problem by developing a generalization of PCA to the Riemannian manifold setting. In particular, they define the intrinsic mean, compute the principal components by maximizing a variance criterion, define the geodesic sub-manifolds (counterparts of linear subspaces) and devise a way to project on them. They apply the so-called principal geodesic analysis (PGA) to a population of hippocampi in a schizophrenia study.

Frequency-Domain Representations

An alternative approach that considers the Fourier transformation of a contour and uses a probability distribution on the Fourier coefficient was adopted in [Staib & Duncan 1992] (and extended by [Chakraborty *et al.* 1996]). A Gaussian probability distribution was learned for each parameter. This representation was also used by [Székely *et al.* 1996] who however modeled the variations of the shape using principal component analysis (PCA) in the Fourier coefficient space. This work is extended in [Kelemen *et al.* 1999], where the parametric representation of shapes uses series of spherical harmonics (SPHARM) and PCA is performed in the shape parameters space as well. The applications dealt with the segmentation in 2D and 3D of brain and cardiac structures. [Davatzikos *et al.* 2003] represent shape using a wavelet transform of the landmark coordinates. This frequency analysis permits to obtain coarse-level shape information, as well as fine-level details, thanks to the wavelets coefficients in each band. A statistical hierarchical model is then built, by applying PCA to the coefficients in each band separately. An ASM-based search strategy was applied to the segmentation of the corpus callosum in MR images, as well as the segmentation of hand images. [Nain *et al.* 2007] and [Yu *et al.* 2007] extended this approach to 3D via spherical wavelets. Their method is however limited to 3D shapes that have a spherical topology (genus-0 manifolds). [Essafi *et al.* 2009] overcome this limitation by utilizing diffusion wavelets [Coifman & Maggioni 2006] as a hierarchical shape representation. They apply the orthomax algorithm to the obtained coefficients which leads to a sparse shape parametrization.

Shape Descriptors-based Models

Other approaches associate to landmarks individual shape descriptors that can have geometric-invariant properties. For instance, [Coughlan *et al.* 2000] define a deformable template of the hand that is invariant to translation and rotation. The landmarks are associated with the normal vector to the contour at their location. When matching the template to an image, two geometric Gaussian constraints (with mean values from the template) are imposed to the model, and combined with the extracted image features (corners, histograms

of gradient orientations,etc...). The first one concerns the angle between the normals associated to two successive points on the contour. The second one is related to the angle between two successive points on the contour and the local normal vector. The problem is defined as a maximum a posteriori (MAP) optimization in a Bayesian framework. The structure of the prior is a Markov chain which allows the use of dynamic programming to obtain the global solution without a need for initialization. A similar approach is considered in [Coughlan & Ferreira 2002], where the structure of the prior is not a tree, but a complete graph. The minimization of the Markov Random Field (MRF) defined energy is done using loopy belief propagation.

Attribute vectors were also used to describe shape geometry by [Shen & Davatzikos 2000] (as a 2D version followed by a 3D version in [Shen *et al.* 2001]). For each landmark, they used normalized areas of local triangles (respectively normalized volumes of local tetrahedrons in 3D) formed by connecting each control point to its local neighboring points. The snake internal energy reflects the difference between the model attributes and those of the evolving contour.

[Belongie *et al.* 2002] design a different descriptor for each landmark, called "shape context", by attaching to each point a log-polar coordinates histogram of all the relative positions with respect to the other shape landmarks. A distance is defined based on this descriptor to estimate the similarity between shapes. It is possible to obtain a similarity-invariant version of the descriptor using a scale normalization, and rotation-invariance is ensured through a local Frenet frame instead of the absolute frame for coordinates calculations.

The work of [Manay *et al.* 2006] can be viewed as a continuous generalization of "shape context". The authors define integral invariants as shape descriptors and provide a distance for shape matching. [Hong *et al.* 2006] use this descriptor for object segmentation.

From Single Templates to Statistical Pictorial Structures

Incorporating priors in the form of a (deformable) shape template is an idea that retained the attention of many researchers. It goes back to the work of [Fischler & Elschlager 1973] who used a mass-spring template as a represen-

tation of faces (by modeling the inter-parts dependencies). [Yuille *et al.* 1989] matched "hand-crafted" models of face parts (eye, mouth) to the image to extract objects. Two-dimensional templates were used as a shape representation in [Tagare 1997]. These templates were closed contours, similar to the object to be segmented. They were associated with orthogonal pre-computed curves that constrained the deformations of the contour during the search. An average contour model have been used to represent shapes in [Chen *et al.* 2002], after rigid alignment of the training examples. In the case of a class of shapes that exhibits large variations, the training examples were classified in different clusters and an average shape model was considered for every cluster. More recently, [Schoenemann & Cremers 2007] developed a graph-theoretical algorithm for segmentation and alignment of shapes that produces a global minimum to the optimization problem. The algorithm uses an elastic shape prior that is invariant to scale and translation changes, and is able to match a single object template to the image based on edge information. The algorithm can benefit from parallelization. It is also possible to induce rotation-invariance by matching successively the image to the template rotated by different angles.

[Felzenszwalb & Huttenlocher 2000] develop an efficient optimization framework of the pictorial structures model (introduced in [Fischler & Elschlager 1973]). The best match of the model to an image is the one that gives good confidence for each part while satisfying inter-parts constraints. They require the object model to have a tree structure, which allows the use of dynamic programming to minimize the objective function in linear time (rather than quadratic time, thanks to a restriction on the pairwise cost function) and obtain a global optimum. A person model and a car model are designed and applied for detection. The paper introduces also a Bayesian formulation of the problem as a maximum a posteriori (MAP) estimation. This framework allows for learning the model parameters statistically in [Felzenszwalb & Huttenlocher 2005] using a maximum likelihood method. Hence, the algorithm chooses the parameters that best represent the appearance of each part of the model in the training set, as well as the optimal tree structure that best explains the data.

2.1.2 Our Proposed Method

In this work, we endeavor to propose a shape model that overcomes the shortcomings of the existing approaches. First, global models often suffer from their lack of flexibility. For instance, while aiming to prevent invalid shapes, PDM will fail to capture new shape variations if the training set is too small, and the object projection on the learned subspace can be too constraining. Adding local degrees of freedom is subsequently a desired property. Moreover, we would like to be able to model non-linear variations, such that the representation is usable for a large class of shapes. A generic formulation that widens the range of applications to an arbitrary object dimension (2D or 3D) is advantageous, while the ability to properly represent the data is of equal importance. Another relevant aspect to be addressed is the tractability of the inference related to the shape model. Eventually, it is widely accepted that shape is the remaining information of the data after filtering location, orientation and scale effect.

Subsequently, we propose a statistical shape model (SSM) that is similarity-invariant. We adopt a landmark-based representation, where the global shape information is expressed as a combination of local (sparse) interactions. Our representation is associated with a particular model structure that properly captures the geometry of the object and exhibits useful mathematical properties. The model parameters are deduced from the data, through manifold learning techniques. Our derivation permits to define a shape prior as an energy term that is decomposable into pairwise terms. When exact inference is not possible (curse of dimensionality), the latter property opens up new horizons for efficient optimization using state-of-the art MRF methods.

More precisely, the chord lengths of the object, defined by pairs of landmarks, and normalized by the object scale, are used as a description of the shape. These normalized inter-point distances bequeath similarity-invariance to our representation. A statistical model of the shape is then defined by estimating a probability density function for each normalized chord length (NCL) from a training set of images. Akin to the PDM setting, we require correspondences between the control points. However, there is no need to register the shape since the invariance property is intrinsic to our representation. Hence, the model encodes the -possibly nonlinear - variations of the shape, regardless

of its location, orientation and scale.

In the following step, we observe that considering the whole set of chords is redundant. A smaller set of such edges, with the distance information, is sufficient to capture the shape geometry. However, a particular structure of the retained edges is needed to ensure the geometric modeling coherence. It is hence interesting to see that the k-tree structure (and also the k-fan structure - definitions are given later, see Fig. 2.3 for examples) is equivalent to the fully connected structure in the case of perfect distance matching. Nonetheless, for a given set of control points, different instances of such a structure can be defined. On top of these numerous possibilities there is also the variability that a change in the parameter k generates. Since different k-trees (whether they have the same k or not) would lead to different models, with various degrees of data fidelity, we aimed to learn the k-fan structure as well as the number of fans k from the data, such that it is best represented by the model.

To this end, we use a manifold learning technique, called shape maps [Langs & Paragios 2008], that combines the strengths of diffusion maps [Coifman & Lafon 2006] and model description length. This technique embeds the training shape population into a metric space, where the Euclidean distances between the embedded landmarks account for their interdependency in the data. Then, we use an unsupervised clustering technique (based on linear programming (LP)) [Komodakis *et al.* 2008a] to group the similarly-behaving control points in an automatically-defined number k of clusters. Then, the k-tree construction stems from the obtained clusters.

From a graph-theoretic point of view, k-tree graphs belong to the larger class of chordal graphs, that inherit the property of the perfect elimination scheme. This paves the way to the use of dynamic programming approaches for exact inference, as well as for sampling from the posterior probability distribution. Moreover, since our formulation is based on normalized inter-points distances, it exhibits pairwise energy terms, that are adequate for the use of efficient MRF optimization techniques.

2.1.3 Related Work

Our approach relates to previous art through the use of inter-points-based representations, k-trees and graph structure learning.

Chord-based Representations

The use of chord-length representations of shapes has been to some extent explored in the literature. Distributions of random chord lengths were used for shape matching and recognition in the work of [Smith & Jain 1982] and [You & Jain 1984]. They compute for each shape all inter-boundary points distances and normalize them by the length of the longest chord. They estimate then a probability distribution using all the computed normalized chord lengths (NCL). Then they use a Kolmogorov-Smirnov statistical test to compare two shapes. Despite their interesting results, [Mallows & Clark 1970] had showed earlier that this shape descriptor is not information preserving, as the global distributions of chord lengths can be equivalent for two different (non-identical) shapes. These early methods confirm the richness of this mathematical representation. Nevertheless, note that our statistical model (individual statistics of NCL over a training population) is different from what they used as we will show later.

A representation based on the distances between pairs of landmarks was also considered in the context of PDM [Cootes *et al.* 1992] [Taylor & Cooper 1990][Cooper *et al.* 1991]. A shape is hence represented as a vector of all chord lengths, and PCA is applied to obtain a shape model as a mean shape and eigenmodes. The authors explain in [Cootes *et al.* 1995] the ability of this rigid-motion-invariant representation to capture nonlinear shape variations (as bending shapes), opposite to conventional PDM. Before further commenting this approach, we need to introduce the following concepts.

Definition 2.1. Euclidean Distance Matrices (EDM).
A matrix $D = [d_{ij}^2] \in \mathbb{R}^{N\times N}$ is an EDM if there exist N points $(x_1, \ldots, x_N) \in \mathbb{R}^n$, for some $n \geq 1$, such that $\forall\ 1 \leq i, j \leq N, d_{ij} = \|x_i - x_j\|_2$. The set of points $(x_1, \ldots, x_N)$ is called a realization of D in $\mathbb{R}^n$.

EDM are invariant with respect to translation, rotation and reflexion [Dattorro 2005]. Hence, two sets of points $(x_1, \ldots, x_N)$ and $(p_1, \ldots, p_N)$ have the same EDM if and only if they are related by an isometry. This observation justifies the use of EDM (as being a "strong constraint") to represent shapes. According to [Dattorro 2005], the set of EDM is not a linear space, but is a convex cone. Moreover, note the following classical result of Euclidean

geometry.

Theorem 2.1 ([Schoenberg 1935][Young & Householder 1938]). *A matrix* $D = [d_{ij}^2] \in \mathbb{R}^{N\times N}$ *admits a realization* $(x_1, \ldots, x_N)$ *in* $\mathbb{R}^n$ *but not in* $\mathbb{R}^{n-1}$ *if and only if the* $(N-1)\times(N-1)$ *symmetric matrix* $A = [a_{ij}]_{1\leq i,j\leq N-1}$ *defined by:*

$$a_{ij} = \frac{1}{2}(d_{1(i+1)}^2 + d_{1(j+1)}^2 - d_{(i+1)(j+1)}^2)$$

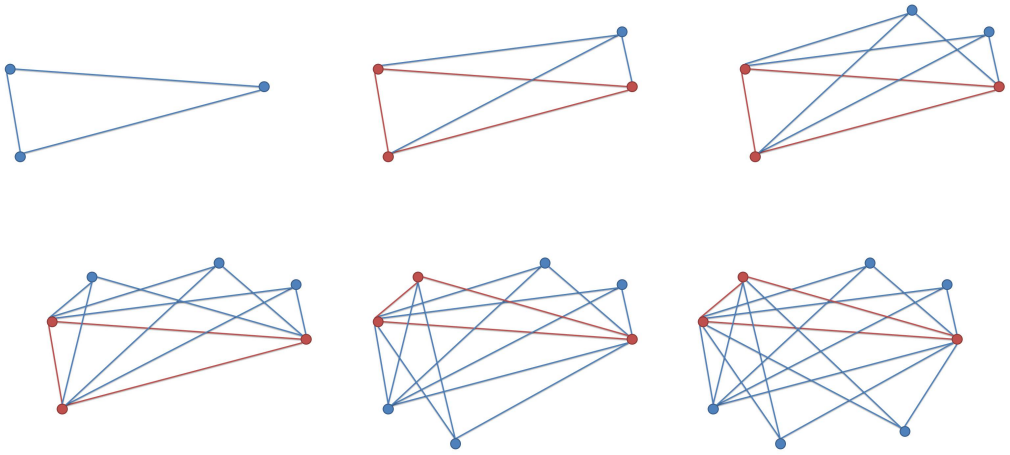
is positive semidefinite and has rank n.

The proof of this theorem can be found in [So 2007], and is reported in Appendix B. We can see that the set of EDMs that admit a realization in $\mathbb{R}^n$ is a sub-manifold of a convex cone. For this reason, applying PCA on the set of EDMs may be questionable. With this linearity assumption, the computed chords may be non-Euclidean in n-D, as pointed out in [Cootes *et al.* 1992]. Furthermore, the produced matrices may violate the EDM assumption. The authors try to cope with this problem by approximating the produced chord lengths using a steepest gradient descent optimization. This local search may generate plausible shapes for simple cases, but is doomed to fail for more complex objects.

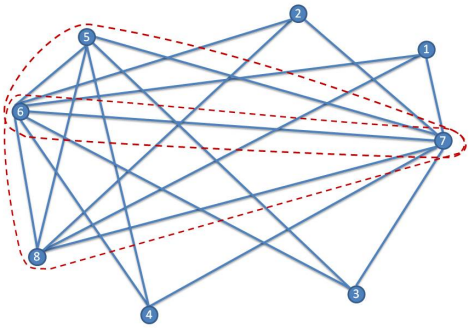
Our approach is also related to another line of work by [Caetano *et al.* 2006] that deals with graph matching. Based on the Euclidean geometry theory, they define a matching criterion between two graphs with respect to the induced distortion, using the inter-graph nodes distances. While their approach does not deal with statistical shape modeling, they show an interesting result that is inherent to the graph structure. Before proceeding, let us give a few definitions. We denote a complete graph with n vertices by K_n. A k-tree is defined recursively as follows:

Definition 2.2. k-trees, base k-cliques.
K_k is a k-tree. A k-tree $\mathcal{T}_{n+1}$ on $n+1$ vertices can be generated from a k-tree $\mathcal{T}_n$ on n vertices, by adding a new vertex to $\mathcal{T}_n$ and connecting it with k edges to some k-clique (a clique with k vertices) of $\mathcal{T}_n$. This latter clique is called then a base k-clique.

Fig. 2.1 illustrates the previous definition, and Fig. 2.3 shows, among other chordal graphs, examples of k-trees. As it will be detailed later,



(a)



(b)

Figure 2.1: The definition 2.2 gives a recursive procedure to build a k-tree. (a) The steps of building a 3-tree are shown. In each step, the current base 3-clique is depicted in red. (b) The obtained graph has 2 base 3-cliques. A simplicial elimination order is given.

[Caetano *et al.* 2006] prove the global rigidity of k-trees in $\mathbb{R}^{k-1}$, and subsequently show that the k-tree model is equivalent to the fully connected (complete graph) model for the case of exact graph matching in $\mathbb{R}^{k-1}$ [(1)]. Note that the use of a shape model based on this structure to segment an image can be thought of as matching a "statistical graph" to some graph ex-

[1] Chordal graphs will be defined subsequently in definition 2.3 (section 2.1.3), and the notion of rigidity will be clearly introduced in definition 2.5 (section 2.2.2)

tracted from the image. Hence, the latter property is of great interest to our approach. [Caetano *et al.* 2006] explain however that in case of position jitter (inexact matching case), the k-tree model is only an approximation to the complete model, while being extremely robust. They also discuss the problem of the k-tree selection as different k-trees can result in different matching accuracies in the inexact matching case. In our method, we do not restrict ourselves to k-trees in $\mathbb{R}^{k-1}$, but consider the n-trees structure ($n \geq k$). We learn n as well as the structure of the tree from the data in a (statistical) manifold learning framework. The work of [Caetano *et al.* 2006] discusses geometric aspects as well as graph-theoretic optimization aspects. The k-trees being chordal graphs, they permit to solve general NP-hard problems in polynomial time (precisely, they use the junction tree algorithm to solve a MAP problem).

Chordal Graphs

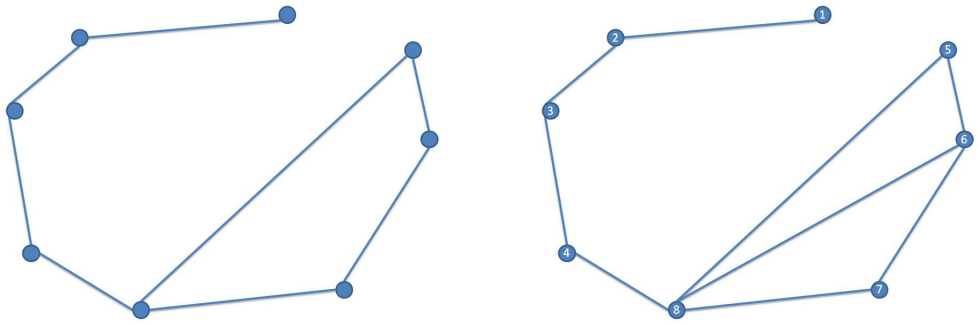
Let us first give a few definitions:

Definition 2.3. Chordal Graph
A graph is chordal if it does not contain any chordless cycle of length at least 4. A chord is an edge that connects two non successive vertices in a cycle.

Definition 2.4. Simplicial Elimination Ordering (SEO)
A vertex of a graph is called simplicial if its neighbors form a clique (a clique on n vertices is K_n, i.e. a complete subgraph with n vertices). A simplicial elimination ordering of a graph (SEO), or a perfect elimination scheme, is an ordering $v_1, \ldots, v_n$ of its vertices such that v_i is simplicial in the subgraph induced by $\{v_i, \ldots, v_n\}$.

An important graph theory result is that a graph is chordal if and only if it has a perfect elimination scheme. An example of a chordless graph and a chordal graph, with a possible SEO, are given in Fig. 2.2. Other example of chordal graphs are shown in Fig. 2.3.

[Felzenszwalb 2005] introduces a representation of 2D shapes (with piecewise-smooth boundaries and no holes) using triangulated polygons (by the means of constrained Delaunay triangulation on the vertices that form



(a) (b)

Figure 2.2: (a) A chordless graph with 8 vertices. (b) A chordal graph with 8 vertices, and a simplicial elimination order (SEO).

the boundary of the shape). This representation has the following nice property: the dual graph of the triangulated polygon (which is a 2-tree) is a tree which bequeaths to the shape vertices the perfect elimination order property (see Fig. 2.3(c)). This property permits the definition of segmentation of objects in images as a minimization problem (using edge information as data term and constraining the deformation of each triangle from the template to the image to be as close to a similarity transform as possible - as smoothness / prior term) that can be solved exactly and globally using dynamic programming, by quantizing the possible locations of the vertices (on a grid for instance). However, complexity is cubic to the number of grid points and the model can not be extended in a straightforward manner to 3D (for instance for medical imaging applications). In a second step, the shape model is extended to a statistical representation, by learning for each triangle a mean and a covariance using a local Procrustes analysis (for each triangle). This work is similar in some aspects to the pictorial structure formulation [Felzenszwalb & Huttenlocher 2005] as both structures (in both works) are chordal graphs, which permits to solve the MAP problem in polynomial time where the complexity is related to the size of the maximum clique in the graph. Thanks to this structure, [Felzenszwalb 2005] define a Bayesian formulation and estimate the parameters and the tree structure

from the data in a maximum likelihood (ML) fashion. They are also able to sample shapes from the posterior probability distribution thanks to this same property. The suggested model is similarity-invariant, deals naturally with occlusions and noise and can handle multiple segmentations. In the same line of development, [Crandall *et al.* 2005] extend the pictorial structure [Felzenszwalb & Huttenlocher 2005] by defining another class of chordal graphs, namely k-fans. These graphs are exactly k-trees with a single base (also called reference part), and we use in our approach the same graph structures (see Fig. 2.3(d), Fig. 2.3(f)). They also extend their Bayesian framework to this case, and learn the statistical parameters (Gaussian distributions on the relative landmarks positions) as well as the reference part (and hence the k-fan structure) from the data. They apply their approach in 2D object recognition. However, in practice, they did not consider similarity invariance (while it can be defined for $k \geq 2$ with respect to the reference part, as they precisely explain), and restricted application to $k \in \{0, 1, 2\}$. This choice is probably guided by the wish of controlling the polynomial complexity of the MAP problem. In our case, the geometric aspects dictate to use $k \geq n + 1$ in $\mathbb{R}^n$ (global graph rigidity issue as explained in the previous section), k is learned from the training set of shapes and efficient approximate optimization techniques [Komodakis *et al.* 2008b] [Kolmogorov 2006] are used to cope with the curse of dimensionality.

A related segmentation approach in medical image analysis is presented by [Seghers *et al.* 2007a]. They define a 2D shape prior using landmarks that form a closed curve. This chain structure enables the use of dynamic programming. Three different prior statistics were tested. The first shape prior estimates a Gaussian distribution of the relative positions of successive landmarks on the curve, and is hence translation-invariant. The second model estimates a Gaussian distribution of the relative positions of successive landmarks on the curve in a local frame. The third model estimates a conditional Gaussian distribution of the relative positions of successive landmarks on the curve with respect to the previous pair of landmarks. The last two models are also rotation-invariant. While the search using the first prior is quadratic in the number of grid points, it is cubic for the other two priors. This framework is extended to 3D in [Seghers *et al.* 2007b] using the first proposed prior with

a mesh of the shape. A heuristic search method is used for the optimization, and is called iterative dynamic programming. More details of this segmentation approach will be discussed in the next chapters. Here again, the authors make an explicit assumption about the graph structure, and this issue is not addressed.

Other Related Approaches

We now comment a few additional related papers. [Gu *et al.* 2007] assume a Gaussian Markov Random Field (GMRF) prior over the landmarks. This prior is fully represented by the mean landmark positions as well as the precision matrix (the inverse of the covariance matrix) over the landmarks. The zero entries in the precision matrix fully determine the graph structure, and account for the conditional independence between the corresponding landmarks. Unfortunately, the precision matrix is often ill-defined in practice (where the number of training examples can be smaller than the problem dimension). They circumvent this structure learning problem by using a Lasso regression technique [Meinshausen & Bühlmann 2006]. This leads to an interesting sparse representation. However, the Gaussian assumption is a shortcoming of this model in case of non-linear shape variations.

[Donner *et al.* 2010] use a MRF shape representation, associated with symmetry-based interest point local descriptors derived from Gradient Vector Flow (GVF). The statistical prior is encoded using Gaussian distributions of the graph edge lengths, and relative graph edge orientation with respect to the landmarks descriptors orientations. However, this rigid-motion invariant shape model does not rely on the training data to learn the graph structure.

Eventually, [Glocker *et al.* 2009] use a learning scheme that is similar to what we present, to deduce from a training set smoothness constraints to be included in a registration algorithm. They use a free form deformation (FFD) approach, and learn the co-dependencies of pairs of control points of the deformation grid. These co-dependencies are defined using the Bhattacharyya divergence between the deformation distribution of each control point. Then, an unsupervised clustering technique [Komodakis *et al.* 2008a] is applied to cluster the control points, according to their relative pairwise Bhattacharyya divergence. The deformation constraints are defined in a part-based model

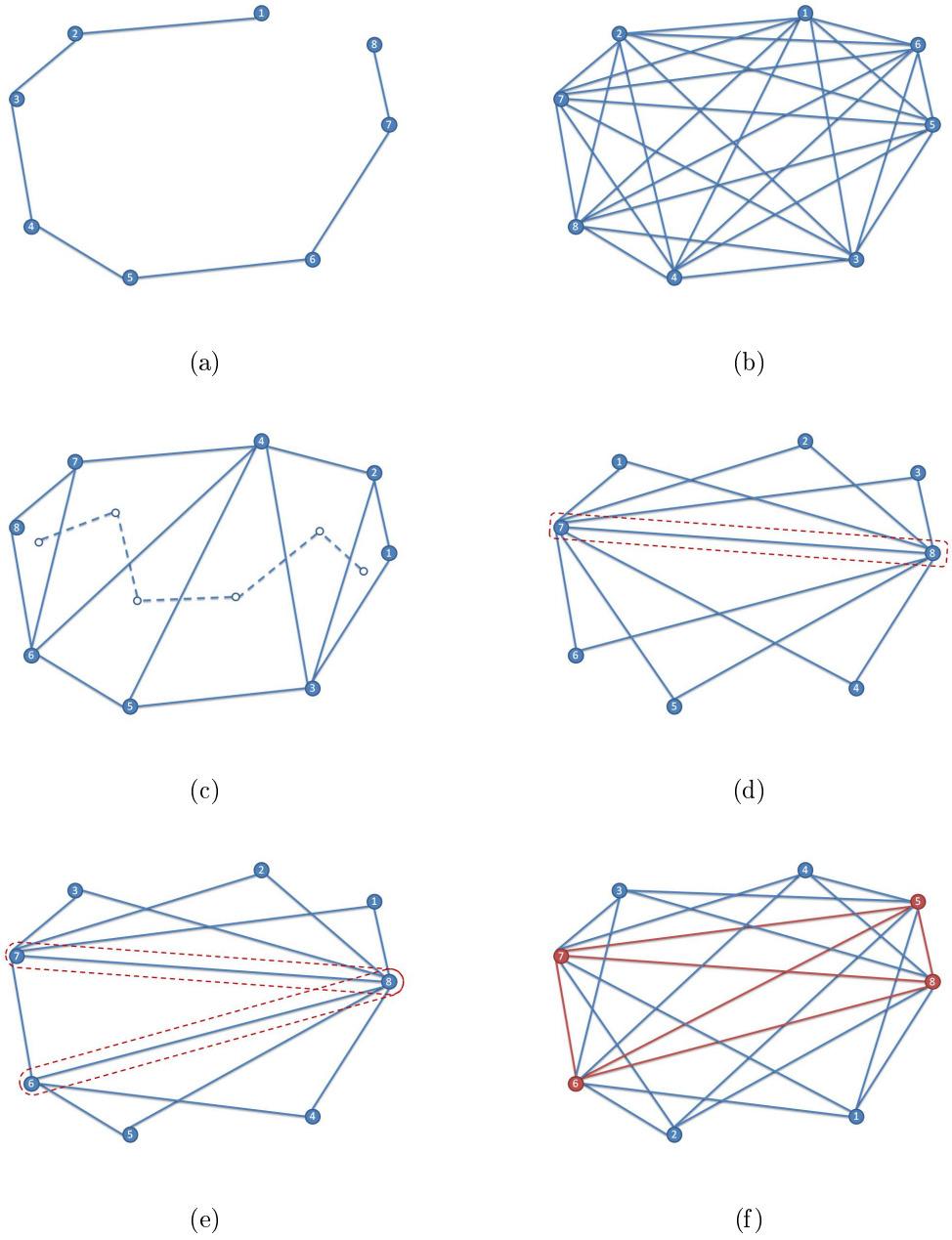


Figure 2.3: Several chordal graph structures with 8 vertices, and a SEO for each graph. (a) A tree. (b) The complete graph. (c) A triangulated polygon and its dual graph. (d) A 2-fan graph - the reference vertices are shown. (e) A 2-tree that is not a 2-fan - the reference cliques are emphasized. (f) A 4-fan.

manner using an incomplete graph, where the cluster centers form a complete graph, and where each cluster forms a complete graph as well. Our approach differs from their in two aspects. We use shape maps [Langs & Paragios 2008] to model the joint behavior of two control points, and this measure takes into account the interactions between this pair of landmarks and the remaining ones in the training set. Hence the shape maps technique is able to bring more information about the co-dependencies than the Bhattacharyya divergence. Then, a major difference lies in the rationale behind the final structure of the graph. We obtain a k-fan graph, and this choice takes into consideration geometric constraints that will be explained in section 2.2.2. The graph proposed by [Glocker *et al.* 2009] is not guaranteed to have the same global rigidity property, which is important for our shape modeling. On the contrary, the FFD registration requirements are different from these shape considerations. Their model permits to deal with the ill-posedness of the registration problem while allowing discontinuities for the deformations of neighboring control points belonging to different clusters.

2.2 A Normalized Chord-Length Statistical Shape Model

We describe here our statistical shape model (SSM) and detail the inherent geometric, manifold learning and graph-theoretic aspects.

2.2.1 Second-Order Approximation and Chord-Length

Knowledge-based segmentation methods are based on the definition of a model which is then combined with image support towards object extraction. Traditional approaches consist of representing the shape using a number of landmarks and learning their behavior using a training set.

In this work, we consider that a shape is defined by its boundary (a curve in 2D, or a surface in 3D) as well as a set of control points. Knowing the positions of these controls points, the shape boundary can be (approximately) retrieved by the use of an interpolation method. The control points can in general belong to the boundary, or not. In practice however, it is natural that such

landmarks are chosen to lie on the contour of the object, which exhibits very often distinctive features. Therefore, we will assume in the following such a hypothesis, meaning that the control points are boundary points.

We represent a shape as a Markov Random Field (MRF). Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a graph where the nodes $\mathcal{V} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ represent the control points positions, and where the set of edges $\mathcal{E}$ represents the conditional dependencies between the landmarks. We assume for the time-being that this graph is complete, that is we take into account all the pairwise dependencies between the landmarks. Hence, $\mathcal{E} = \{(i,j) \mid 1 \leq i < j \leq n\}$. Pursuing the reasoning in [Cremers & Grady 2006] and [Seghers *et al.* 2008], we express the shape probability in terms of co-occurrence probabilities. Using the Bayes rule, we can for example write:

$$\begin{aligned} p(\mathbf{x}_1, \ldots, \mathbf{x}_n) &= p(\mathbf{x}_1)\, p(\mathbf{x}_2 \ldots, \mathbf{x}_n | \mathbf{x}_1) \\ &= p(\mathbf{x}_1)\, p(\mathbf{x}_2|\mathbf{x}_1)\, p(\mathbf{x}_3 \ldots, \mathbf{x}_n | \mathbf{x}_1, \mathbf{x}_2) \\ &= \ldots \\ &= p(\mathbf{x}_1)\, p(\mathbf{x}_2|\mathbf{x}_1)\, p(\mathbf{x}_3|\mathbf{x}_1, \mathbf{x}_2) \cdots p(\mathbf{x}_n | \mathbf{x}_1, \ldots, \mathbf{x}_{n-1}) \quad . \end{aligned} \tag{2.1}$$

First, seeking a representation that does not depend on an absolute position, the probability of co-occurrence of a pair of control points can be then written as the probability of their difference vector, or:

$$p(\mathbf{x}_i, \mathbf{x}_j) = p(\mathbf{x}_i - \mathbf{x}_j) \quad . \tag{2.2}$$

This hypothesis also leads to a uniform probability distribution for the absolute points positions or, a constant $p(\mathbf{x}_i)$, for all $1 \leq i \leq n$. It also means that the conditional probabilities are equal for any pair of landmarks, or $p(\mathbf{x}_i|\mathbf{x}_j) = p(\mathbf{x}_j|\mathbf{x}_i)$, $\forall\ 1 \leq i, j \leq n$. [Seghers *et al.* 2008] proceed similarly to induce independence with respect to absolute location.

We will consider the second order approximation of the shape probability, assuming that the positions of two landmarks is independent from any third landmark, or:

$$p(\mathbf{x}_i, \mathbf{x}_j | \mathbf{x}_k) \approx p(\mathbf{x}_i, \mathbf{x}_j), \ \forall k \neq i, j \quad . \tag{2.3}$$

Then, applying this property to equation (2.1) by choosing particular point pairs leads to:

$$p(\mathbf{x}_1, \ldots, \mathbf{x}_n) \propto p(\mathbf{x}_2|\mathbf{x}_1)\, p(\mathbf{x}_3|\mathbf{x}_2) \cdots p(\mathbf{x}_n|\mathbf{x}_{n-1}) \quad . \tag{2.4}$$

Equation (2.4) is one particular choice of decomposing the joint probability. This choice corresponds to specifying a spanning tree of the graph $\mathcal{G}$. Considering all such possibilities is equivalent to performing the previous simplification using all possible spanning trees of $\mathcal{G}$. Let us multiply all these resulting equations, we obtain thanks to the simplification (2.3):

$$p\left(\mathbf{x}_1, \ldots, \mathbf{x}_n\right)^{\tau(\mathcal{G})} \propto \prod_{i<j} p\left(\mathbf{x}_i, \mathbf{x}_j\right)^{\nu} \quad , \tag{2.5}$$

where $\tau(\mathcal{G})$ is the number of all possible spanning trees of $\mathcal{G}$, and ν is the number of appearances of each pair in all these possible equations. This number is obviously the same for each pair in a complete graph. Let us simplify the above equation using double counting. A spanning tree is a sequence of $n-1$ elements (or edges). Then the total number of elements used to form all the spanning trees is $(n-1) \times \tau(\mathcal{G})$. There are $\frac{n(n-1)}{2}$ edges in a complete graph, and each edge appears ν times in all the spanning trees, which makes a total of $\frac{n(n-1)}{2} \times \nu$ elements. Then, $\nu = \tau(\mathcal{G}) \times \frac{2}{n}$. This gives a simple form to the product (2.5), or:

$$p\left(\mathbf{x}_1, \ldots, \mathbf{x}_n\right) \propto \prod_{i<j} p\left(\mathbf{x}_i, \mathbf{x}_j\right)^{\frac{2}{n}} \quad . \tag{2.6}$$

We would like to enforce invariance properties to our shape model, with respect to translations, rotations, and global scale changes. Such similarity-invariance characteristics enable to dismiss object changes due to the pose variations and permit to learn the intrinsic variability of the object. Discarding the transformation parameters in the inference step from the unknowns is also an advantage that makes an invariant shape model more tractable and easier to use.

Invariance with respect to rotation is equivalent to ignoring any orientation direction, that to consider the Euclidean norm of the difference vector, or:

$$p\left(\mathbf{x}_i, \mathbf{x}_j\right) = p\left(\left\|\mathbf{x}_i - \mathbf{x}_j\right\|\right) \quad . \tag{2.7}$$

Eventually, let us consider $\widetilde{d} = \frac{2}{n(n-1)} \sum_{i<j} \left\|\mathbf{x}_i - \mathbf{x}_j\right\|$ the mean Euclidean distance between the control points. This mean chord length can be considered as an estimate of the shape scale. Hence, dividing the chord lengths by $\widetilde{d}$ gives a representation that is invariant to scale changes as well, or:

$$p\left(\mathbf{x}_i, \mathbf{x}_j\right) = p\left(\frac{\left\|\mathbf{x}_i - \mathbf{x}_j\right\|}{\widetilde{d}}\right) \quad . \tag{2.8}$$

Strictly speaking, $\widetilde{d}$ depends on all the control points. Therefore, the ambiguity carried by equation (2.8) will be explained later in the algorithmic sections. By combining equations (2.6) and (2.8) we define the graph probability as:

$$p\left(\mathbf{x}_1,\ldots,\mathbf{x}_n\right) \propto \prod_{i<j} p\left(\frac{\|\mathbf{x}_i-\mathbf{x}_j\|}{\widetilde{d}}\right)^{\frac{2}{n}} \quad , \tag{2.9}$$

Upon taking the negative log of the above equation, the graph energy is defined, up to an additive constant ν', as:

$$E_{shape}\left(\mathbf{x}_1,\ldots,\mathbf{x}_n\right) \approx -\frac{2}{n}\sum_{i<j}\log\left(p\left(\frac{\|\mathbf{x}_i-\mathbf{x}_j\|}{\widetilde{d}}\right)\right)+\nu' \quad . \tag{2.10}$$

Let us consider now a set $\mathcal{S} = \{s_1,\ldots,s_m\}$ of m instances of the object, where each example is represented using n control points, i.e. $s_u = \{x_1^u,\ldots,x_n^u\}, \forall u \in \{1,\ldots,m\}$. Hence, $\forall i \in \{1,\ldots,n\}$, the set $\mathcal{X}_i = \{x_i^1,...,x_i^m\}$ represents instances of the i^{th} control point of the shape. In practice, this training set is obtained by manually labeling the landmarks for each instance of the shape, or by deducing the landmarks from the registration between a labeled shape and a set of non-labeled shapes. Note that an alignment of the shape is not needed before the learning phase since our representation is intrinsically invariant to similarity transforms.

Then, given a statistical model, we learn from the training set the probability density distributions of the relative positions of the control points $p_{ij} \equiv p\left(\mathbf{x}_i,\mathbf{x}_j\right)$. These $\frac{n(n-1)}{2}$ densities enable us to describe the information contained in the training set. In practice, we used a Gaussian kernel density estimation to compute the probability distributions.

One can question the use of the NCL representation for shape modeling. Recall the observation made in section 2.1.3 with respect to the ability of EDM to model shape in an isometric-invariant manner [Dattorro 2005]. The scaling factor $\widetilde{d}$ that we add simply extends the result to the similarity-invariant case. However, this representation suffers from redundancy and could be computationally expensive and hard to optimize during inference. We describe in the following the learning technique we use to remove the redundancy from our model, to obtain a sparser representation. But before proceeding, we discuss some geometric aspects that will enlighten the redundancy encapsulated by the complete graph.

2.2.2 Graph Rigidity and the Exact Matching Problem

This section is built on the work of [Caetano *et al.* 2006]. We first recall the result that they provide, and then extend it to a more general case that is of particular interest in our modeling.

The Point Pattern Matching Problem

Let $\mathcal{P}_s = (p_1, \ldots, p_s)$ a set of s "source" points in $\mathbb{R}^n$, and $\mathcal{Q}_t = (q_1, \ldots, q_t)$ a set of t "target" points in $\mathbb{R}^n$ where in general $s \leq t$. The point pattern matching problem consists in finding a mapping $m : \mathcal{P}_s \to \mathcal{Q}_t$ that minimizes a distortion measure between $\mathcal{P}_s$ and $m(\mathcal{P}_s)$. This can be formulated mathematically as minimizing:

$$\mathcal{D}_{K_s}(m) = \sum_{i=1}^{s} \sum_{j=1}^{s} \rho(d_{ij}^s, d_{m(i)m(j)}^t) \ , \tag{2.11}$$

where ρ is a dissimilarity measure, and $d_{ij}^s = \|p_i - p_j\|$, $d_{ij}^t = \|q_i - q_j\|$ are Euclidean distances. In general, due to the noisy data, it is not possible to achieve a perfect matching, i.e. an exact matching. An exact matching is a mapping m such that $d_{ij}^s = d_{m(i)m(j)}^t$ for all $1 \leq i, j \leq s$. Hence, in the perfect matching case problem, ρ is chosen to be a binary function that checks the equality of its arguments. The pattern point matching can be written as a weighted complete graph matching, where the edge weights correspond to the Euclidean distance between the embedded vertices in $\mathbb{R}^n$. In the general case, if only a subset $\mathcal{E}_s$ of the pairs of source points in $\mathcal{P}_s$ are used to evaluate the matching distortion, this corresponds to the graph matching problem formulated on the graph $\mathcal{G}_s = (\mathcal{P}_s, \mathcal{E}_s)$, or minimizing:

$$\mathcal{D}_{\mathcal{G}_s}(m) = \sum_{(i,j) \in \mathcal{E}_s} \rho(d_{ij}^s, d_{m(i)m(j)}^t) \ . \tag{2.12}$$

In the remainder of this section, we will use indifferently the terms *graph* or *graph embedding*.

Global Graph Rigidity

[Caetano *et al.* 2006] define the global rigidity of a graph in an intuitive manner.

Definition 2.5. A graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ in $\mathbb{R}^n$ is said to be globally rigid if the lengths of the edges in $\mathcal{E}$ determine uniquely the lengths of the edges of the complement graph (the "missing" edges).

Then, they show that a k-tree with all base k-cliques in general position in $\mathbb{R}^{k-1}$ (i.e. do not lie in a $k-2$-dimensional vector subspace) is globally rigid in $\mathbb{R}^{k-1}$. They use this property to prove that in the perfect matching case in $\mathbb{R}^{k-1}$, the k-tree model is equivalent to the complete model.

Theorem 2.2. *In the exact matching case in $\mathbb{R}^{k-1}$, if $\mathcal{G}_s = (\mathcal{P}_s, \mathcal{E}_s)$ is a k-tree, then a mapping m which minimizes $\mathcal{D}_{\mathcal{G}_s}(m)$ also minimizes $\mathcal{D}_{K_s}(m)$.*

We now extend this property to a larger case of graphs that will appear to be useful later in the manifold learning part. To this end, we invoke the following result, the proof of which is reported in Appendix C:

Lemma 2.1 ([Cai & Maffray 1993])**.** *Every k'-tree $\mathcal{T}'$ with $k' \geq k \geq 1$ possesses a spanning k-tree $\mathcal{T}$ with the property (P) that every k'-clique of $\mathcal{T}'$ contains a k-clique of $\mathcal{T}$.*

Then, it follows that the property proven by [Caetano *et al.* 2006] for the k-trees in $\mathbb{R}^{k-1}$ is generalizable. We suppose that all base k-cliques are in general position in $\mathbb{R}^{k-1}$. Then:

Proposition 2.1. *In the exact matching case in $\mathbb{R}^{k-1}$, if $\mathcal{G}_s = (\mathcal{P}_s, \mathcal{E}_s)$ is a k'-tree, with $k' \geq k$, then a mapping m which minimizes $\mathcal{D}_{\mathcal{G}_s}(m)$ also minimizes $\mathcal{D}_{K_s}(m)$.*

Proof. The mapping m minimizes $\mathcal{D}_{\mathcal{G}_s}(m)$. Then for all $(i,j) \in \mathcal{E}_s$, $d^s_{ij} = d^t_{m(i)m(j)}$. $\mathcal{G}_s$ being a k'-tree, with $k' \geq k$, then (lemma 2.1) there exists a k-tree $\mathcal{G}^k_s = \left(\mathcal{P}_s, \mathcal{E}^k_s\right)$, such that $\mathcal{E}^k_s \subset \mathcal{E}_s$. Subsequently, for all $(i,j) \in \mathcal{E}^k_s$, $d^s_{ij} = d^t_{m(i)m(j)}$, which is equivalent to saying that the mapping m minimizes $\mathcal{D}_{\mathcal{G}^k_s}(m)$. $\mathcal{G}^k_s$ being a k-tree in $\mathbb{R}^{k-1}$, then (theorem 2.2) the mapping m also minimizes $\mathcal{D}_{K_s}(m)$, which completes the proof. □

To summarize, a k-tree with all subsets of size $n+1$ of the base k-cliques in general position in $\mathbb{R}^n$, with $k > n$, is equivalent to the complete graph model in the exact matching case problem. This result shows that such structures encode sufficient information and dependencies to correctly model the

shape. It also suggests that the complete graph model is redundant, and hence can be made sparser. Moreover, as [Caetano *et al.* 2006] show in their experiments, different k-tree models lead to varying accuracy in the non exact graph matching case. For this reason, a method of selecting a particular k-tree with optimal properties is advantageous in their case. In the next step of our modeling we tackle the redundancy problem, and we build a k-tree graph structure where the parameter k is learned from the data, and where the particular structure of the graph is related to the training shape population.

2.2.3 Removing Redundancy from the Model

The task of eliminating the redundancy in the model, while preserving its ability to represent the data, is related to the minimum description length principle on one hand, and can be thought as a spectral clustering problem on the other hand. We aim to obtain as compact a model as possible assuming that the high dimensional data space can be approximated by a lower dimensional embedded manifold, which reduces the dimension of the problem significantly. In other words, we would like to probe the population of training shapes for statistical dependencies between the control points variations.

Shape maps [Langs & Paragios 2008] handle precisely these two aspects, and are learned from the data in a way closely related to the diffusion maps [Coifman & Lafon 2006], but using the compactness of models that describe subsets of the entire data instead of the spatial distances or similarities between individual points. Therefore, we compute the shape map of the control points, using the training set $\{\mathcal{X}_1, \ldots, \mathcal{X}_n\}$, and then we cluster the control points according to their mutual shape map distances. We start first by reviewing the theoretical foundations of diffusion maps.

Diffusion Maps

Diffusion maps is a technique based on diffusion processes that aims to define a metric, namely the diffusion distance, that measures meaningfully the connectivity of data sets, with respect to their geometry. An efficient representation of the geometric structures complexity is given in [Coifman & Lafon 2006], through the eigenfunction of Markov matrices. Formally, let X be the (dis-

crete) data set, and let $\kappa : X \times X \to \mathbb{R}$ be a symmetric and positivity preserving kernel. Then a reversible Markov chain on X is constructed, with $\eta(x,y) = \frac{\kappa(x,y)}{\delta(x)}$ being the anisotropic transition kernel, and $\delta(x) = \sum_y \kappa(x,y)$. The associated Markov matrix P (with entries $\eta(x,y)$) encode geometric information about the data and permits its analysis in a multi-scale fashion. The kernel $\eta_t(x,y)$ (which corresponds to the entries of the t^{th} power of P) represents the transition probability from the node x to the node y in t time steps. The diffusion process reflects random walks on the Markov chain according to these transition probabilities. After the diffusion process, the posterior distribution $u \mapsto \eta_t(x,u)$ characterizes the connectivity of the node x with respect to the other nodes. A weighted L^2 distance on these distributions defines the diffusion distance between two nodes x and y at time t, or:

$$\begin{aligned} D_t(x,y)^2 &= \|\eta_t(x,.) - \eta_t(y,.)\|_{L^2}^2 \\ &= \sum_u \frac{(\eta_t(x,u) - \eta_t(y,u))^2}{\pi(u)} \quad , \end{aligned} \tag{2.13}$$

where $\pi(u) = \frac{\delta(u)}{\sum_v \delta(v)}$ is the probability of u in the unique stationary distribution. The larger the number of high probability paths of length t connecting the nodes x and y, the lower will be $D_t(x,y)$. Then, the authors relate the diffusion distance to the eigendecomposition of the operator P. Let $\lambda_1 \geq \cdots \geq \lambda_i \geq \ldots$ be the eigenvalues of P, and let $\{\Psi_1, \ldots, \Psi_i, \ldots\}$ be its eigenfunctions. Then, a tractable expression of the diffusion distance is given by:

$$\begin{aligned} D_t(x,y)^2 &= \sum_i \lambda_i^{2t} (\Psi_i(x) - \Psi_i(y))^2 \\ &\approx \sum_{i=1}^{N} \lambda_i^{2t} (\Psi_i(x) - \Psi_i(y))^2 \quad , \end{aligned} \tag{2.14}$$

and a good approximation is obtained by considering the top N eigenfunctions.

Shape Maps

Shape maps [Langs & Paragios 2008] bridge the gap between behavior analysis and manifold learning. It uses the diffusion map technique with an appropriate kernel κ. Let us first recall the notations of section 2.2.1.

The data are represented by a graph, where the vertices/control points are $\mathcal{V} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$. The pool of training data is $\{\mathcal{X}_1, \ldots, \mathcal{X}_n\}$, where $\mathcal{X}_i = \{x_i^1, ..., x_i^m\}$ represents instances of the i^{th} landmark of the shape.

Shape maps use the description length (DL) principle to assess the mutual coherence of the landmarks behavior. The authors compute the compactness of small sub-models of landmarks subsets, that capture the non-rigid local variations, in order to define a measure of similarity between landmarks, in terms of statistical dependence. The description length consists of the cost $\mathcal{L}$ of transmitting a model $\mathcal{M}$ and the data D encoded with the model: $\mathcal{L}(D, \mathcal{M}) = \mathcal{L}(\mathcal{M}) + \mathcal{L}(D|\mathcal{M})$. For each pair of landmarks, $(\mathbf{x}_i, \mathbf{x}_j)$, and for a defined kernel size s,

$$\mu_s(i,j) = \min_{\mathcal{M}}(\mathcal{L}(\mathcal{M})|(i,j) \subset \mathcal{M} \ \& \ |\mathcal{M}| = s) \tag{2.15}$$

is the minimum description length of models encompassing the two landmarks and $s-2$ other landmarks. The considered sub-models were Gaussian, and were determined using PCA on the corresponding s landmarks after alignment of the data. If $\mathbf{x}_i$ and $\mathbf{x}_j$ belong to a cluster of coherently behaving landmarks, then this will be detected by a low value of $\mu_s(i,j)$. In practice, we considered $s = 3$ and evaluated the DL of all sub-models of this size.

Next, a Markov chain that reflects the model structure is created, with a node corresponding to every landmark $\mathbf{x}_i$, through the definition of the kernel $\kappa(i,j) = e^{-\frac{\mu_s(i,j)}{\varepsilon}}$. The diffusion maps technique provides an embedding $\mathbf{\Psi}_t : \mathcal{V} \to \mathbb{R}^N$ of the landmarks in an N dimensional Euclidean space:

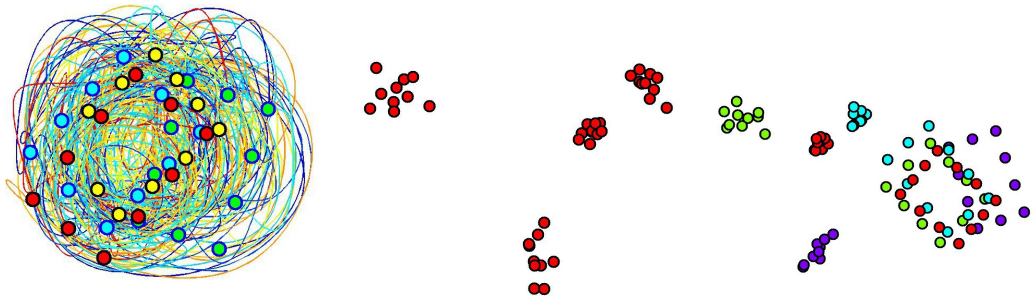
$$\mathbf{x}_i \mapsto \mathbf{\Psi}_t(\mathbf{x}_i) = \begin{pmatrix} \lambda_1^t \Psi_1(i) \\ \vdots \\ \lambda_N^t \Psi_N(i) \end{pmatrix} \ . \tag{2.16}$$

Thereby, the diffusion distance in equation (2.14) can be expressed as:

$$D_t(\mathbf{x}_i, \mathbf{x}_j) = \|\mathbf{\Psi}_t(\mathbf{x}_i) - \mathbf{\Psi}_t(\mathbf{x}_j)\| \ . \tag{2.17}$$

An implementation of this method is available online[2]. [Langs & Paragios 2008] use a synthetic example that illustrates clearly the use of shape maps. We take this experiment up, and use it to illustrate

[2] http://people.csail.mit.edu/langs/Georg_Langs/ShapeMaps.html



(a) (b) (c)

Figure 2.4: A synthetic example illustrating the use of shape maps and unsupervised clustering. (a) The generated synthetic data: 4 boxes with varying aspects rotate around each other independently. (b) The shape maps embedding of the data is shown using the first 3 dimensions. We can see that 4 clusters appear, which corresponds to the generated data. (c) The unsupervised LP-based clustering (section 2.2.4) is performed using the obtained diffusion distance. It finds automatically the number of clusters and correctly identifies them, which permits to separate the boxes in the data.

the following unsupervised clustering method as well (Fig. 2.4). A synthetic data of 4 boxes with varying size attributes is generated. These objects rotate independently around each others (Fig. 2.4(a)). The shape maps technique is applied to analyze the points trajectories and provides an embedding of the data in a Euclidean space. This mapping is shown in Fig. 2.4(b), where $N = 3$ for visualization purposes.

In terms of the joint model, low distance values suggest high compactness for a representation that encompasses landmarks $\mathbf{x}_i$ and $\mathbf{x}_j$. We considered in practice a time step $t = 1$. Having obtained these behavior-explaining distances, we would like to cluster the control points accordingly. A recent clustering algorithm [Komodakis *et al.* 2008a] was used for this final task, and is described in the section 2.2.4. The obtained clusters reflect the interdependencies between the control points, and refer to the parts of the object that have highly-correlated relative displacements.

2.2.4 Unsupervised Clustering Using Linear Programming

Clustering refers to the process of organizing a set of objects into groups, where the members of each group are as similar to each other as possible. More formally, a common definition for clustering is the following one: suppose we are given a set of objects $\mathcal{V} = \{v_1, \ldots, v_n\}$ endowed with a distance function d that can measure the dissimilarity between any two objects $v_i, v_j \in \mathcal{V}$. In such a case, the goal of clustering is to choose K objects, say, $\{c_1, c_2, \ldots, c_K\}$ from $\mathcal{V}$ (these will be referred to as the *clusters centers*), so that the obtained sum of distances between each object and its closest center is minimized, or:

$$\min_{c_1,c_2,\ldots,c_K} \sum_{v_i \in \mathcal{V}} \min_{c_k} d(v_i, c_k) \quad . \tag{2.18}$$

A common drawback of many popular clustering techniques (such as the K-means algorithm) is that they need to be given the number K of clusters beforehand. However, this is problematic as this number is very often not known in advance. To address this issue, we will let this number vary as well and change the objective function of clustering so as to assign a penalty (denoted by $d(v_i, v_i)$) whenever an object v_i is chosen as a cluster center, or:

$$\min_{K,c_1,c_2,\ldots,c_K} \sum_{v_i \in \mathcal{V}} \min_{c_k} d(v_i, c_k) + \sum_{c_k} d(c_k, c_k) \quad . \tag{2.19}$$

Another very bad symptom of many clustering techniques is that they are particularly sensitive to initialization. For instance, the K-means algorithm (which is one of the most commonly used clustering techniques) is doomed to fail if its initial cluster centers happen not to be near the actual cluster centers. To address this very important issue, we have used a novel clustering method. The main idea behind our method is to first formulate the clustering

as a linear integer program as follows:

$$\min \sum_{i=1}^{n} \sum_{j=1}^{n} d(v_i, v_j) x_{ij} \tag{2.20}$$

$$\text{s.t.} \sum_{j=1}^{n} x_{ij} = 1, \ \forall i \tag{2.21}$$

$$x_{ij} \leq x_{jj}, \ \forall i \neq j \tag{2.22}$$

$$x_{ij} \in \{0, 1\}, \ \forall i, j \tag{2.23}$$

In the above formulation, the binary variable x_{ij} (with $i \neq j$) indicates whether object v_i has been assigned to cluster center v_j or not, while the binary variable x_{jj} indicates whether object v_j has been chosen as a cluster center or not. It is then very easy to prove that the above linear integer program is actually equivalent to minimizing the objective function (2.19) of clustering. To this end, it suffices to observe that (2.21) simply expresses the fact that each object v_i can be assigned to exactly one cluster center v_j, while (2.22) simply expresses the fact that if any object v_i has been assigned to an object v_j, then v_j must be chosen as cluster center. To obtain an approximately optimal solution to the above integer program, we will then rely on first solving its linear programming relaxation and then "rounding" the relaxed solution in an appropriate manner. More details about the formulation of the problem and its optimization are given in [Komodakis *et al.* 2008a]. In the validation section of [Komodakis *et al.* 2008a], it is added that a constant penalty cost, roughly set to the median of the distances $d(v_i, v_j)$ is used in the experiments. We also considered the same penalty value for our tests.

In our case, the set of objects correspond to the control points $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, and the distance d corresponds to the aforementioned diffusion distance D_t in section 2.2.3. As an example of application, in the previously introduced synthetic case (Fig. 2.4(c)), this method is able to find the actual number of clusters used to generate the data, and clusters it correctly.

2.2.5 The Shape Model

The unsupervised clustering technique described in section 2.2.4 is applied using the aforementioned shape maps distances (section 2.2.3). The process

output is a number k of clusters (in practice, $k \geq 3$ in 2D, and $k \geq 4$ in 3D), and a set of k cluster centers, chosen among the control points. Then, the final step to build our shape model consists in defining the structure of the graph, in a principled manner.

A cluster of landmarks is defined as a set of control points of highly correlated behavior. Thus, such a group of points carries a high amount of redundant information in terms of shape variability. This observation drives us to factor out the redundancy using the obtained clusters, through the center landmarks. The cluster centers being the best representatives of their respective groups, we make the mild assumption that for a given control point, that is not a cluster center, the joint landmark-cluster information can be encoded as landmark-cluster center information. Hence, the landmarks will be considered to be conditionally independent, knowing the cluster centers. The global coherence of the model is kept thanks to the reference part, that is the cluster centers, that form a clique of size k. The obtained graph is a k-tree with a single base k-clique. It is called a k-fan, following [Crandall *et al.* 2005]. The novelty here consists in the method that defines automatically from the training data the number of clusters and their centers. To each one of these pairs $(\mathbf{x}_i, \mathbf{x}_j)$ we associate a probability density distribution p_{ij} learned from the training set as previously stated in section 2.2.1. Our k-fan graph, being chordal, benefits from the simplicial elimination ordering (SEO) property, and permits theoretically the development of a Bayesian framework for segmentation for instance. Dynamic Programming-based approaches ensure finding a global optimum. However, the maximum clique size of a k-tree being $k+1$, these polynomial time algorithms are not usable in practice for huge memory requirements and extreme slowness. For these reasons, we resort to approximate optimization techniques for inference on the k-fan graph.

We will denote the k-fan graph by $\mathcal{G}_k = (\mathcal{V}, \mathcal{E}_k)$. Without loss of generality, we can assume that the base k-clique corresponds to the k first nodes in $\mathcal{V}$. Let $\mathcal{R}_k$ be the subset of edges in $\mathcal{E}_k$ that correspond to the base k-clique. The k-fan model assumes that a node $\mathbf{x}_i$ is conditionally independent from a node $\mathbf{x}_j$ (with $k+1 \leq i < j \leq n$) knowing the base k-clique. To derive the joint probability of our shape model, we use the same reasoning presented in section 2.2.1, by taking into account this conditional independence assumption.

Similarly to equation (2.1), using the Bayes rule, we have:

$$\begin{aligned} p(\mathbf{x}_1, \ldots, \mathbf{x}_n) =& p(\mathbf{x}_1)\, p(\mathbf{x}_2|\mathbf{x}_1) \cdots p(\mathbf{x}_k|\mathbf{x}_1, \ldots, \mathbf{x}_{k-1}) \\ & p(\mathbf{x}_{k+1}|\mathbf{x}_1, \ldots, \mathbf{x}_k) \cdots p(\mathbf{x}_n|\mathbf{x}_1, \ldots, \mathbf{x}_{n-1}) \quad . \end{aligned} \tag{2.24}$$

The aforementioned independence assumption allows the following simplification of equation (2.24):

$$\begin{aligned} p(\mathbf{x}_1, \ldots, \mathbf{x}_n) =& p(\mathbf{x}_1)\, p(\mathbf{x}_2|\mathbf{x}_1) \cdots p(\mathbf{x}_k|\mathbf{x}_1, \ldots, \mathbf{x}_{k-1}) \\ & p(\mathbf{x}_{k+1}|\mathbf{x}_1, \ldots, \mathbf{x}_k)\, p(\mathbf{x}_{k+2}|\mathbf{x}_1, \ldots, \mathbf{x}_k) \cdots p(\mathbf{x}_n|\mathbf{x}_1, \ldots, \mathbf{x}_k) \quad . \end{aligned} \tag{2.25}$$

Then, by applying the second order assumption in (2.3), we can for example write:

$$\begin{aligned} p(\mathbf{x}_1, \ldots, \mathbf{x}_n) \propto & \overbrace{p(\mathbf{x}_2|\mathbf{x}_1)\, p(\mathbf{x}_3|\mathbf{x}_2) \cdots p(\mathbf{x}_k|\mathbf{x}_{k-1})}^{\text{subset of } \mathcal{R}_k} \\ & \underbrace{p(\mathbf{x}_{k+1}|\mathbf{x}_1)\, p(\mathbf{x}_{k+2}|\mathbf{x}_1) \cdots p(\mathbf{x}_n|\mathbf{x}_1)}_{\text{subset of } \mathcal{E}_k \setminus \mathcal{R}_k} \quad . \end{aligned} \tag{2.26}$$

Equation (2.26) corresponds to specifying a spanning tree of the base k-clique, and choosing for each node that does not belong to the base k-clique a neighbor $\mathbf{x}_i$, with $1 \leq i \leq k$. Upon multiplying all the equations that result from this construction, we obtain the following expression (which is a modified version of the equation (2.5)):

$$p(\mathbf{x}_1, \ldots, \mathbf{x}_n)^{\xi} \propto \prod_{(i,j) \in \mathcal{R}_k} p(\mathbf{x}_i, \mathbf{x}_j)^{\nu} \prod_{(i,j) \in \mathcal{E}_k \setminus \mathcal{R}_k} p(\mathbf{x}_i, \mathbf{x}_j)^{\nu'} \quad . \tag{2.27}$$

Note that the pairs (i, j) with $k+1 \leq i < j \leq n$ do not appear in this expression due to their conditional independence assumption. A node in the base k-clique has a degree $n-1$ (i.e. $n-1$ neighbors), whereas a node that does not belong to the base k-clique has a degree k. Therefore, ν and ν' are different in general. Let ζ be the number of spanning trees of the base k-clique. Then we clearly have that $\xi = \zeta k^{n-k}$, since there are $n-k$ non-reference nodes having a degree equal to k. Let η be the number of appearances of an edge in $\mathcal{R}_k$ in the total number of spanning trees of the base k-clique. Then, as showed in section 2.2.1, $\eta = \frac{2}{k}\zeta$. Now, it is easy to count the number of edge appearances ν and ν' in the product (2.27). On one hand, we can see that

Algorithm 1 Build the NCL Shape Model

1: Apply shape maps to the training data.
2: Cluster the control points $\mathbf{x}_i$ using the LP-based unsupervised clustering algorithm, through their pairwise diffusion distances $D_t(\mathbf{x}_i, \mathbf{x}_j)$. The output of this step is k clusters, and k prototype centers.
3: Build the k-fan graph $\mathcal{G}_k = (\mathcal{V}, \mathcal{E}_k)$ that represents the data by linking each landmark $\mathbf{x}_i$ to all the k cluster centers $\mathbf{x}_j$.
4: For every pair of control points $(\mathbf{x}_i, \mathbf{x}_j)$ such that $(i,j) \in \mathcal{E}_k$, learn a Gaussian kernel density estimation of the probability density function $p(\mathbf{x}_i, \mathbf{x}_j) = p\left(\frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{\widetilde{d}}\right)$.

$\nu = \eta k^{n-k}$, which leads to $\nu = \frac{2}{k}\xi$. On the other hand, $\nu' = \zeta k^{n-k-1}$ which gives $\nu' = \frac{1}{k}$. Then we can write equation (2.27) more explicitly:

$$p(\mathbf{x}_1, \ldots, \mathbf{x}_n) \propto \prod_{(i,j)\in\mathcal{R}_k} p(\mathbf{x}_i, \mathbf{x}_j)^{\frac{2}{k}} \prod_{(i,j)\in\mathcal{E}_k\setminus\mathcal{R}_k} p(\mathbf{x}_i, \mathbf{x}_j)^{\frac{1}{k}} \quad . \tag{2.28}$$

To simplify the notations, we introduce the weighting γ_{ij} defined by:

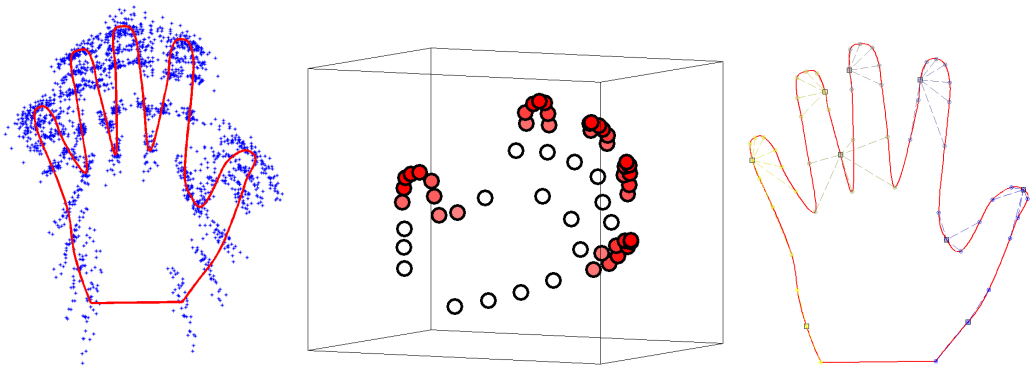
$$\gamma_{ij} = \begin{cases} \frac{2}{k} & \text{if} \quad (i,j) \in \mathcal{R}_k \\ \frac{1}{k} & \text{if} \quad (i,j) \in \mathcal{E}_k \setminus \mathcal{R}_k \end{cases} \tag{2.29}$$

The shape energy in equation (2.10) can now be approximated as:

$$E_{shape}(\mathbf{x}_1, \ldots, \mathbf{x}_n) \approx -\sum_{(i,j)\in\mathcal{E}_k} \gamma_{ij} \log\left(p\left(\frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{\widetilde{d}}\right)\right) \quad . \tag{2.30}$$

Algorithm 1 summarizes the construction of our shape model. Recall that the thread leading to our shape model is threefold:

- Geometry: the k-fan MRF has an optimal geometric property in the exact matching case, that is related to its global rigidity.
- Manifold learning: the k-fan MRF particular structure, and the number k of fans are deduced from the population of training shapes.
- Graph theory - optimization: although with very limited usefulness, due to the curse of dimensionality, the k-fan MRF benefits from the advantages of the larger class of chordal graphs.



(a) (b) (c)

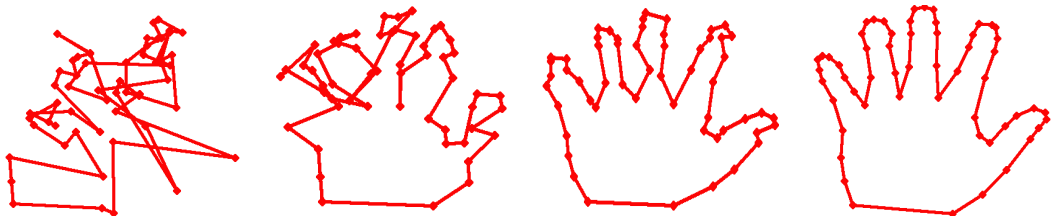
Figure 2.5: Hand model construction. (a) The considered training set. (b) The projection of the control points on the first 3 shape map dimensions. The density of the landmarks is estimated by a Gaussian kernel and is color-coded. (c) Control points clustered in 9 clusters: centers are represented by squares.

2.2.6 Experimental Validation

Drawing Samples from the Learned Distribution

A first application that is usually done after building a statistical shape model consists in generating synthetic shapes. This can be done by estimating the probability distribution over a grid, and then drawing sample from this distribution. We formulate the probability estimation similarly to [Felzenszwalb & Huttenlocher 2005]. The probability of our shape model is given by equation (2.28). We will first compute the marginal probability of the reference part (the base k-clique) $p(\mathbf{x}_1, \ldots, \mathbf{x}_k)$. Then we will sample a location for the landmarks belonging to the base k-clique. Next, we will sample from $p(\mathbf{x}_i|\mathbf{x}_1, \ldots, \mathbf{x}_k)$ a location for the non reference landmarks. We start by marginalizing (2.28) over the non reference part:

$$p(\mathbf{x}_1, \ldots, \mathbf{x}_k) \propto \sum_{k+1} \cdots \sum_{n} p(\mathbf{x}_1, \ldots, \mathbf{x}_n) \quad . \tag{2.31}$$



(a) (b) (c) (d)

Figure 2.6: Hand shape prior applied to random points.(a)–(d) Deformation of a random point cloud according to the shape prior term.

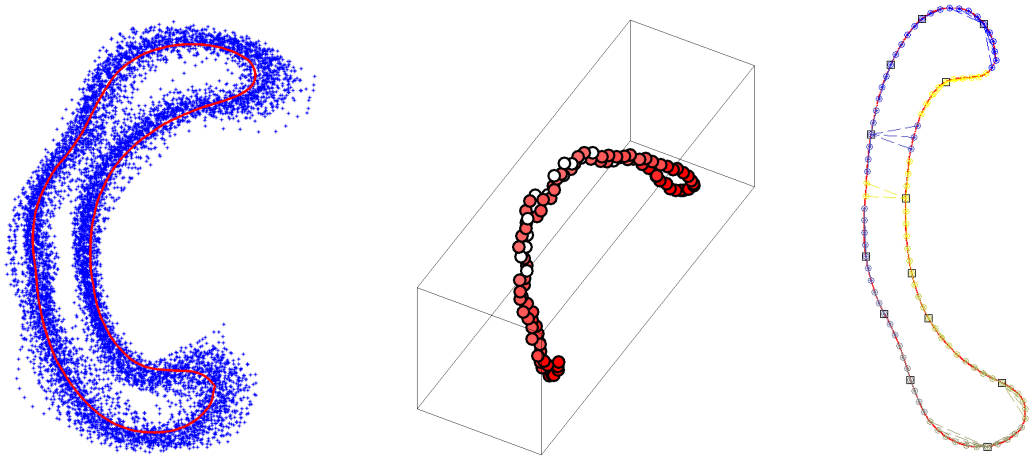
The exponential number of operations described by the above equation can be reduced to polynomial, by rearranging the terms following a SEO, or:

$$p\left(\mathbf{x}_1,\ldots,\mathbf{x}_k\right) \propto \left(\sum_{k+1}\prod_{i=1}^{k} p\left(\mathbf{x}_{k+1},\mathbf{x}_i\right)^{\gamma_{i,k+1}}\right)\cdots \left(\sum_{n}\prod_{i=1}^{k} p\left(\mathbf{x}_n,\mathbf{x}_i\right)^{\gamma_{in}}\right)\left(\prod_{(i,j)\in\mathcal{R}_k} p\left(\mathbf{x}_i,\mathbf{x}_j\right)^{\gamma_{ij}}\right) \quad . \tag{2.32}$$

The previous expression is fairly easy to obtain, as the depth of the dual tree corresponding to the k-fan (or its junction tree) is equal to 2. The conditional probability of a landmark knowing the reference points is:

$$p\left(\mathbf{x}_i|\mathbf{x}_1,\ldots,\mathbf{x}_k\right) \propto \prod_{j=1}^{k} p\left(\mathbf{x}_i,\mathbf{x}_j\right)^{\gamma_{ij}} \quad . \tag{2.33}$$

Sampling on a grid of size m runs in $O(nm^{k+1})$ polynomial time, and requires $O(nm^k)$ memory space. This is why in practice it is unfeasible to perform such a sampling. We resort to another method to assess the ability of our method to capture the shape information, by applying a priori-constrained perturbations to a randomly located set of random points. Let $(\mathbf{p}_1,\ldots,\mathbf{p}_n)$ be a set of random points in $\mathbb{R}^N$. We propose to verify the information encoded by our shape prior through constrained displacements of this set of points. Formally, let $\mathcal{D}_i = \{\mathbf{d}_{i1},\ldots,\mathbf{d}_{im}\}$, $1 \leq i \leq n$ be a finite set of candidate displacements, and let $\mathcal{L}_i = \{l_{i1},\ldots,l_{im}\}$ be a set of associated labels. Assigning a label l_{ij} to a point $\mathbf{p}_i$ is equivalent to displacing the latter by $\mathbf{d}_{ij}$. The best perturbation



(a) (b) (c)

Figure 2.7: Corpus callosum model construction. (a) The considered training set. (b) The projection of the control points on the first 3 shape map dimensions. The density of the landmarks is estimated by a Gaussian kernel and is color-coded. (c) Control points clustered in 13 clusters: centers are represented by squares.

according to the shape model refers to the displacements that minimize:

$$-\sum_{(i,j)\in\mathcal{E}_k}\gamma_{ij}\log\left(p\left(\frac{\|\mathbf{x}_i+\mathbf{d}_{ii'}-\mathbf{x}_j-\mathbf{d}_{jj'}\|}{\widetilde{d}}\right)\right)\quad. \tag{2.34}$$

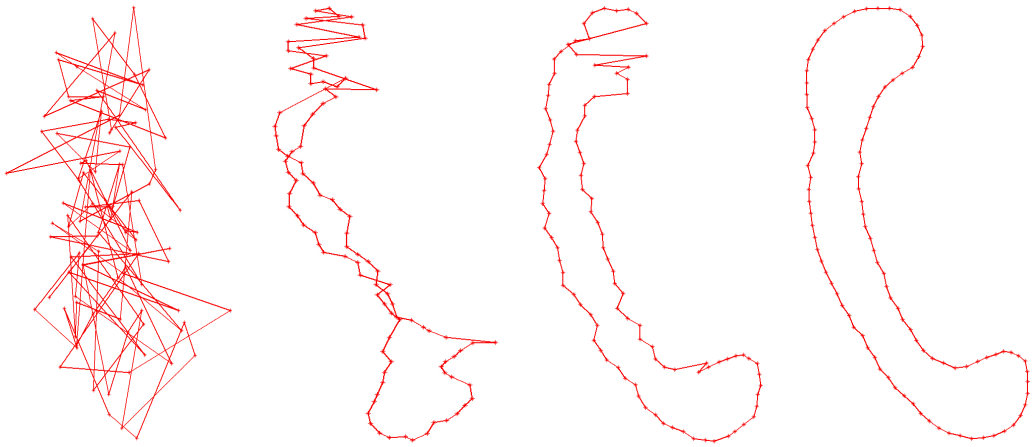
This quadratic assignment problem $\min_{i',j'}\sum_{(i,j)\in\mathcal{E}_k}V_{ij}(l_{ii'},l_{jj'})$ is NP-hard in the general case, and can be solved approximately using [Komodakis *et al.* 2008b] or [Kolmogorov 2006]. An implementation of these methods is available online[3,4].

The minimization problem (2.34) can be solved iteratively. The solution of the iteration i initializes the iteration $i+1$. After a few iterations, the shape does not evolve any more.

In practice, applying shape prior constraints to an initial set of random control points leads to a "mean" instance of the learned object. The use of such prior in the segmentation framework is explored in the following.

[3] http://www.csd.uoc.gr/~komod/FastPD/

[4] http://vision.middlebury.edu/MRF/code/



(a) (b) (c) (d)

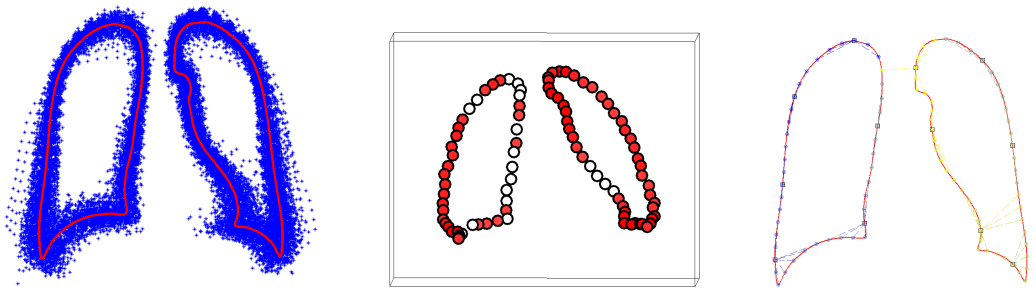
Figure 2.8: Corpus callosum shape prior applied to random points.(a)–(d) Deformation of a random point cloud according to the shape prior term.

Applying the Shape Prior to a Set of Random Points

We applied our shape model in 2D and in 3D, in the medical and non-medical settings. The obtained results are shown in the following.

Hands dataset Fig. 2.5(a) shows the hands annotated database provided by [Stegmann & Gomez 2002]. It consists of 40 labeled hand contours with 56 control points. Correspondences of the landmarks are guaranteed by manual labeling. Fig 2.5(b) displays the shape maps embedding of the data, and the unsupervised clustering output is depicted in Fig. 2.5(c). The results are intuitively satisfactory as the rigidly moving sub-components of the object (the fingers) are separated in coherent groups. A prior-constrained perturbation of randomly located 56 landmarks leads to a hand shape as shown in Fig. 2.6.

Corpus Callosum dataset The corpus callosum dataset that we used contains 90 2D MRI images manually annotated. Each corpus callosum contour consists of 100 points. An illustration of the data is shown in Fig. 2.7(a). We used all these landmarks to train our shape model, and we obtained the shape maps embedding in Fig. 2.7(b). Next, we performed the unsupervised LP-based clustering method that provided 13 clusters as shown in Fig 2.7(c).

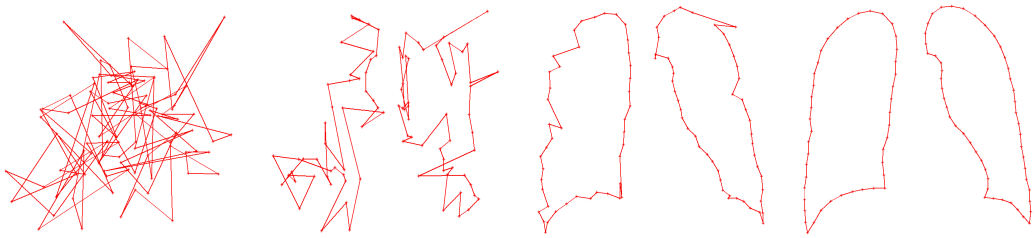


(a) (b) (c)

Figure 2.9: Joint lungs model. (a) The considered training set. (b) The projection of the control points on the first 3 shape map dimensions. The density of the landmarks is estimated by a Gaussian kernel and is color-coded. (c) Lungs landmarks are clustered in 12 clusters: centers are represented by squares.

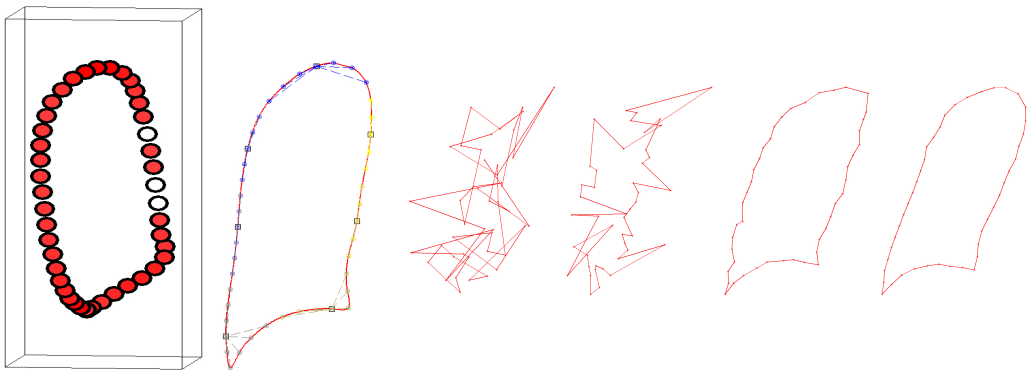
The last step of the validation consisted of applying the shape prior to a set a randomly located points. We used Gaussian distributions to models the pairwise probability functions. As expected, the shape prior guides the landmarks towards a corpus callosum shape. Some iterations of this contour evolution are presented in Fig. 2.8.

Right Lung and Left Lung dataset We applied our method to the shape modeling of the right and left lungs from 2D radiographic (X-ray) images. We used a publicly-available dataset



(a) (b) (c) (d)

Figure 2.10: Joint lungs shape prior applied to random points.(a)–(d) Deformation of a random point cloud according to lungs the shape prior term.



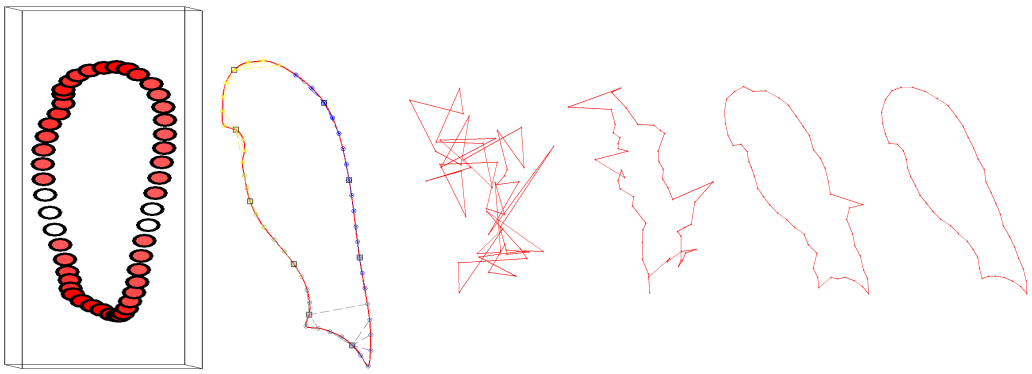
(a) (b)

Figure 2.11: A separate model of the right lung. (a) Shape maps embedding (left) and unsupervised clustering (right). We obtain 7 clusters. (b) Applying the learned shape model to random points.

[van Ginneken *et al.* 2006][Shiraishi *et al.* 2000][5] of 247 images of healthy and non healthy subjects (presenting nodules). The database contains gold standard segmentations from radiologists, that provided a delineation of the organ. Gold standard segmentation masks are hence available as well as corresponding landmark positions lying on the contour. We used the 44 available landmarks to learn the NCL shape model. Fig 2.9(b) shows the obtained shape maps embedding. The clustering leads to the result presented in Fig. 2.9(c) which shows the 13 automatically determined clusters for the right lung. The clustering was used to define the 13-fan structure of the graph, which contained 481 edges (compared to the 946 edges of the corresponding complete graph - almost one edge out of two was pruned). The corresponding prior-constrained deformation validation is presented in Fig. 2.10. We also modeled separately the right lung (Fig. 2.11) and the left lung (Fig. 2.12).

Left Ventricle dataset The Computed Tomography (CT) cardiac dataset is courtesy of *Siemens Corporate Research* and was used by [Grady *et al.* 2005] to illustrate their interactive graph-based segmentation methods. It consists of 25 3D image volumes, with varying voxel size, approximately equal to $0.2 \times 0.2 \times 0.5$ mm^3. These data were segmented using an interactive tech-

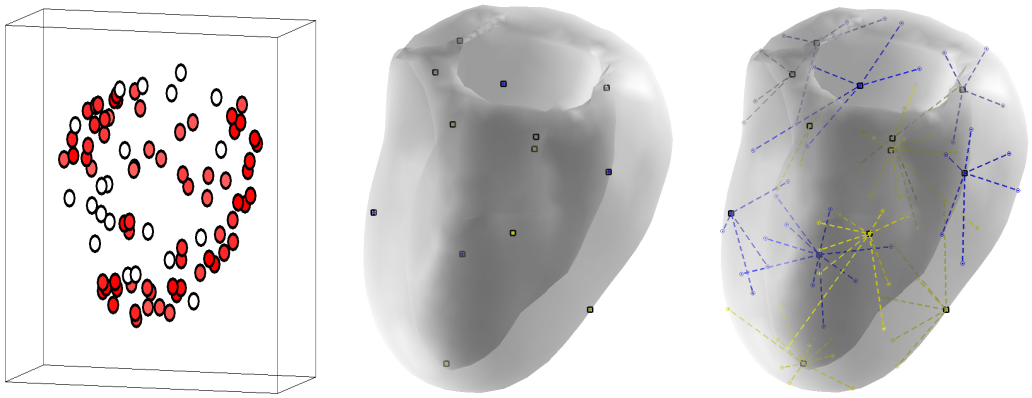
[5] http://www.isi.uu.nl/Research/Databases/SCR/



(a) (b)

Figure 2.12: A separate model of the left lung. (a) Shape maps embedding (left) and unsupervised clustering (right). We obtain 9 clusters. (b) Applying the learned shape model to random points.

nique [Grady 2006]. One particular volume was identified as a "mean" volume, and all the remaining ones were registered to it using [Taron *et al.* 2009]. Precisely, the smoothed segmentations masks of the left ventricle (LV) were used to generate signed distance maps (SDM) that were deformed using continuous free form deformations (FFD). The mean volume LV segmentation mask was



(a) (b) (c)

Figure 2.13: Modeling the left ventricle. (a) The obtained shape maps embedding. (b) The clusters centers are represented on the left ventricle surface. (c) The clustering result.

then manually labeled, and 90 control points were placed on the myocardium surface as well as the atrium surface. Afterwards, the transformations provided by the registration step were applied to estimate the correspondences of the control points, and generate the training dataset. The control points are associated with a meshed surface, that is interpolated using thin plate splines (TPS). The shape maps method gives an embedding of the landmarks that is represented in Fig. 2.13(a). The clustering step ends to 13 clusters as depicted in Fig. 2.13(c). The prior-constrained perturbation validation of the LV model is shown in Fig. 2.14.

In terms of computational complexity, we executed the "perturbation moves" algorithm 50 successive times using a 3 GHz dual-core processor. We report the running time, using respectively the complete graph and the k-fan graph, for the hand example and the left ventricle example in the Table 2.1.

Table 2.1: Running time comparison between the complete graph and the k-fan graph.

	Hand example	LV example
Complete Graph runtime (s)	34($\pm$4)	51($\pm$5)
k-fan Graph runtime (s)	11($\pm$1)	26($\pm$2)
Number of Labels	361	105
# edges Complete Graph	1540	4005
# edges k-fan Graph	459	1079

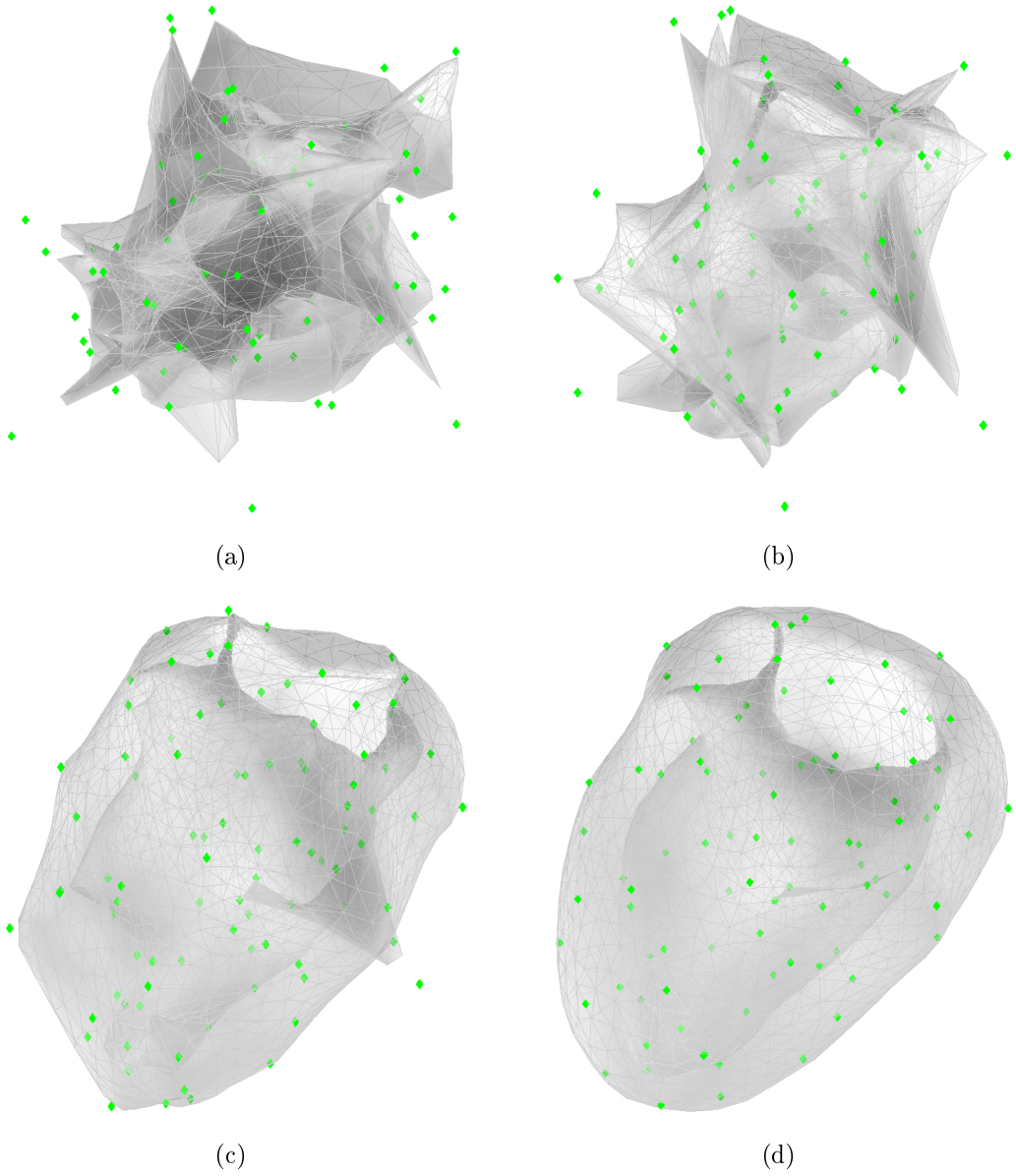


Figure 2.14: Left ventricle shape prior applied to random points.(a)–(d) The mesh is updated through thin plate spline (TPS) interpolation of the control point (in green) displacements. These displacements are guided by the shape prior.

Chapter 3

Knowledge-based Segmentation

Contents

3.1 Introduction

The use of prior models is very popular in image segmentation. The most common approach consists of associating the model with the observations and then seeking for the optimal model parameters, that is a compromise between the observed image and the prior constraints. In this section, we will exploit the normalized chord length shape model towards model-based segmentation. The underlying graph structure is determined from the data, and the shape model parameters are learned from the training shape population as well. Given such a model, one has first to introduce data support and then efficient optimization techniques for the inference of the shape parameters, given an unseen image (an image that does not belong to the training dataset). Our motivation is to propose a flexible framework that allows the use of different image cues, and permits efficient optimization thanks to the recent developments in the field.

This chapter is structured as follows: we review in section 3.1.1 the optimization techniques related to pairwise Markov random fields (MRFs). In section 3.1.2, we detail different segmentation techniques, and describe specifically the graph-based methods. We will particularly discuss the approaches that combine MRFs with shape priors, showing that it is is a tedious procedure in general, due to the fact that graph-based methods assume weak connectivity while shape priors often require global knowledge of the shape parameters. In an attempt to incorporate a statistical shape prior model within the MRF framework, we propose in section 3.2 a segmentation method that combines our learned prior knowledge with a region-based data term, through the use of Voronoi diagrams. The resulting algorithm features an object contour (or surface) that evolves towards the solution, in the spirit of deformable contours and surfaces. Experiments of this method are shown in a 2D computer vision application, and in a 3D medical image analysis setting. In order to illustrate the flexibility of our framework, we detail in section 3.3 a slightly different approach that shares ideas with graph matching and point pattern matching. In this algorithm, the data term is related to the responses of classifiers based on distinctive image feature vectors. In both cases, we show that our statistical shape model guarantees regular and robust segmentation.

3.1.1 Discrete MRFs and Optimization

MRFs were introduced to the computer vision domain by [Geman & Geman 1984]. Since their seminal work, MRFs have been successfully applied to the modeling of a wide range of vision applications, such as image restoration, image completion, segmentation, stereo vision, tracking and motion estimation, registration, etc [Boykov & Funka-Lea 2006]. These applications often seek the estimation of the model variables with respect to some observations. This is often formulated as a maximum a posteriori (MAP) estimation problem, and under mild assumptions, it leads to minimizing the following MRF energy:

$$E(\mathbf{l}) = \sum_{p \in \mathcal{V}} V_p(l_p) + \sum_{(p,q) \in \mathcal{E}} V_{pq}(l_p, l_q) \ , \tag{3.1}$$

where $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is the considered MRF, $\mathcal{V}$ is the set of graph variables, $\mathcal{E}$ is a local neighborhood system associated with this graph that can be spatially varying, and $l_p \in \mathcal{L}$ is a possible state of the variable $p \in \mathcal{V}$. This problem is also interpreted as a labeling problem in the discrete case, where the unary potentials $V_p(l_p)$ indicate/measure the quality of the labeling with respect to the observations for a given node, and the pairwise potentials $V_{pq}(l_p, l_q)$ aim to establish coherence between the labeling of connected nodes. Minimizing such function is a challenge in the general case, as this problem is proven to be NP-hard. However, numerous optimization techniques were developed to tackle this problem from different points of view. For instance, some methods guarantee an exact optimal solution of the problem (3.1) for a very restricted set of cases. Others provide nearly optimal solutions, with predefined confidence, if some constraints are satisfied. We present in the following an overview of these methods.

Early Methods

Iterated conditional modes (ICM), simulated annealing (SA), highest confidence first (HCF) and mean field annealing (MFA) are among the first algorithms considered to solve the aforementioned MRF minimization problem. ICM [Besag 1986] is a greedy approach that iteratively updates the labels of the MRF nodes. At each iteration, a node is allowed to change its label to the

one that provides the local largest decrease of the objective function assuming the remaining of the labels constant from the previous iteration. These updates are repeated for all nodes until convergence to a local minimum, i.e. until no label update gives any decrease of the energy. This method is sensitive to initialization. HCF [Chou & Brown 1990] is a deterministic algorithm that attempts to improve the ICM by updating the nodes in a certain order. It introduces an additional label to describe an "uncommitted" state. At the beginning of the algorithm, all nodes are "uncommitted". The algorithm defines the so-called stability for each node, and at each iteration, only the node with the smallest stability is allowed to commit, or update its label. Once a node is committed to a label, the potentials of the connected nodes are also updated and decisions for them can be taken with more confidence. SA [Kirkpatrick *et al.* 1983][Černý 1985] is also an iterative algorithm, where a temperature parameter T plays an important role (in analogy with the slow cooling/annealing phenomenon in metallurgy). Iteratively, the SA heuristic choses whether to stay in the current state or to move to a neighboring one in a probabilistic manner. If moving to a new state reduces the energy, then the move is accepted. Otherwise, the move is accepted with a certain probability that depends on the temperature T, and decreases over time, i.e. as the temperature tends to zero. Intuitively, for large values of T, the system is allowed to move with certain freedom, such that it is able to escape local minima. This freedom decreases when T is close to zero, and the probabilistic moves become almost deterministic. Asymptotically, SA finds the global optimum, which in practice would require an infinite time. Note also that SA implementations converge to a local minimum. Moreover, [Greig *et al.* 1989] gave binary labeling examples where the solution provided by SA is far from the optimum. MFA [Bilbro *et al.* 1989] is an algorithm that is similar in spirit to SA. It proceeds however by estimating the mean field of each node at decreasing temperatures, rather than with a sampling scheme and probabilistic moves. The mean field of the MRF converges to the MAP when T tends to zero. This popular minimization algorithm resorts in practice to approximations as it needs to estimate the partition function, which is computationally intractable. This is another instance of algorithms with asymptotic optimal properties, which converge however to local minima when used in real-world

applications. We turn now to present another class of methods that belongs to combinatorial optimization techniques and exhibits efficiency properties as well as the ability to achieve a global optimum under certain conditions.

Graph-Cut - Binary Case

[Greig *et al.* 1989] were the first to use a s/t graph cut algorithm in the context of MAP estimation of a MRF in computer vision. They successfully applied it to the restoration of binary images and showed that it outperforms SA. A decade later, the rediscovery of graph cut techniques in the computer vision community started a new research trend. This was due to the seminal work of [Boykov *et al.* 1999] where an efficient implementation of the max-flow/min-cut problem was introduced in the context of vision. These ideas were popularized in [Boykov & Jolly 2001] where the use of the graph cut technique for interactive n-dimensional image segmentation was introduced. In this subsection, we start first by describing the basics of graph cut algorithms. In order to simplify the presentation, let us first consider a simple case where the label set consists of two labels.

The binary labeling problem is represented by a directed graph where each node represents a variable, and where edges connect neighboring nodes. These nodes are called internal nodes, and the corresponding edges are called internal links or n-links. In addition, the graph contains two terminal nodes called the source s and the sink t. These two special nodes correspond to the two classes of the binary labeling. The graph contains also terminal edges, or t-links, that connect each internal node with both terminal nodes. Each edge is associated with a cost, or a capacity. The cost of a n-link defines the penalty of assigning the corresponding two nodes to different classes, whereas the cost of a t-link represents the penalty of assigning the node to the corresponding terminal. For the case of the MAP estimate of a binary MRF, defined by (3.1), the n-link costs correspond to $V_{pq}(l_p, l_q)$ and the t-links to $V_p(l_p)$, where $\mathcal{L} = \{s, t\}$.

An s/t cut on such a graph partitions the nodes in two disjoint classes $\mathcal{S}$ and $\mathcal{T}$ such that $\mathcal{S}$ contains the source terminal s and $\mathcal{T}$ contains the sink terminal t. The cost of such a cut is equal to the sum of the costs of directed edges (p, q) such that $p \in \mathcal{S}$ and $q \in \mathcal{T}$. Hence, a MAP binary labeling

is equivalent to a cut with a minimum cost, or a min-cut. The theorem of [Ford & Fulkerson 1962] proves that the min-cut problem can be equivalently solved by finding the maximum flow from the source s to the sink t, through the capacitated edges. It also states that this max-flow saturates a set of edges that correspond to a min-cut. The max-flow and min-cut values are equal. Different low (polynomial) complexity algorithms for solving the min-cut/max-flow problem exist, and can be classified in two major classes: "push-relabel" methods [Goldberg & Tarjan 1988] and "augmenting paths" methods [Ford & Fulkerson 1962]. [Boykov & Kolmogorov 2004] give a description of these methods, and compare their performance on a set of computer vision applications. They also present a modified "augmenting paths" algorithm that produces the best results in their experiments. As far as the quality of the obtained solution is concerned, [Kolmogorov & Zabih 2004] discuss the set of energy functions that can be minimized using graph cuts. Hence, a global optimum is only guaranteed when the energy function (3.1) is submodular, meaning that $V_{pq}(0,0) + V_{pq}(1,1) \leq V_{pq}(0,1) + V_{pq}(1,0),\ \forall (p,q) \in \mathcal{E}$.

For non-submodular functions, the problem remains NP-hard in general. However, [Kolmogorov & Rother 2007] review a method called quadratic pseudo-Boolean optimization (QPBO) [Hammer *et al.* 1984] that provides partial optimal solutions in the arbitrary case. Essentially, the algorithm builds an appropriate s/t graph, where each variable $p \in \mathcal{V}$ is represented by two nodes, z and $\bar{z}$. Ideally, the label of $\bar{z}$ should be the negation of the labeling of z, i.e. $l_{\bar{z}} = 1 - l_z$. The new energy corresponding to the constructed graph is submodular, and is hence minimized using min-cut/max-flow. A consistent labeling of z and $\bar{z}$ (i.e. $z \in \mathcal{S}$ and $\bar{z} \in \mathcal{T}$, or the converse) provides a labeling for p. The variables corresponding to an inconsistent labeling of z and $\bar{z}$ remain unlabeled. This algorithm provides hence a partial solution, but with the guarantee that the labeled variables belong to the optimal solution.

One can put the importance of graph cut methods into perspective as they are limited to the binary case, and come with restrictions on the energy function. However, as we shall see in the following, the binary case is closely related to the multi-label setting.

Graph-Cut - Multi-label Case

Several works in the late 90's showed the high relevance of s/t graph cut with respect to the multi-label problem. [Roy & Cox 1998] were the first to use an s/t graph cut for the general labeling case. Building on their work, and using appropriate graph constructions, [Boykov *et al.* 1998] proposed a method to exactly minimize the problem (3.1) when the pairwise potentials are linear functions of the labels. [Ishikawa 2003] shows subsequently that convex pairwise interactions can be handled as well. [Boykov *et al.* 2001] further generalize the result to the case where V_{pq} is a metric, showing that a good approximation of the optimal solution can be found within guaranteed bounds. The algorithms that they proposed, α-expansion and $\alpha-\beta$-swap are efficient and perform well in practice, even in the non-metric case [Rother *et al.* 2005] (yet losing optimality guarantees). We give now some details about these algorithms.

The proposed expansion move and swap move algorithms in [Boykov *et al.* 2001] try to iteratively improve the current labeling. In an α-expansion move, any node is only allowed to change its label to α. In an $\alpha-\beta$ swap, only nodes with labels α (respectively β) are allowed to change their label to β (respectively α), and β (respectively α) only. Both algorithms find in each iteration the move that minimizes (locally) the energy. The number of these moves being exponential, the authors translate this minimization problem into a min-cut / max-flow problem in an appropriately constructed graph. Hence, during each of these iterations, an s/t graph cut is performed. A complete set of iterations over all the labels $\mathcal{L}$ is called a cycle. The algorithms are guaranteed to converge in a finite number of cycles to a local minimum. For α-expansion, this local minimum is within a known factor of the global optimum.

Loopy Belief Propagation

Belief propagation (BP) [Pearl 1988] is an iterative algorithm that propagates local messages between the nodes of the MRF. We present here an intuitive description of its mechanism. At each iteration, a node sends messages to its neighbors, and receives messages from them. The algorithm keeps repeating

these iterations until a fixed point is reached, i.e. all the messages stabilize.

In each iteration, a node p sends a message $\{m_{pq}(l_q)\}_{l_q \in \mathcal{L}}$ to the node q. This message expresses how likely the node q should be assigned a label l_q according to the node p. This message considers two aspects. First, using the interaction term $V_{pq}(l_p, l_q)$, the node p examines the likelihood of a label l_q for the node q if it gets the label l_p. Then, it assesses the likelihood of getting a label l_p, based on the evidence $V_p(l_p)$ and the messages $m_{rp}(l_p)_{r \neq q}$ it receives from its neighbors (other than q) about the likelihood of l_p.

After exchanging such messages sufficiently, the messages converge. Then a belief $b_p(l_p)$ of assigning a label l_p to a node p can be computed, as a combination of the evidence $V_p(l_p)$, and the messages $m_{rp}(l_p)$ that p received from its neighbors. One can distinguish between two kinds of BP algorithms: the sum-product algorithm that is used to estimate marginals, and the max-product algorithm that gives MAP estimates. In both cases, BP is exact on acyclic graphs, or trees, and only needs to pass messages in a forward way (from the leaves to the root) and then in a backward way to converge. However, when the graph contains loops, there are no such guarantees, and the algorithm may not converge [Yedidia *et al.* 2003] [Kschischang *et al.* 2001]. In practice, loopy belief propagation (LBP) has successfully been applied even with cyclic graphs.

Linear Programming Relaxation

The linear programming (LP) relaxation of the MAP-MRF problem (3.1) is a prominent approach to find approximate solutions. The latter can be cast to an equivalent integer program:

$$\min \sum_{p \in \mathcal{V}} \left(\sum_{l \in \mathcal{L}} V_p(l) x_p(l) \right) + \sum_{(p,q) \in \mathcal{E}} \left(\sum_{l,l' \in \mathcal{L}} V_{pq}(l,l') x_{pq}(l,l') \right) \quad , \tag{3.2}$$

such that:

$$\sum_{l} x_p(l) = 1 \qquad \forall p \in \mathcal{V} \tag{3.3}$$

$$\sum_{l} x_{pq}(l, l') = x_q(l') \qquad \forall l' \in \mathcal{L},\ \forall (p,q) \in \mathcal{E} \tag{3.4}$$

$$\sum_{l'} x_{pq}(l, l') = x_p(l) \qquad \forall l \in \mathcal{L},\ \forall (p,q) \in \mathcal{E} \tag{3.5}$$

$$x_p(.),\ x_{pq}(.,.) \in \{0, 1\}\ . \tag{3.6}$$

The binary variables $x_p(.)$ and $x_{pq}(.,.)$ that are introduced in the above functions act as indicator functions, and specify which label is assigned to a node or a pair of nodes. Hence, $x_p(l) = 1$ is equivalent to assigning label l to node p, and $x_{pq}(l, l') = 1$ is equivalent to assigning label l to node p and label l' to node q. The constraint in equation (3.3) guarantees that each node is assigned one and only one label, and constraints (3.4) and (3.5) ensure consistency between the unary and pairwise binary variables.

The LP relaxation consists in relaxing the integer constraints (3.6) to the constraints $x_p(.) \geq 0$ and $x_{pq}(.,.) \geq 0$. This LP relaxation was first introduced by [Schlesinger 1976], as pointed out in [Werner 2007]. The dual program that corresponds to the (primal) linear program can be written as:

$$\max \sum_{p \in \mathcal{V}} y_p\ , \tag{3.7}$$

such that:

$$y_p \leq \min_{l \in \mathcal{L}} h_p(l) \qquad \forall p \in \mathcal{V} \tag{3.8}$$

$$y_{pq}(l) + y_{qp}(l') \leq V_{pq}(l, l') \qquad \forall (l, l') \in \mathcal{L},\ (p,q) \in \mathcal{E}\ , \tag{3.9}$$

with $h_p(.) = V_p(.) + \sum_{\substack{q \\ (p,q) \in \mathcal{E}}} y_{pq}(.)$. The quantity $h_p(.)$ is called height variable. Optimizing the obtained linear program using conventional techniques such as the simplex algorithm or interior-point methods is often not practical in the context of computer vision problems because of their very large scale. For this reason, the above-mentioned methods would be very slow. This motivated the development of approximate but efficient LP optimization techniques. These methods either use the primal-dual paradigm, or proceed in the dual domain to maximize a lower bound of the energy. We give in the following a description of the most prominent approaches.

Primal-Dual approaches and Fast-PD [Komodakis & Tziritas 2007] apply the primal-dual principle to solve approximately the linear program (i.e. the relaxed version of (3.6)), while providing suboptimality bounds. Their algorithm requires that the pairwise costs are positive (which is less restrictive than the metricity requirement of α-expansion for instance). By weak duality, the optimal cost of the primal-dual problem lies between the costs of any pair of primal and dual feasible variables. Therefore, the primal-dual principle states that controlling their cost gap by a factor f provides an f-approximation to the optimal integral solution. Moreover, feasible primal dual solutions that satisfy the relaxed complementary slackness condition provide an f-approximation to the optimal integer solution. Subsequently, the primal-dual schema can be applied to find a solution that is an f-approximation to the optimum. This schema is devised to generate pairs of integral primal and dual solutions until the last pair forms a primal and a dual feasible solutions that satisfy the relaxed primal complementary slackness conditions. [Komodakis & Tziritas 2007] propose an algorithm that follows the primal-dual schema. They maintain during optimization an active primal labeling $x_p(.)$, and dual variables $y_{pq}(.,.)$, such that the complementary slackness conditions are satisfied. To this end, they iteratively update the primal and dual variables to meet in particular the height complementary condition $h_p(x_p) = \min_{l \in \mathcal{L}} h_p(l)$. The update of the dual variables takes place in a selective manner. In each c-iteration, only c-label variables $y_{pq}(c)$ are changed, to satisfy the constraints as much as possible, which in turn would produce new $h_p(c)$ height values. The optimal update of the c-variables is found by solving a min-cut max-flow problem in an appropriately constructed capacitated graph. At this point, there might still be nodes violating the height complementary condition with respect to the label c. This is taken care of by updating accordingly the primal variables corresponding to those nodes. After each update, the dual cost is guaranteed to decrease. The algorithm converges to a pair of primal-dual solutions that satisfy the complementary slackness conditions (when no update is done after $|\mathcal{L}|$ c-iterations), and hence provide an f-approximation to the solution. The value of f can be estimated from the pairwise costs, which are (only) assumed to be positive, as $f = \max_{pq} \frac{\max_{l \neq l'} V_{pq}(l,l')}{\min_{l \neq l'} V_{pq}(l,l')}$. Besides the theoretical guarantees on the approxima-

tion quality, per-instance suboptimality bounds are provided for each application. These bounds proved to be in practice much tighter than the theoretical ones. It also turns out that this Primal-Dual method is a generalization of α-expansion.

[Komodakis *et al.* 2008b] further extend their method to obtain more efficiency and provide a very fast primal dual algorithm, called FastPD. They essentially build upon their work in [Komodakis & Tziritas 2007], by additionally making the link between the primal-dual gap and the number of path augmentations during the min-cut/max-flow optimization. Indeed, the speed of the algorithm depends heavily on the efficiency of solving these max-flow problems, which in turn is related to the number of augmenting paths per max flow. They show that the primal-dual gap is an approximate upper bound on the number of such paths. A modification of their previous algorithm ensures that the primal-dual gap always decreases (and is never destroyed, even temporarily). This first adjustment leads to a clear decrease in the number of augmentation paths and hence to a substantial speed up of the algorithm, while maintaining it optimality properties. Moreover, they design a max-flow algorithm that is adapted to their graph (that has few s-links) and provide an incremental graph construction method which yields an additional acceleration. In practice, FastPD performed 3 to 9 times faster than α-expansion.

Message-Passing Approaches Tree-reweighted (TRW) methods are another class of approaches that aim to approximately solve the MAP-MRF problem. While relying on the passing of messages similarly to BP, these methods are closely related to LP. Indeed, they attempt to solve the LP relaxation of the MRF, by maximizing the dual of the LP. [Wainwright *et al.* 2005] introduced the tree-reweighted message passing algorithm (TRW-MP). They show that by combining concave subproblems, they can provide a concave lower bound on the MRF energy. These subproblems can be chosen as trees. They show that the message passing algorithm leads to a tight bound if and only if the collection of trees shares a common optimum. They also prove that the maximization of this lower bound (that is independent from the choice of the trees) is equivalent to optimizing the dual of the LP relaxation of the MAP MRF problem. However, TRW-MP does not have any

guarantee of convergence, or increase of the lower bound. [Kolmogorov 2006] proposes a sequential version of the tree-reweighted algorithm, TRW-S, that has convergence guarantees. However, the author shows that the algorithm may converge to a point that is not the optimum of the LP lower bound. The optimal properties are only shown for the binary submodular case [Kolmogorov & Wainwright 2005].

[Komodakis *et al.* 2007] introduced a new message-passing method based on dual decomposition (DD) [Bertsekas 1999], in order to address the above-mentioned limitations. More explicitly, they aimed to solve the dual LP problem, i.e. maximize the lower bound. The dual decomposition technique consists of decomposing the Lagrangian dual of the intractable original MAP MRF problem, into less complex subproblems, thanks to a Lagrangian relaxation. [Komodakis *et al.* 2007] consider tree subproblems that can be easily solved. Then, the optimization (using the projected subgradient method) amounts to an iterative "message passing" between the "master" problem and the "slave" problems. Iteratively, each easy subproblem is solved according to the updated MRF potentials, and the result is sent to the master problem that updates the potentials according to the new results. Moreover, the authors show that the Lagrangian relaxation problem is equivalent to the LP relaxation of the MAP MRF problem. Thus, they establish the link between their method, the TRW methods and LP. Unlike TRW approaches, this DD-based algorithm is guaranteed to maximize the lower bound of the LP relaxation.

Tightness of the Relaxation The above-mentioned class of methods try to optimize the dual of the LP relaxation. Interestingly, this dual energy is a lower bound on the MAP MRF energy. Hence, one can hope that by solving the LP, a good approximation to the original integer program can be obtained. This observation raises the question of the tightness of the relaxation. Indeed, a loose relaxation would lead to a poor approximation. Researchers have tried other types of relaxations, besides the LP one. For instance, a quadratic programming (QP) relaxation is used by [Ravikumar & Lafferty 2006], and a second order cone programming (SOCP) relaxation was proposed by [Muramatsu & Suzuki 2003] for the binary case and extended to the general case by [Kumar *et al.* 2006]. [Kumar *et al.* 2009]

show that the QP relaxation and the SOCP relaxation are equivalent in terms of tightness. Furthermore, they show that they do not compete favorably with the LP relaxation, meaning that the latter leads to better approximations of the MAP-MRF problem.

[Komodakis & Paragios 2008] present a work that deals with the relaxation tightness. They act on the lower bound of the MRF energy by defining tighter LP-relaxations, through the use of a new set of pairwise potentials that satisfy certain constraints, called virtual potentials. They give a characterization of dual LP energies that are not tight, as the existence of the so-called inconsistent node cycles with respect to the labeling. They propose an iterative algorithm to repair these inconsistent cycles, by successively adjusting the virtual potential to appropriate values, and applying the augmenting DAG algorithm [Werner 2007] to impose necessary optimality conditions. When using this algorithm, the lower bound on the MRF energy does not decrease. These iterations are repeated until there are no more inconsistent cycles. This can be thought of as a series of relaxations of gradually increasing tightness. In practice, this algorithm leads to better optima. It especially performed well in difficult MRF problems, with respect to state-of-the art techniques.

[Szeliski *et al.* 2008] provide a comparison between different MRF energy minimization methods, using benchmark problems. The algorithms that are compared are ICM, max-product LBP, a LBP implementation derived from TRW-S, α-β-swap, α-expansion and TRW-S.

3.1.2 Image Segmentation Techniques

We briefly sketch the most prominent image segmentation techniques in the literature. We note that in addition to the methods that we will detail in the following, there is another class of segmentation techniques that is mainly applied in medical image analysis, namely atlas-based segmentation. Often described as segmentation by registration, their principle consists of registering a reference image, which is called atlas, to a query image. A ground truth segmentation is associated with the reference image, and a segmentation of the query image is obtained by a warping operation, based on the resulting deformation field. A more sophisticated approach that goes beyond the use of simple single-image atlas considers the multi-atlas segmentation problem.

Typically, every atlas is fitted to the query image, and a final segmentation is obtained in a multi-classifier labeling manner by voting for instance. To capture the shape changes due to anatomical and morphological variations, more advanced atlas construction techniques were developed, and involved incorporating probabilistic information. Research about atlas-based segmentation has mainly focused on registration algorithms and atlas construction. A review of these different methods is beyond the scope of this work. However, the reader can be referred to a detailed description in [Rohlfing *et al.* 2005].

We give now an overview of deformable models and level sets for segmentation, active shape and active appearance models, and eventually graph-based segmentation approaches. We end this last part by discussing the endeavors that researchers made to introduce shape priors in the MRF framework, and commenting the approaches that are particularly related to our work.

Several detailed reviews on the topic of image segmentation techniques exist. [McInerney & Terzopoulos 1996], [Jain *et al.* 1998] and [Montagnat *et al.* 2001] describe the deformable contours methods. [Cremers *et al.* 2007] survey level set approaches, and in particular the statistical ones. A discussion about the developments of the active shape model (ASM) and active appearance model (AAM) algorithms can be found in [Heimann & Meinzer 2009]. [Boykov & Funka-Lea 2006] describe the history and developments of the graph cut method, and as previously indicated, [Szeliski *et al.* 2008] compare several discrete MRF optimization methods on a benchmark of computer vision applications.

Deformable Models and Level Sets

[Kass *et al.* 1988] introduced the active contours concept, or *snakes*, in image segmentation. These deformable models evolve towards the object to be segmented, and are guided by an energy that combines an internal term and and external one. The internal energy is based on the (parametrized) curve regularity and imposes smoothness constraints to the evolving contour, whereas the external energy stands for the image-driven forces that attract the curve to the image location with high gradient values. This 2D edge-based approach was then generalized to the 3D case in [Terzopoulos *et al.* 1988]. This method is sensitive to initialization, and can get trapped in local min-

ima. [Cohen 1991] introduces an additional term in the energy that represents "balloon" forces. This inflating (or deflating) term aims to avoid the local minima by helping the curve to pass them. Another criticism that is commonly formulated to active contours is related to their inability to allow for topology changes. [Delingette 1999] attempt to overcome this limitation by introducing the deformable simplex meshes. In such meshes, each vertex has exactly the same number of neighbors. This non-parametric surface representation is able to model objects of arbitrary topologies. The author proposes an algorithm that adapts the topology of an initial surface to the object to be extracted from the image. [Montagnat *et al.* 1999] use a 2-simplex mesh to segment 3D cylindrical echocardiographic images. This work is extended in [Montagnat *et al.* 2003] to handle 3D ultrasound image sequences. By using reparametrization tricks, [McInerney & Terzopoulos 1999] are also able to propose a topology-adaptable version of *snakes*. Alternatively, level sets [Osher & Sethian 1988] are an elegant framework that naturally handles topology changes. The work of [Malladi *et al.* 1995], [Kichenassamy *et al.* 1995] and [Caselles *et al.* 1997] benefit from these aspects, as they introduce the geodesic active contours, which can be viewed as geometric (parameter-free) counterparts of classical snakes.

Opposite to the previously-mentioned methods, other variational approaches rely on image region characteristics. [Mumford & Shah 1989] define a functional which allows to segment the image into homogeneous parts. They approximate the image by a piece-wise smooth function. The discontinuities of the approximating function are only allowed on a curve with its length being minimized within the functional. [Chan & Vese 2001] incorporate this approach in the level set framework, and design the so-called "active contours without edges". [Zhu & Yuille 1996] use statistical techniques of region growing by considering an MDL criterion. Their approach has also links to the snakes/balloon formulation. [Yezzi Jr. *et al.* 2002] implement this method in the level sets framework.

[Paragios & Deriche 2000] introduce the geodesic active regions Bayesian formulation that combines image region statistics with edge-based terms within the level set framework. The use of both region and edge information was considered earlier in [Chakraborty *et al.* 1996] without benefiting however

from the power of the level-set framework. They apply the Green's theorem to the region term, such that is can be expressed on the boundary of the evolving contour. Their approach features a shape model that uses the probability distributions of the Fourier coefficients representing the contour. However, their method depends on the initial conditions as it is optimized using a conjugate-gradient algorithm. The same idea of applying the divergence theorem was used by [Unal *et al.* 2004] where an active polygonal deformable model was defined based on image region information.

Continuous efforts have been made since then to improve these methods by developing knowledge-based variational formulations. Representing shape using a signed distance maps (SDMs) was a natural way of achieving this goal within the level set framework (section 2.1.1). Linear shape models (section 2.1.1) as well as their non-linear kernel-based counterparts (section 2.1.1) were also used in a variational framework.

Active Shape and Active Appearance Models

Active shape models (ASM) have become increasingly popular since their introduction by [Cootes & Taylor 1992]. This technique learns the shape of a class of objects using PCA (section 2.1.1). A shape $\mathbf{s}$ is hence approximated by its parameters vector $\mathbf{b}$ which is the orthogonal projection of the residual from the mean on the eigenbasis $\mathbf{E}$, or $\mathbf{b} = \mathbf{E}^T(\mathbf{s} - \bar{\mathbf{s}})$. An appearance model per landmark is associated to the shape model, and is used to enforce image-driven prior knowledge during search. This appearance model is related to the boundary of the object: intensity profiles, centered in the point location, and orthogonal to the contour (and hence capturing internal and external object features) are extracted from the training data. Mean profiles and covariance matrices are then computed, and the quality of a new profile is assessed by the Mahalanobis distance. Profiles of derivatives were also used in practice, and these different profiles were often normalized. ASM is an iterative algorithm that uses a local search method, due to the large size of the search space in general. At an iteration t, the evolving shape $\mathbf{s}_t$ in the image is expressed using its parameters $\mathbf{b}_t$ and a similarity transformation T, or $\mathbf{s}_t = T_t(\bar{\mathbf{s}} + \mathbf{E}\mathbf{b}_t)$. Then, for each landmark, the locally optimal candidate position along the normal to the contour is identified using the local appear-

ance model (intensity profile). The output of this step is a vector of optimal displacements $\mathbf{ds}_t$. The transformation T_t and then the shape parameters $\mathbf{b}_t$ are updated accordingly (while applying limit constraints on $\mathbf{b}_t$). Hence, a new valid instance of the shape $\mathbf{s}_{t+1}$ is obtained. These steps are repeated till convergence. The convergence depends on the initial conditions. Since ASM is a local search method, the initialization needs to be close enough to the solution. The robustness and the speed of the algorithm are improved through the use of coarse-to-fine search. For each level of the image corresponding pyramid, a different appearance model is built in the training stage.

Alternative ASM appearance models The flexibility of the ASM framework allows the use of different features, which paves the way to substantial improvements. For instance, Gabor wavelets were used in [Jiao *et al.* 2003], and the resulting feature distributions were modeled using Gaussian mixture models. [Langs *et al.* 2006] describe the object appearance using steerable features [Freeman & Adelson 1991]. [van Ginneken *et al.* 2002] suggest to use locally orderless image features [Koenderink & Van Doorn 1999]. In this method, a given image is fed to a bank of multiscale Gaussian derivative filters. Then, first statistical moments are extracted from local histograms in the filtered images (the histograms extents are related to the corresponding scale). These moments represent the considered features. [van Ginneken *et al.* 2002] used in practice all derivatives up to the second-order, five scales for Gaussian filtering, and the two first moments, and applied ASM to the segmentation of chest radiographs. An optimal set of features per landmark is then extracted using a forward-backward feature selection scheme.

Moreover, other non-linear methods apart from Gaussian mixtures were used to model the appearance features. [de Bruijne *et al.* 2003] propose a method based on k-nearest neighbors (kNN) to evaluate the probability of a given profile. During the training phase, profiles are sampled inside and outside the object in the vicinity of the landmarks positions. Hence the pool of profiles contains "true" profiles and "false" profiles, corresponding respectively to correct and wrong landmark positions. Then during the search phase, the kNN of a candidate profile assesses its quality by voting, and a probability value can be determined as the fraction of "true" votes. A similar approach

was adopted by [van Ginneken *et al.* 2002] to evaluate the goodness of candidate positions during search. The difference from [de Bruijne *et al.* 2003] lies in the use of a weighted kNN voting (the weights are inversely proportional to the distance to the landmark position), which establishes for a given point the probability of belonging to the inside or to the outside of the object. Then, the optimal position is defined by minimizing a cost function that combines such individual probabilities along profiles orthogonal to the contour and centered at the tested positions. The same rationale is applicable with different classifiers. Successful methods are presented in [Li *et al.* 2004a] and [Li & Ito 2005] and benefit from the strengths of the Adaboost algorithm [Freund & Schapire 1997].

Alternative ASM search schemes Other efforts to develop the ASM algorithm mainly considered the search part. [de Bruijne & Nielsen 2004] use a point distribution model (PDM) in a stochastic inference framework. They reformulate the segmentation task as a maximum likelihood (ML) problem, that is optimized using particle filtering [Isard & Blake 1998]. Iteratively, shape samples are drawn from the PDM. Then, particle weights are computed using a likelihood term (a kNN classifier is used for pixel probability estimation akin to [de Bruijne *et al.* 2003]). According to these weights, a random sampling of new particles is performed, followed by a small random perturbation of duplicated shapes. This process results in the multiplication of likely particles, whereas the bad shapes disappear gradually. Asymptotically, the algorithm converges to the maximum likelihood estimate. However, the optimal shape is approximated in practice by the strongest local mode of the particle distribution (that can be identified using the mean shift algorithm [Comaniciu & Meer 2002]) after a sufficient number of iterations. The use of a large number of initial samples bequeaths to the algorithm additional robustness against local minima and more independence with respect to the initial conditions, with the downside of increasing the computational cost. The authors extend their approach in [de Bruijne & Nielsen 2005] to the segmentation of multiple shape instances in the same image, by defining interaction constraints between neighboring shapes, and adjusting the data term. [Tu *et al.* 2004] and [Qu *et al.* 2008] are other examples of successful

combination of PDM with particle filtering approaches.

An alternative research direction deals with handling outlier candidates during the ASM search. In fact, the standard ASM search algorithm considers a least-square criterion with respect to the residual between the model and the data. The minimization of such criteria leads to optimal results only when the Gaussian noise assumption holds, which is not always true in practice. [Rogers & Graham 2002] tackle this problem by considering two options: M-estimators and random sampling techniques. M-estimators aim to alleviate the effect of outliers. This leads to a weighting strategy of the landmarks that depends on the size of their residual. The random sampling approach selects repeatedly small random subsets of landmarks, and uses them to estimate the shape parameters. Random sample consensus (RANSAC) [Fischler & Bolles 1981] is an algorithm that follows this rationale. It evaluates the obtained parameter estimation by the amount of its consistency with respect to the data, and returns as final estimation the parameters with the largest consensus set. [Abi Nahed *et al.* 2006] handle the outliers issue in a different way. They extract from the image to be segmented a set of candidate points for each landmark. Then, they use a robust point matching algorithm [Chui & Rangarajan 2000] to find the best correspondences between a legal shape instance and the pool of candidates, while allowing outlier rejection.

Incorporating MRF regularization in the search step is another method to tackle the independent estimation of the landmarks optimal displacements in the standard ASM. [Behiels *et al.* 1999] introduce a regularization term that constrains the relative displacements of consecutive landmarks on the object curve. The optimal landmarks are the one determined by minimizing a cost function that combines the aforementioned pairwise regularization terms, and individual landmark terms. A minimum cost path search algorithm (using dynamic programming, thanks to the tree structure of the open 2D contour) solves the problem. The method exhibits improvements compared to ASM, as far as outliers handling is concerned. More recently, [Tresadern *et al.* 2009] use a hybrid shape model that combines PDM and MRF. The unary MRF potentials correspond to a data term that describes the goodness of the candidate features, and the pairwise potentials account for the relationship between the relative displacements of pairs of landmarks. This regularization term is

given by a Mahalanobis distance, where the mean and covariance parameters are learned from the data. In both methods, the solutions are then regularized using the global model. Alternative developments of ASM are related to the shape modeling part and question the Gaussianity assumption (section 2.1.1) or seek a sparser representation (section 2.1.1).

Active Appearance Models AAM [Cootes *et al.* 2001] can be thought of as ASM with region-based image features. Besides modeling the shape variations, AAM encode the texture of the object and its variations. Hence, similarly to PDM that can synthesize new shapes, AAM are able to generate new images of the learned object. However, this generative model combines shape and appearance description into one linear system, such that the parameter vector controls both shape and appearance variations.

The first step consists in warping the training shape images to a common mean frame, using a triangulation algorithm. The intensity values are then stacked into texture vectors, which are used to perform PCA after intensity normalization. This leads to expressing the shape appearance as $\mathbf{g} = \overline{\mathbf{g}} + \mathbf{E}_g\mathbf{b}_g$, where $\overline{\mathbf{g}}$ is the mean normalized texture vector, $\mathbf{E}_g$ is the eigenbasis and $\mathbf{b}_g$ is the vector of texture parameters. The shape variations are expressed akin to the ASM setting as $\mathbf{s} = \overline{\mathbf{s}} + \mathbf{E}_s\mathbf{b}_s$. Then, the shape and texture (independent) parameters are combined in a common parameter vector $\mathbf{b} = \begin{pmatrix} \mathbf{W}_s\mathbf{b}_s \\ \mathbf{b}_g \end{pmatrix}$, where $\mathbf{W}_s$ is a diagonal matrix that weights each shape parameter, accounting for the units differences between the shape and texture models. Such a vector is generated for each training example, and then PCA is applied on the obtained pool of vectors. This yields the appearance model $\mathbf{b} = \mathbf{E}_c\mathbf{c}$, where the appearance parameter vector $\mathbf{c}$ controls both shape and texture variations:

$$\begin{aligned} \mathbf{s} &= \overline{\mathbf{s}} + \mathbf{E}_s\mathbf{W}_s^{-1}\mathbf{E}_{cs}\mathbf{c} \\ \mathbf{g} &= \overline{\mathbf{g}} + \mathbf{E}_g\mathbf{E}_{cg}\mathbf{c} \ , \end{aligned} \tag{3.10}$$

where $\mathbf{E}_c = \begin{pmatrix} \mathbf{E}_{cs} \\ \mathbf{E}_{cg} \end{pmatrix}$.

The AAM search algorithm is fairly different from its ASM counterpart. The AAM search corresponds to an image interpretation problem. It aims to

minimize the difference between the given image and a version synthesized by the appearance model. An instance of the model in the image is defined by a similarity transformation, an appearance parameter vector $\mathbf{c}$ and a texture transformation parameter vector. All these parameters are described in a unique parameter vector $\mathbf{p}$. Let $\mathbf{g}$ be the model-generated texture, and let $\mathbf{g}_{img}$ be the part of the image texture inside the shape region. Then, minimizing the residual $\mathbf{r}(\mathbf{p}) = \mathbf{g} - \mathbf{g}_{img}$ amounts after a first order Taylor expansion to expressing the parameters update as a function of the residual, or:

$$\delta\mathbf{p} = -\mathbf{R}.\mathbf{r}(\mathbf{p}) \ . \tag{3.11}$$

The key approximation of AAM is that the matrix $\mathbf{R}$ is assumed to be fixed. This matrix is learned from the training set using numeric differentiation, and hence provides a priori knowledge on the model parameters adjustment during the search. This approximation leads to a tractable algorithm that solves an initially-tough high-dimensional optimization problem.

Different modifications to the standard AAM exist in the literature, we briefly report a few of them. [Matthews & Baker 2004] introduce a new computationally-efficient analytical gradient descent algorithm to optimize the AAM search and demonstrate its applicability for face tracking. [Andreopoulos & Tsotsos 2008] extend this latter approach, initially presented in 2D, to the 3D case and apply it to the segmentation of cardiac MRI images. [Donner *et al.* 2006] use the canonical correlation analysis algorithm (CCA) in order to capture the correlation between the parameter update and the image residual. This alternative method to compute the matrix $\mathbf{R}$ yields more precise parameter prediction than the standard numeric differentiation method and allows speeding up the search by a factor of four. [Gross *et al.* 2006] build on their previous work in [Matthews & Baker 2004] to develop a robust AAM with respect to occlusions.

Graph-based Methods and Related Work

[Boykov & Jolly 2001] introduce an interactive method of n-dimensional image segmentation based on graph cut. The user provides hard segmentation constraints by specifying some pixels as belonging to the object or the background. Then, the segmentation problem is cast into a min-cut/max-flow

problem, where the internal nodes of the graph represent the image pixels, and the source and sink node refer to the object and background classes. The internal graph links represent the neighboring system in the image, and their weights define a discontinuity penalty that is related to the pixel intensity differences (a boundary-related term). The terminal weights represent the penalty of assigning a pixel to one class or the other. These penalties represent the likelihood of the pixel intensity given the intensity distribution of the labeling class. The hard constraints are enforced by pairs of $(0, K)$ weights, where K is a large penalty. The segmentation of the image is given by the min-cut. This work is further reviewed in [Boykov & Funka-Lea 2006]. The authors survey the developments and variations of their method and discuss the historical steps that led to a ubiquity of the graph cut approach. Although the multi-label case is NP-hard, an extension using multi-way graph cut can be done in practice using α-expansion for instance. Different improvements were added to the standard graph cut formulation [Boykov & Jolly 2001]. For instance, [Blake *et al.* 2004] propose to learn the regional parameters from color images using a Gaussian Mixture Markov Random Field (GMMRF) generative model. This idea is combined in the "Grabcut" algorithm [Rother *et al.* 2004] with an iterative minimization, where from an initial loose initialization provided by the user, the parameters of the model are updated and then the segmentation is estimated using graph cut, until convergence. This algorithm alleviates the user interactions while providing good performance. [Li *et al.* 2004b] use watershed pre-segmentation to increase the efficiency of the algorithm. Such pre-segmentation provides super-pixels that are used to build the graph.

In a different line of work, [Li *et al.* 2006] propose a method to segment multiple 3D "terrain-like" surfaces. Their algorithm can combine edge-based costs and region based weights. They define several geometric constraints to enforce surface smoothness and inter-dependencies between the different objects. They use an image unfolding operation to handle tubular structures. The optimization is performed by an s/t min-cut in an appropriately constructed graph to take into account both image costs and geometric constraints. They applied their method to 3D medical image segmentation of the diaphragm as well as the inner and outer walls of vascular structures.

[Zabih & Kolmogorov 2004] propose a pixelwise segmentation method

based on clustering. Their framework is general and applies clustering to image pixels according to their feature vectors, while imposing spatial coherence. Their algorithm is applicable to parametric as well as non-parametric clustering methods, as long as the cluster quality function meets a linearity criterion with respect to the image pixels. The algorithm is twofold, similarly to the expectation maximization algorithm (EM). Iteratively, the clusters parameters are fixed, and the best corresponding pixel labeling is found by solving a graph cut using α-expansion. Then, given the pixel labels, new cluster parameters are estimated.

[Shi & Malik 2000] point out that the graph cut segmentation method is biased towards small region cuts that minimize the length of the partition boundary. They tackle this problem using a graph spectral approach, by defining a new grouping criterion. This introduced dissociation measure is called normalized cut. They obtain an approximate solution to the problem by defining a relaxed formulation, expressed as a generalized eigenvalue system. [Grady 2006] introduced a new graph-theoretical algorithm for multi-label interactive image segmentation, known as the random walker algorithm. The method requires regions seeds defined by the user. The algorithm can be intuitively thought of as a random walker that starts from each pixel and arrives to each seed with a given probability. Then the largest probability values determine a labeling (a segmentation) of the image. The random walk is biased by the image structure, which is encoded by the weights of the edges connecting the pixels. [Grady 2006] provides a tractable means of solving the random walker problem by formulating it as a Dirichlet problem. By the use of combinatorial operators, the problem boils down to solving a system of linear equations. However, the size of this system depends on the size of the corresponding image, which can be very large in practice, in the 3D medical imaging setting for instance. Such an observation prohibits the use of direct linear system solvers, and motivates the use of iterative solvers. The random walks algorithm provides a global solution to the problem, and unlike graph cut, does not suffer from the metricity bias or the "shrinking bias". However, for the seeded image segmentation problem, this algorithm is biased towards the spatial locations of the seeds. [Grady 2005] extends the previous work to alleviate the need for user initialization by using prior intensity prob-

ability densities. This modification allows the segmentation of disconnected objects belonging to the same class. A further improvement is presented in [Singaraju *et al.* 2008] where directed edges are used to construct the graph, resulting in directional asymmetric weights. This new formulation allows the improvement of the seeded segmentation results, by balancing the bias towards seeds locations.

Although the random walker formulation seems to be different from the graph cut formalism, it turns out that these two approaches are closely related. [Sinop & Grady 2007] present a unifying framework for seeded segmentation algorithms, the p-norm of the spatial gradient of a potential function is minimized, subject to Dirichlet boundary conditions, specified by the seeds. They show that using the ℓ_1 norm is equivalent to graph cut, whereas using the ℓ_2 norm leads to the random walker algorithm. [Singaraju *et al.* 2009] build on this work and interpret the potential function as an MRF with continuous valued variables. The MAP estimate corresponds to the minimization of the the defined general cost function in [Sinop & Grady 2007]. They tackle this problem by using interactive least square techniques. Moreover, by choosing an appropriate value of p (between 1 and 2), a trade-off between the graph cut behavior and the random walker behavior can be achieved. They show in the experiments that with $p = 1.5$, drawbacks of both methods are minimized.

As we have already noted, one drawback of graph cut methods for image segmentation is related to their bias towards shorter boundaries, which is known as the "shrinking bias". In particular, common techniques miss thin elongated structures during segmentation. This observation motivated the work of [Vicente *et al.* 2008], who tackled this problem by adding constraints to the energy minimization problem. These constraints are intuitively translated into an interactive segmentation framework. Starting from an imperfect s/t segmentation, where these structures are missing, the user would click on pixels that must be connected to the object (through thin components). The ensuing minimization problem turns out to be NP-hard. However, the authors develop a practical optimization algorithm called DijkstraGC, that combines aspects from the Dijkstra algorithm and graph cuts. The proposed approach performs well in practice, and the resulting interactive segmentation framework can handle thin image structures. An alternative (and slower)

minimization strategy based on dual decomposition (DD) was considered as it provides a lower bound on the optimal energy in general, and the optimal solution in a special case. This property was used to verify that DijkstraGC also achieved an optimal performance in these cases.

An alternative way to improve the graph cut technique involves incorporating shape priors. [Funka-Lea *et al.* 2006] apply the binary graph-cut technique to the segmentation of the entire heart in computed tomography (CT) images. This entire heart segmentation serves visualization purposes as it allows a 3D reconstruction that exhibits the coronary arteries. An automatic seed detection sets the hard unary potential constraints. A Potts interaction model is used, and an additional pairwise constraint is introduced. This so-called "blob" energy term encourages graph cuts that produce a shape which edges are orthogonal to the rays coming from the seed region. A dilation of the graph-cut output image mask gives the final result.

[Veksler 2008] adds a new pairwise regularization term to the standard s/t graph cut image segmentation formulation. This term corresponds to a generic shape prior that is related to a large class of shapes, namely the star-shaped objects. The prior term assumes knowing a point inside the object that is a center of the star shape. This seed point is often provided by the user in interactive image segmentation tasks. Subsequently, she shows that the defined prior can be used as a "ballooning" force that biases the segmentation towards longer partition boundaries. This can be achieved by an appropriate parameter choice and allows the author to counter the "shrinking bias".

[Das *et al.* 2009] adopt a similar approach aiming to incorporate a shape prior in the graph cut segmentation. They define a class of shapes called "compact" shapes. These shapes satisfy some regularity properties on a 4-neighborhood grid with respect to an interior point, which is assumed in practice to be close to the center of the object and specified by the user. Then, the standard boundary-related pairwise potential is modified such that shapes that violate the "compactness" condition are highly penalized. Moreover, a "bias" parameter is introduced similarly to [Veksler 2008]. By varying its values, one can encourage the cut towards longer object boundaries, and hence compensate for the "shrinking bias".

The work of [Zeng *et al.* 2008] is in the line of continuous effort to incor-

porate global priors into graph cut segmentation. Their approach modifies the max-flow algorithm to enforce topological consistency between an initialization and the final segmentation. Their work is based on the concepts of topology of digital images and simple points [Bertrand 1994]. They prove that the defined topology cut problem is NP-hard. However, they propose an algorithm that converges to a local minimum of the energy function. In the experiments, they successfully apply their method and show examples of topology-preserving segmentations. Although the last three methods account for shape characteristics, their prior remains generic, and they do not allow specifying a particular class of shapes. This is exactly what other researchers worked towards.

[Slabaugh & Unal 2005] use novel terminal weights as a formulation of an elliptic shape prior. This geometric primitive adds regularity constraints to the segmentation. They propose an iterative framework, where a binary elliptic mask is repeatedly formed knowing the current ellipse position in the image, mean image intensity inside and outside the ellipse are computed, an s/t graph cut is performed such that the regional term is related to the computed mean intensity values, and the shape prior term is enforced by the binary elliptic mask, using the t-weights. Then, a new ellipse is fitted to the result of the graph cut, until convergence. Similar ideas are also used in the work of [Lin *et al.* 2005].

An alternative attempt to enforce prior knowledge in the graph cut interactive segmentation setting is presented in [Freedman & Zhang 2005]. The authors use a fixed template shape prior, represented by an unsigned distance map Φ. Then, for every neighboring pixels p and q in the image, an additional n-link weight is defined as $\Phi(\frac{p+q}{2})$. This penalty encourages the cut to take place on the zero level of the distance map Φ, i.e. the learned contour. To deal with the pose estimation problem, [Freedman & Zhang 2005] rely on the user input to compute a rigid transformation that maps the distance map to the image, using Procrustes Analysis. However, the user input is very often insufficient, which impairs the quality of the rigid alignment. Then, scale variations are handled in a brute force manner. The use of this template prior shows improvements in the result of the graph cut-based segmentation.

[Zhu-Jacquot & Zabih 2007] and [Zhu-Jacquot & Zabih 2008] introduce a

method that combines statistical or parametric shape priors with graph cut in an EM framework, and apply it to the segmentation of the LV in 2D MR image. In [Zhu-Jacquot & Zabih 2007], they apply PCA to the SDMs of training shapes to learn a statistical shape prior, whereas in [Zhu-Jacquot & Zabih 2008], they use a geometric parametrized prior (concentric circles). In both cases, the shape prior is represented by a set of parameters, and the image intensity is modeled using a mixture of Gaussians. In the expectation step, the shape parameters as well as the photometric model are updated using a gradient descent, and in the maximization step, the pixels of the image are labeled according to the shape prior and the intensity model constraints using graph cut (or α-expansion if there are more than 2 classes). The process is iterated until convergence.

Another approach that incorporates a statistical shape prior into an iterative graph cut segmentation process is proposed in [Malcolm *et al.* 2007a]. They use KPCA to learn a statistical shape model from binary images representing the training set. The algorithm repeats the following operations. First, object and background histograms are computed from the current pixel labeling. Then, a pre-image of the projection in the feature space of the current segmentation is formed, using an approximate algebraic expression. This pre-image is used to introduce a Bayesian shape prior in a unary regional image term. Then, a graph cut is performed and a new pixel labeling is produced. The algorithm iterates these operations, starting from an initial segmentation provided by the user, until convergence.

[Ali *et al.* 2007] propose a method to segment 2D Dynamic Contrast Enhanced Magnetic Resonance Imaging (DCE-MRI) slices. Their approach combines graph cut and shape priors. They build from the aligned training slices a "shape image" or a template that is composed of three regions: the kidney, the background, and an area of variability, where the probability of belonging to one class decays exponentially as the distance to the considered class increases. Then, this probabilistic template is used during segmentation, by incorporating a shape prior term for each pixel, as a terminal weight. To this end, a preliminary alignment of the query image with respect to the training images is necessary, and is performed using a rigid transformation.

[Vu & Manjunath 2008] propose a segmentation algorithm based on a 2D

shape prior within the binary graph cut framework. They use a template shape prior (a binary mask) and enforce the corresponding constraint using a shape distance derived from [Chan & Zhu 2005]. Essentially, in the discrete setting, this distance evaluates the region of disagreement between two binary shape images. This representation allows arbitrary topologies. Hence, the distance to the template of some labeling can be decomposed over the image pixels. Thus, [Vu & Manjunath 2008] define the shape prior term as terminal weights. They also resort to a normalization technique to obtain invariance properties with respect to rigid motion in their model. Their framework allows the segmentation of multiple overlapping instances of the object, through the so-called multi-phase graph cut. This algorithm solves iteratively a graph cut for each object instance, while accounting for the overlap between objects through an appropriate choice of the data term (which is also encoded as terminal weights). Each iteration is exactly solved by the min-cut/max-flow algorithm, due to the submodularity of the regularizing pairwise terms. However, the multi-phase iterative graph cut algorithm is not guaranteed to converge to the global optimum. An extension of the approach allows the use of multiple shape templates to encode the prior knowledge. Successful segmentation results under noise and occlusion are achieved.

Another template-based approach is introduced in [Schoenemann & Cremers 2007]. This method benefits from global optimization based on graph theory. Their method seeks a 2D curve such that it is close to strong gradient locations and similar to a template curve. They also seek to enforce invariance with respect to similarity transformations. They define an energy that is the sum of three terms. The first one integrates a decreasing function of the norm of the gradient along the curve. The second term is translation-invariant, and measures how well the curve is aligned with the template. In the perfect case, the obtained mapping should be a diffeomorphism. Then, a third term is introduced as a penalization of the mapping. The problem is then cast as finding the best curve and the best mapping that minimize the defined energy functional. All these terms are actually normalized by the length of the sought curve to eliminate bias toward shorter boundaries. This ratio of integral along the curve paves the way to the use of the minimum ratio cycles (MRC)

minimization technique [Jermyn & Ishikawa 2001]. Then, inspired by the work in [Ishikawa & Jermyn 2001], scale invariance is enforced through the use of a product graph. In this appropriately built graph, the authors solve for a MRC optimally and efficiently, by parallelizing computations. Their model is not however rotation-invariant. This is addressed by matching to different rotated instances of the template.

More recently, [Ben Ayed *et al.* 2009] introduce a novel graph-cut-based algorithm to segment 2D left ventricle (LV) images in magnetic resonance (MR) sequences. In their method, the user manually segments the first frame, by drawing a single contour that partitions the image into two disjoint sets (the object and the background), and the algorithm propagates the segmentation to the remaining ones. They argue that this user intervention alleviates the need for intensive shape training of the geometric and photometric features. Using this s/t min-cut formulation, they show improvements over similar variational approaches. More specifically, they define geometric and photometric constraints using distribution matching, rather than pixelwise costs. They define the optimal segmentation (or binary labeling) in an unseen image frame as the one that maximizes the Bhattacharyya coefficient of the color and shape object distributions with respect to the one of the manually segmented frame. The image intensity distribution is defined using a kernel density estimate. The shape of the object is described by the collection of normalized distances separating all its interior points and a centroid point. A distance distribution is also defined using a kernel density estimation. In addition to these distribution matching terms, a smoothness constraint that is related to the length of the partition boundary is considered. The obtained minimization criterion is then relaxed using a first-order approximation of the Bhattacharyya coefficient. The resulting energy is defined as the sum of unary and pairwise potentials, that are suitable to s/t graph cut minimization. Moreover, a global optimum is guaranteed as the pairwise penalties are submodular.

Discussion & Related Work The aforementioned graph-based segmentation methods that made use of a (possibly statistical) shape prior essentially model the image pixels using a MRF. Hence, the MRF variables account for pixel labels. This representation narrows the possibil-

ities of incorporating a shape model while being computationally expensive. With the exception of very few techniques ([Ben Ayed *et al.* 2009], [Schoenemann & Cremers 2007]), the overwhelming majority of the approaches define the shape prior at the pixel level, through the t-links weights. To this end, an implicit shape representation through a distance map [Freedman & Zhang 2005], or the use of a binary shape image [Vu & Manjunath 2008] is required. A single template representation is insufficient to capture large shape variations, and the use of numerous templates to enforce variability might be limited compared to statistical approaches, and comes with a large computational burden. This template representation applies also to the statistical approaches, where a probabilistic shape image [Ali *et al.* 2007], or a pre-image of the projection of the segmentation on learned subspace [Malcolm *et al.* 2007a] were used. We also notice that in many cases, the coupling between the statistical shape parameters and the MRF variable is complex, and is therefore accounted for through the use of iterative optimization schemes, leading to suboptimal solutions. We note also that the need of aligning the template to the image is an ill-posed problem because the pose of the object to be segmented is a priori unknown. In addition, the feasibility of extending these methods to the 3D case is not always guaranteed. Nevertheless, the flexibility of the graph cut framework allows these different methods to combine several image cues, from regional features to boundary evidence. A different modeling approach uses the MRF to represent parts of the object, or boundary landmarks. These methods are more closely related to our work.

[Zhang *et al.* 2004] introduce a technique to segment human bodies using a graphical model. They represent the articulated shape using a triangulated silhouette, with a specified vertex elimination order. They augment this representation by a parametrization that specifies articulation joint angles. The graphical model encodes the dependencies among the nine considered angles, whereas each silhouette landmark is considered to be dependent on its parent triangle. The root triangle corresponds to the face part. After learning the parameters from a large database of training data, the obtained translation-invariant model is combined with several visual cues (edge gradient map, background subtraction, skin color, etc.) to perform human body segmenta-

tion. The inference is realized using particle filtering. Although the presented results are appealing, this method suffers from its application-specific design. Hence, it is not generalizable in a straightforward manner to other applications, or to higher dimensions.

[Felzenszwalb 2005] adopts a different approach, that is related to the spirit of the pictorial structures work [Felzenszwalb & Huttenlocher 2005]. A triangulated polygons representation of 2D shapes is proposed, its dual structure is a tree (the model is in fact a 2-tree, which is a chordal graph). Such a template is used to segment an unseen image, using boundary information. Benefiting from the perfect elimination scheme property of their graph, they are able to provide a global solution to the segmentation problem. The shape prior is enforced in each triangle clique through a constraint on the local induced transformation, which is required to be as close to a similarity as possible. The resulting segmentation algorithm is similarity-invariant and handles naturally occlusions.

[Seghers *et al.* 2007a] propose a segmentation method that shares ideas with the approach in [Felzenszwalb 2005]. A 2D contour is represented by a set of landmarks structured in a chain. For each landmark, a classifier is trained based on a locally orderless image feature vector, following [van Ginneken *et al.* 2002]. Segmentation is then cast into the problem of selecting the best candidate points found by the classifier, while satisfying the statistical shape prior constraints. The latter are defined by the Gaussian distributions of the relative positions of two successive landmarks on the chain. A higher order prior that conditions this relative position with respect to the one of the previous landmark pair on the chain was also used. Inference was done using a dynamic programming approach, namely the minimal cost path (MCP) between the "root" and the "leaf" of the chain. Successful segmentation results of the lungs in standard posterior-anterior chest radiographs were shown. A more complex model, like for instance a closed contour, or a connected graph, makes the use of dynamic programming impossible. Moreover, this exact inference is not extendable as it is to the 3D case (without making approximations), to handle minimal cost surfaces ([Grady 2010] has recently surveyed the work of [Sullivan 1990] and proposes a method that tackles the problem). Therefore, [Seghers *et al.* 2007b] resort to approximate optimiza-

tion techniques to extend their work to the 3D case, and showed results of liver segmentation in CT images. A heuristic approach called iterative dynamic programming is used, where the problem is solved on randomly extracted paths from the meshed surface using DP, and the solutions are combined using a voting scheme. In [Seghers *et al.* 2008] mean field annealing (MFA) was also considered for optimization.

Another classifier-based discrete detection method is proposed in [Donner *et al.* 2010]. Their approach is based on symmetry interest points and descriptors, that are extracted from gradient vector flow (GVF) fields [Xu & Prince 1998]. Given a binary edge map, the gradient vector flow is nearly equal to its gradient when the latter is large, and varies smoothly in homogeneous regions. GVF minima locations are of a great interest as they correspond to local symmetry points. [Donner *et al.* 2010] use this property to define their MRF model. The MRF nodes correspond to landmarks at local symmetry locations, and the edges encode inter-landmark dependencies. Local descriptors using local GVF around the interest points are used to define the unary potentials, as well as the pairwise shape prior. The latter uses relative edge / descriptor orientation as well as the length of the edges. Therefore, the shape prior is invariant with respect to rigid motion. During inference, candidate points are extracted from the unseen image using its GVF. The model is matched to the image as a MAP-MRF problem that is solved using the MAX-SUM diffusion algorithm introduced by [Schlesinger 1976] and reviewed in [Werner 2007]. They apply their method to the segmentation of 2D hand X-rays and 2D spine MRIs [Donner *et al.* 2007b]. To make the training phase easier and lessen the manual annotation needs, they propose in [Donner *et al.* 2009] a technique to learn the model in a weakly-supervised manner. Note however that their approach does not handle the object boundaries. Thus, it can be considered as techniques to be used more for detection than for segmentation.

[Kumar *et al.* 2010] present a knowledge-based segmentation approach that bridges the gap between graph-cut methods and pictorial structures. On one hand, the image is modeled by a MRF, and a binary labeling $\mathbf{l}$ that specifies the object and the background is sought. On the other hand, they use a layered pictorial structure (LPS) model to represent articulated objects, where

the parts are connected in a complete graph (for a better spatial consistency). The layer term refers to a depth number associated with each part, in order to allow overlap between the different components. A non-articulated object was modeled as a single part. Each part was specified by a set of exemplars, forming the so-called set of exemplars model (SOE), as well as a pose parameters vector. The authors provide a Bayesian formulation to the problem using a graphical model representation, that adds essentially a shape variable $\mathcal{S}$ (referring to the LPS) on top of the MRF representing the image. In addition to the standard s/t graph cut regional and boundary terms, the defined energy features a shape prior term, that is computed using a SDM of the prior shape instance. Inference in a new image $\mathcal{I}$ is performed by maximizing the posterior $p(\mathbf{l}|\mathcal{I})$, instead of the joint probability $p(\mathbf{l}, \mathcal{S}|\mathcal{I})$. This is done by marginalizing over the shape prior. This marginalization is performed approximately and efficiently through a sampling strategy. Indeed, the authors suggest to use the generalized expectation maximization algorithm (EM). In the E-step, BP is applied to estimate the probability of the shape knowing the current image labeling, in order to perform sampling. During the M-step, maximizing the posterior turns out to be efficiently approximated by solving only one s/t graph cut, where the shape prior cost corresponds to a linear combination of the costs induced by the sampled shape instances in the E-step, weighted by their respective likelihoods. The so-termed OBJCUT algorithm converges to a local minimum.

The aforementioned line of research considers top-down modeling using MRFs or graphical models. While [Felzenszwalb 2005] succeeds in exactly optimizing the energy functional he defines, the method does not generalize to higher dimensions. More complex modeling that would overcome this limitation as in [Kumar *et al.* 2010] makes the use of approximations inevitable. Moreover, the complexity of the method would highly increase in the 3D case since it is related to the number of image pixels, and the search space for the pose estimation will be substantially bigger. The approaches proposed in [Seghers *et al.* 2007a] and [Donner *et al.* 2010] rely exclusively on edge-based cues or local features. In some applications, the use of regional cues can be of a great interest.

In the following sections, we introduce a nearly fully-automatic segmenta-

tion framework that permits to plug-in various image cues, as in the standard graph cut setting [Boykov & Jolly 2001]. The use of regional image information is introduced in section 3.2, and classifier-based segmentation is explored in section 3.3, in a relatively similar manner to the work of [Seghers *et al.* 2007a] and [Donner *et al.* 2010]. Our formulation does not make any assumption about the dimensionality of the image. It is hence extendable to the 3D case, at a relatively low cost. The complexity of our method is related to the number of landmarks, and as we do not explicitly model the pose parameters, the increase in the dimensionality of the search space does not affect the tractability of the problem. Moreover, for the special case of regional image features, we benefit from the Green's theorem to substantially speed up computations. As far as optimization is concerned, we resort to approximate inference and our method leads to a local minimum in general. However, the experiments show that in practice, satisfactory results are obtained, and that the segmentation algorithm is robust to noise and occlusions thanks to the shape prior. In a particular case where the dimension of the problem is small, we show that exact inference is tractable using dynamic programming.

We also mention that our work is related to the registration algorithm of [Glocker *et al.* 2008] (initially introduced in [Glocker *et al.* 2007]) and to the one of [Tang & Chung 2007]. Both methods are based on MRFs where the labeling corresponds to quantized displacements of control points. This idea will be used in the following.

3.2 Region-based Segmentation

The main challenges of knowledge-based segmentation are: (i) appropriate modeling of shape variations, (ii) successful inference between the image and the shape model. Let us consider the simplest possible scenario that aims to detect an object of particular interest from the background in an image $\mathcal{I}$. Then, the prior-based segmentation task corresponds to the estimation of the shape model parameters from the observation. Formally, ones seeks to maximize the posterior probability $p(\mathcal{G}_k|\mathcal{I})$, where $\mathcal{G}_k = (\mathcal{V}, \mathcal{E}_k)$ is the k-fan statistical shape model introduced in section 2.2. By a simple manipulation

using the Bayes rule, the segmentation is expressed as the MAP estimation of the MRF parameters, or:

$$\max \; p\left(\mathcal{G}_k\right) p\left(\mathcal{I}|\mathcal{G}_k\right) = \max \; p\left(\mathbf{x}_1, \ldots, \mathbf{x}_n\right) p\left(\mathcal{I}|\mathbf{x}_1, \ldots, \mathbf{x}_n\right) \; . \tag{3.12}$$

By taking the negative log of the latter equation, we formulate the segmentation problem as an energy minimization problem. We seek optimal landmark positions $(\mathbf{x}_1^*, \ldots, \mathbf{x}_n^*)$ such that:

$$\begin{aligned} (\mathbf{x}_1^*, \ldots, \mathbf{x}_n^*) &= \underset{(\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \Omega^n}{\operatorname{argmin}} \; E\left(\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathcal{I}\right) \\ &= \underset{(\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \Omega^n}{\operatorname{argmin}} \; E_{shape}\left(\mathbf{x}_1, \ldots, \mathbf{x}_n\right) + E_{image}\left(\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathcal{I}\right) \; . \end{aligned} \tag{3.13}$$

The energy $E\left(\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathcal{I}\right)$ corresponding to the landmark positions $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$[1] in the image domain Ω is the sum of a data-related term $E_{image}\left(\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathcal{I}\right)$ expressing the local image cost conditioned on the control points, and a prior term $E_{shape}\left(\mathbf{x}_1, \ldots, \mathbf{x}_n\right)$ expressing the cost of deforming the landmark pairs in $\mathcal{E}_k$ with respect to the prior learned distributions. We explain in this section how we define these two energy terms, and we detail in particular the relationships between the control points and the image domain. We also develop an optimization procedure that enables to solve (3.13) using an efficient discrete optimization algorithm.

3.2.1 Regional Statistics and Image Segmentation

By applying the data-related cost, we seek the optimal separation of the object from the background in terms of visual properties. Let $p_{\mathcal{O}}$ and $p_{\mathcal{B}}$ be the prior densities for these two hypotheses. We suppose that the control points form a closed boundary, or that there is a tractable way of inferring such a boundary from the control points positions, using an interpolation scheme for instance. Then, they partition the image domain Ω into an object domain $\mathcal{O}$ and a background domain $\mathcal{B}$. Note that we consider this simplifying hypothesis for the sake of a simple presentation and lighter notations. In the

[1] This is a notation where no difference is made between the random variables and their values.

general case, the image domain Ω is partitioned into several $\mathcal{O}^r$ objects and a background domain $\mathcal{B}$. The following derivation holds in this case too. Upon considering the negative log of the posterior probabilities, we express the cost $E_{image}(\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathcal{I})$ using the regional statistics [Zhu & Yuille 1996] as follows:

$$\begin{aligned} E_{image}(\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathcal{I}) &= \int_{\mathcal{O}} -\log\left(p_{\mathcal{O}}(\mathcal{I}_y)\right) dy + \int_{\mathcal{B}} -\log\left(p_{\mathcal{B}}(\mathcal{I}_y)\right) dy \\ &= \int_{\mathcal{O}} -\log\left(\frac{p_{\mathcal{O}}(\mathcal{I}_y)}{p_{\mathcal{B}}(\mathcal{I}_y)}\right) dy + \int_{\mathcal{O}} -\log\left(p_{\mathcal{B}}(\mathcal{I}_y)\right) dy \\ &\quad + \int_{\mathcal{B}} -\log\left(p_{\mathcal{B}}(\mathcal{I}_y)\right) dy \\ &= \int_{\mathcal{O}} -\log\left(\frac{p_{\mathcal{O}}(\mathcal{I}_y)}{p_{\mathcal{B}}(\mathcal{I}_y)}\right) dy + \underbrace{\int_{\Omega} -\log\left(p_{\mathcal{B}}(\mathcal{I}_y)\right) dy}_{\text{constant}} . \end{aligned} \tag{3.14}$$

In order to evaluate this component and associate it with the proposed shape representation, we decompose the image domain Ω according to the control points $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ by considering their Voronoi diagram (a survey on Voronoi diagrams is available in [Aurenhammer 1991]). The Voronoi tessellation partitions Ω into cells Ω_i that are associated with the control points $\mathbf{x}_i$. Every cell Ω_i consists of all image pixels closer to $\mathbf{x}_i$ than to any other landmark. Fig. 3.1 gives an example of a Voronoi diagram. Note that the Voronoi tessellation is the dual of the Delaunay triangulation. Hence we can write $\Omega = \cup_{i=1}^{n} \Omega_i$. By intersecting these Voronoi cells with the "object" domain, we obtain the partition $\mathcal{O} = \cup_{i=1}^{n} \mathcal{O}_i$ that relates each pixel of the object part of the image to one control point, with $\mathcal{O}_i = \mathcal{O} \cap \Omega_i$. Note that thanks to the rearrangement of equation (3.14), we only need to consider the "object" domain pixels, and this is equivalent to using all the image information. Thanks to the Voronoi tessellation, one can decompose the global image term (3.14) into sub-terms which are defined at the partition cells as follows, up to an additive constant:

$$E_{image}(\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathcal{I}) = \sum_{i=1}^{n} V_i(\mathbf{x}_i, \mathcal{I}) \tag{3.15}$$

with

$$V_i(\mathbf{x}_i, \mathcal{I}) = \int_{\mathcal{O}_i} -\log\left(\frac{p_{\mathcal{O}}(\mathcal{I}_y)}{p_{\mathcal{B}}(\mathcal{I}_y)}\right) dy \tag{3.16}$$

being the image-related cost associated with the control point position $\mathbf{x}_i$. Clearly, the notation in equation (3.16) is abusive, because the data term $V_i(\mathbf{x}_i, \mathcal{I})$ depends on the positions of all the control points, through the use of the Voronoi cells. The independence hypothesis will be considered under some mild assumptions that will be detailed later. The above defined terms can be calculated very efficiently in practice by combining rasterization techniques and fast integral computing over polygons [Unal *et al.* 2004]. This observation stems from Green's theorem, that permits to compute the integral in (3.16) using the boundaries $\partial\mathcal{O}_i$ of $\mathcal{O}_i$, or:

$$V_i(\mathbf{x}_i, \mathcal{I}) = \int_{\partial\mathcal{O}_i} \widetilde{\mathcal{I}}_\omega d\omega \ , \tag{3.17}$$

where $\widetilde{\mathcal{I}}$ is an integral of the image $-\log\left(\frac{p_\mathcal{O}(\mathcal{I})}{p_\mathcal{B}(\mathcal{I})}\right)$. Green's theorem had also been applied in a similar context in earlier segmentation algorithms, as in [Chakraborty *et al.* 1996] for example. We should note that this term is equivalent to using the entire image domain to determine the image support and can be replaced either using more complex descriptors (this will be explored in section 3.3), or through edge-driven support. Note that equation (3.16) generalizes to the case of multiple object domains $\mathcal{O}^r$ in a straightforward manner:

$$V_i(\mathbf{x}_i, \mathcal{I}) = \sum_r \int_{\mathcal{O}_i^r} -\log\left(\frac{p_{\mathcal{O}^r}(\mathcal{I}_y)}{p_\mathcal{B}(\mathcal{I}_y)}\right) dy \ , \tag{3.18}$$

where $\mathcal{O}_i^r = \mathcal{O}^r \cap \mathcal{O}_i$, and $p_{\mathcal{O}^r}$ is the prior distribution of the color values in $\mathcal{O}^r$. In practice, we used simple Gaussian models in some applications and mixture of Gaussians models in others to model the object and the background intensity distributions. Such a component will perform well if the data support is strong but will fail to deal with noise, clutter, missing parts, etc. The use of prior knowledge on the expected geometry of the shape could address the above mentioned limitations.

3.2.2 Prior Knowledge and Image Segmentation

In the context of our approach, we have defined the shape model as a k-fan graph. Furthermore, we were able to determine an approximate density of this model using a small number of joint probabilities. In order to impose the

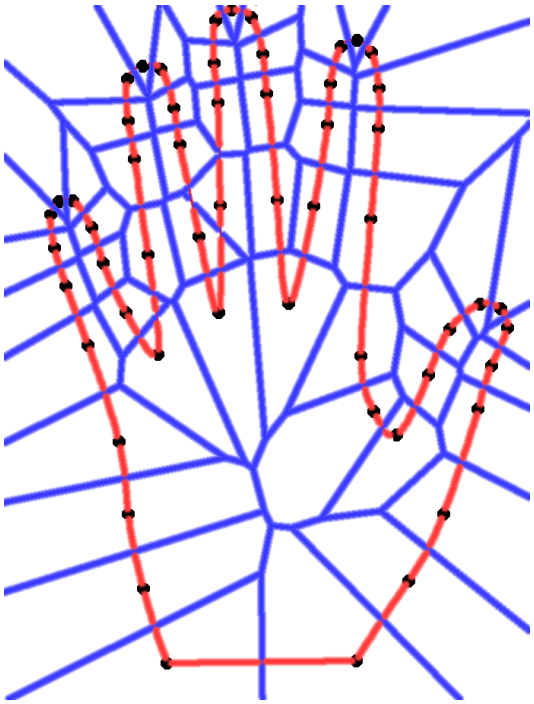


Figure 3.1: Voronoi decomposition of the model domain.

prior, we minimize the cost $E_{shape}(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ that we decompose over all the pairs in $\mathcal{E}_k$, as derived in equation (2.30):

$$E_{shape}(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \sum_{(i,j)\in\mathcal{E}_k} V_{ij}(\mathbf{x}_i, \mathbf{x}_j) \ , \tag{3.19}$$

with

$$V_{ij}(\mathbf{x}_i, \mathbf{x}_j) = -\gamma_{ij} \log\left(p\left(\frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{\widetilde{d}}\right)\right) \ . \tag{3.20}$$

Note that equation (3.20) carries a notation abuse as well. Indeed, the shape scale $\widetilde{d} = \frac{2}{n(n-1)} \sum_{i<j} \|\mathbf{x}_i - \mathbf{x}_j\|$ depends on the positions of all the control points. This issue will be accounted for in the following development. This model allows for the encoding of global dependencies as local combinations of individual pairwise densities.

One can now combine the data term with the prior term towards knowledge-based segmentation, by combining (3.13), (3.15) and (3.19):

$$E(\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathcal{I}) = \alpha \sum_{i=1}^{n} V_i(\mathbf{x}_i, \mathcal{I}) + \sum_{(i,j)\in\mathcal{E}_k} V_{ij}(\mathbf{x}_i, \mathbf{x}_j) \ . \tag{3.21}$$

The factor α that we introduced in the above equation weights the relative importance of the data term with respect to the prior term. Note however that the relative weightings of the base k-clique landmark pairs with respect to the "non-reference" pairs are defined by γ_{ij} (in equation (2.29)).

3.2.3 Energy Minimization

The optimization of the cost function (3.21) in the continuous domain is rather problematic. One can expect that it is not convex and therefore a gradient-based optimization is prone to failure. In order to optimize such an energy functional, we consider recent results from discrete optimization, and adopt a quantization approach similar to those presented in [Glocker *et al.* 2007] and in [Tang & Chung 2007].

We make two assumptions that are most often verified in practice. First, the initial positions of the control points are within the image domain, and that is why we can assume an upper bound on the maximum displacements that would lead to the solution. Second, we consider that the precision required about the solution is specified, which enables to choose a quantization step.

Let us suppose that we are given an initialization of the shape $(\mathbf{x}_1^0, \ldots, \mathbf{x}_n^0)$. Then, a segmentation of the image refers to an optimal displacement of the control points, that minimizes the defined energy. Then, we can approximate the continuous deformations of our shape model towards the solution by a finite set of displacements vectors $\mathcal{D} = \{\mathbf{d}_1, \ldots, \mathbf{d}_z\}$. Let $\mathcal{L} = \{1, \ldots, z\}$ be the set of labels associated to the quantization $\mathcal{D}$ of the displacements. Then, displacing the control point $\mathbf{x}_i^0$ by the vector $\mathbf{d}_{l_i}$ is equivalent to assigning the label $l_i \in \mathcal{L}$ to $\mathbf{x}_i^0$. The minimization of the energy in (3.21) can be hence written as a labeling problem, or:

$$(l_1^*, \ldots, l_n^*) = \underset{l_i \in \mathcal{L}}{\operatorname{argmin}}\ E\left(\mathbf{x}_1^0, \ldots, \mathbf{x}_n^0, \mathcal{I}, (l_1, \ldots, l_n)\right) \quad , \tag{3.22}$$

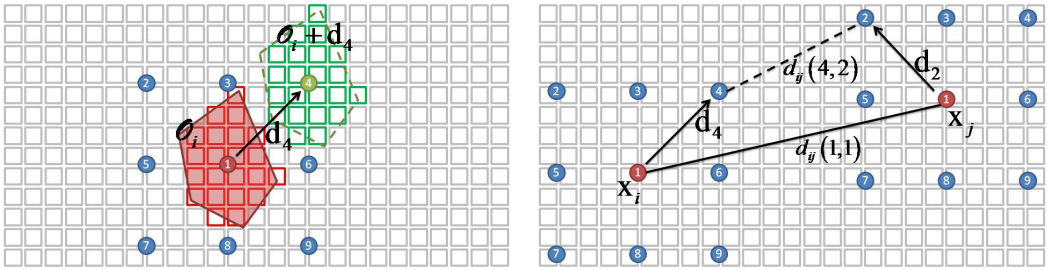
where the data unary terms and the prior pairwise terms in (3.21) are adjusted as follows:

$$V_i(\mathbf{x}_i, \mathcal{I}, l_i) = \int_{\mathcal{O}_i + \mathbf{d}_{l_i}} -\log\left(\frac{p_{\mathcal{O}}\left(\mathcal{I}_y\right)}{p_{\mathcal{B}}\left(\mathcal{I}_y\right)}\right) dy \tag{3.23}$$

$$V_{ij}(\mathbf{x}_i, \mathbf{x}_j, l_i, l_j) = -\gamma_{ij} \log\left(p\left(\frac{d_{ij}(l_i, l_j)}{\widetilde{d}}\right)\right) \quad , \tag{3.24}$$

where $d_{ij}(l_i, l_j) = \left\|\mathbf{x}_i + \mathbf{d}_{l_i} - \mathbf{x}_j - \mathbf{d}_{l_j}\right\|$. We illustrate in Fig 3.2 the way we compute the unary potentials (Fig 3.2(a)) and the pairwise potentials (Fig 3.2(b)). On one hand, equation (3.23) expresses the approximation that we make with respect to the data term computations. The Voronoi cell associated with the control point $\mathbf{x}_i$ displaced by $\mathbf{d}_{l_i}$ is estimated by translating the cell $\mathcal{O}_i$ by the same displacement vector, which is noted as $\mathcal{O}_i + \mathbf{d}_{l_i}$. Intuitively, such an approximation of a higher order term generates small inaccuracies if the displacement range is relatively small. On the other hand, equation (3.24) uses an approximation of the higher order quantity $\widetilde{d}$. Indeed, when estimating the costs of pairs of labels, the shape scale is supposed to be fixed at a value estimated from the initial attributes $(\mathbf{x}_1^0, \ldots, \mathbf{x}_n^0)$. This assumption is also acceptable in the case of small displacements. As we shall see next, starting from an initial shape position, we iteratively solve sequentially several labeling problems based on the formulation in (3.22), where the solution of one step initializes the next iteration. Although the shape scale is fixed within each iteration, it is updated from one step to the other, and the trade-off between the regional image information and the shape prior allows for scale changes. The resulting approach hence deals with translations, rotations and scaling without explicitly modeling them. Let us now detail the iterative search scheme that we propose.

The cardinality of the label set is quite important since on one hand it defines the accuracy of the search, while on the other hand increases the complexity of the algorithm. In order to address these issues, we first consider an approach that is incremental in terms of displacements. To this end, we cope with the accuracy issue, that is closely related to the quantization of $\mathcal{D}$, by using a pyramidal coarse-to-fine approach. Each level of the pyramid corresponds to a quantization step that is refined in the following level. To speed up the convergence in each level of the pyramid, and be consistent with the local displacements assumptions that we mentioned earlier, we also adopt an incremental approach in terms of the label set, where in each iteration t we look for the set of labels that will improve the current solution. This label set refers to a local search neighborhood. More explicitly, we repeatedly



(a) (b)

Figure 3.2: Computing the energy terms: grey squares refer to the image grid, and numbered disks refer to the control points possible labels. The label 1 refers to the current landmark position, i.e. a null displacement. (a) Unary data term: the cost of the label 4 of the landmark $\mathbf{x}_i$ is evaluated by translating the corresponding Voronoi cell $\mathcal{O}_i$ (depicted in red) by $\mathbf{d}_4$ and then computing the integral in (3.23). The translated cell $\mathcal{O}_i + \mathbf{d}_4$ is shown in green. (b) Pairwise shape term: computing the cost $V_{ij}(4,2)$ is done by applying the displacements $\mathbf{d}_4$ and $\mathbf{d}_2$ respectively to $\mathbf{x}_i$ and $\mathbf{x}_j$ and then estimating the resulting distance $d_{ij}(4,2)$.

minimize:

$$E^t\left(l_1(t),\ldots,l_n(t)\right) = \alpha \sum_{i=1}^{n} V_i(\mathbf{x}_i^{t-1}, l_i(t)) + \sum_{(i,j)\in\mathcal{E}_k} V_{ij}(\mathbf{x}_i^{t-1}, \mathbf{x}_j^{t-1}, l_i(t), l_j(t)) \ , \tag{3.25}$$

with

$$\mathbf{x}_i^t = \mathbf{x}_i^0 + \sum_{\tau=1}^{t} \mathbf{d}_{\mathbf{l}_i(\tau)} \ , \tag{3.26}$$

until convergence, with $\mathbf{l}_i(\tau)$ being the optimal label associated with the i^{th} control point at iteration τ.

Recovering the optimal solution of this objective function is known to be an NP-hard problem and the complexity is dependent mostly on the pairwise potentials function, that is not submodular in our case. Hence, we consider an approximate solution to the labeling problem. We presented in section 3.1.1 several methods that permit to tackle the minimization problem, such as α-expansion [Boykov *et al.* 2001], FastPD, TRW-S [Kolmogorov 2006] and loopy belief propagation (LBP). We used in practice the computationally

efficient Fast-PD algorithm [Komodakis *et al.* 2008b] [2]. The resulting method is sketched in algorithm 2.

Algorithm 2 Region-based Segmentation

Initialize the shape at $(\mathbf{x}_1^0, \ldots, \mathbf{x}_n^0)$.
for $\pi = 1$ to π_{max} **do**
  $t = 0$.
  if $\pi > 1$ **then**
    Update $(\mathbf{x}_1^0, \ldots, \mathbf{x}_n^0)$ to the result of the previous pyramid level.
  end if
  repeat
    Update $\widetilde{d}$.
    Update the object boundary (if relevant), using interpolation.
    Compute the Voronoi diagram of the landmarks $\mathbf{x}_i^t$. Deduce the Voronoi object cells $\mathcal{O}_i$.
    Compute the energy terms: the unary potentials and the pairwise potentials.
    Run FastPD to solve (3.25).
    Update the control points according the the optimal labeling, $t \leftarrow t + 1$.
  until Convergence: the labeling stabilizes.
end for

Implementation details The algorithm we propose uses a local search method. Hence in principle, the initialization is required to be quite close to the solution. However, we derived a method to deal with an initialization that is relatively far from the object of interest. We consider a subset of a few control points, that will be used to roughly locate the object by estimating similarity transformations (typically 3 points in 2D, 4 points in 3D).These control points are connected in a complete clique. We apply the same procedure described in algorithm 2 using these landmarks, without restricting the search to a small neighborhood. However, to prevent the distortion of the shape, we project the transformation resulting from each inner loop on

[2] An implementation is available online: http://www.csd.uoc.gr/~komod/FastPD/

the space of similarity transformations, in order to update the landmarks positions. Then, the shape boundary is interpolated using thin plate splines (TPS). After this initialization step, the model is close to the solution, and we apply algorithm 2 using our shape model. An example of this process is shown in Fig .3.3.

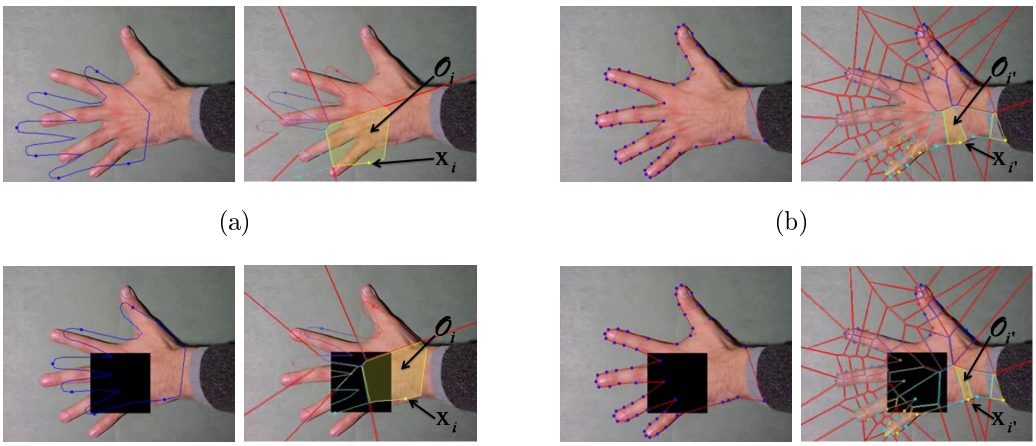


Figure 3.3: Segmentation algorithm inner loop: pairs of images showing the evolving contour (left image of the pair) and the Voronoi diagram (right image of the pair). An "object" Voronoi cell is emphasized in yellow. Two cases: a simple case (a) & (b) and a case with missing parts (c) & (d). (a) & (c) Initialization: only 6 landmarks are used. The resulting transformation is projected on the similarity transformations space. (b) &(d) Convergence using 56 points.

We also note that our method does not impose any explicit constraints with respect to the local transformations that the shape model undergoes during the iterations. While this leads to a simple formulation without any need to quantize rotations for instance, the resulting formulation does not prevent control points displacements that would cause boundary self-intersection or surface flipping. However, thanks to the geometric properties of our shape model, such an undesired behavior should be unlikely, which is verified in practice. Nevertheless, due to our construction, our algorithm is not suitable to handle self-occlusions of an articulated object, as permitted in the layered

pictorial structure [Kumar *et al.* 2010] for instance.

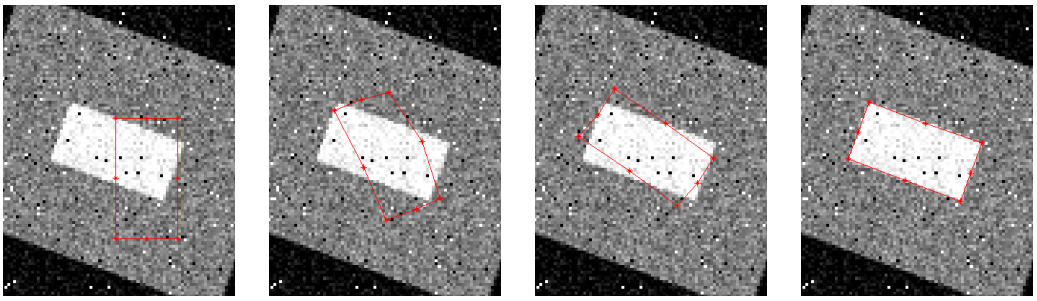
Another observation concerns the energy minimization with respect to the iteration scheme. Due to the approximation made at the level of the data term, the estimated energy within the FastPD algorithm will differ from that computed once an optimal labeling is found. Hence, we carefully controlled these two quantities in order to ensure that the energy decreases monotonically. Any violation of this condition leads to the termination of the algorithm.

3.2.4 Experimental Validation

After preliminary testing on synthetic data, we validated our method by segmenting hands in 2D images, and segmenting the left ventricle in 3D computed tomography (CT) images.

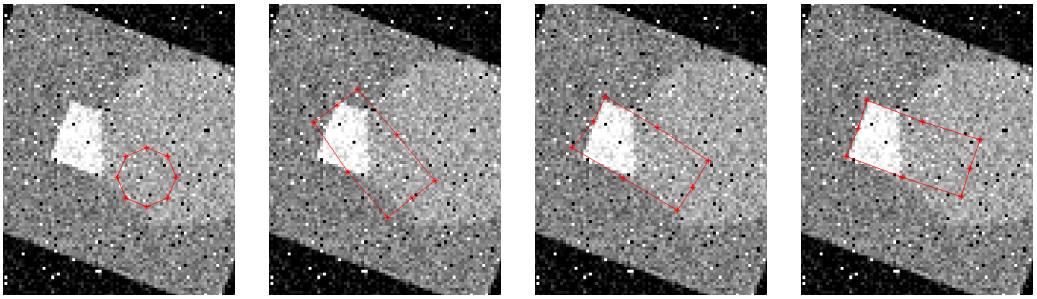
Synthetic data As a first step to assess the performance of the proposed region-based segmentation method, we tested our algorithm on synthetic images that are corrupted by Gaussian and salt and pepper noise, and where the object is occluded. In these preliminary experiments, we considered the simple case of a white rectangle on a black background. We modeled this object using 8 landmarks, and considered a complete graph in this case. Gaussian probability distributions were used to model the pairwise distances, as well as the object and background intensities. Moreover, the scale of the object was supposed to be known in these experiments. Fig. 3.4 shows the result of a simple case. Although rotations are not modeled within our search algorithm, the shape model is able to recover the rotated object through local landmark displacements. A more difficult example in shown in Fig. 3.5 where the object is partially occluded. This is also handled by our method naturally, without a dedicated modeling. Fig. 3.6 simulates the tracking of a moving object that is occluded during its displacement. No dynamic prior has been used.

Hand segmentation In a subsequent experiment to validate the performance of our method, we considered the application of modeling the hand using a 2D 40-example dataset of annotated left hands, showing different relative finger positions, hand sizes, and textures [Stegmann & Gomez 2002]. On each hand contour, 56 landmarks were used to describe the structure. We per-



(a) (b) (c) (d)

Figure 3.4: Easy synthetic case. (a) Initialization. (b)–(c) Intermediate inner-iterations. (d) Convergence.



(a) (b) (c) (d)

Figure 3.5: Synthetic case with occlusion. (a) Initialization. (b)–(c) Intermediate inner-iterations. (d) Convergence.

formed clustering in the diffusion maps space as described in section 2.2.4, using shape maps [Langs & Paragios 2008]. The clustering provided 11 clusters shown in Fig. 2.5(c). The constructed model was used as a shape constraint as shown in Fig. 2.6(a)-2.6(d), and we hereafter apply it in a segmentation setting. We considered a multi-scale implementation of the approach using a gradually increasing number of control points to accelerate convergence, and handle non-local initialization, as explained in the previous section. Fig 3.3 shows the two different stages of the algorithm: the initialization phase and the search phase. It also emphasizes the object Voronoi cell of a control point in each phase. The proposed technique allowed to segmented correctly 37 out of the 40 examples of the database. Examples of the results we obtained are

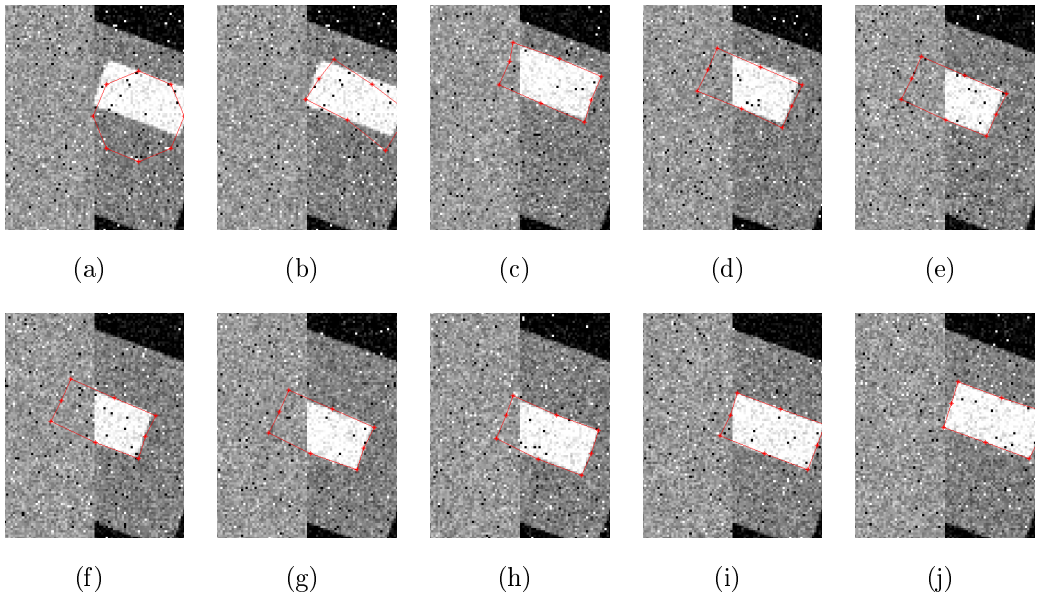


Figure 3.6: Synthetic tracking with occlusion. (a) Initialization. (b)–(j) Results on subsequent frames.

shown in Fig. 3.8(a). We also compared quantitatively our method to AAM segmentation Fig. 3.7 based on the dice overlap coefficient with respect to the ground truth data. We can see in these boxplots that our algorithm performs better quantitatively with examples where the forearm is hidden by a sleeve. In the case of bare forearms, the data term drives the model to "oversegment" the hand in comparison with the ground truth, which explains our results. These "oversegmentations" are visually correct (especially the fingers are correctly segmented) as we can see in Fig. 3.8(a). The three examples where our method did not succeed are particularly difficult to handle for our algorithm because they exhibit occlusion between the fingers, which can cause folding in the evolving contour. The first two segmentation examples shown in Fig. 3.8(b) demonstrate that our method can sometimes handle such difficulties, but this is not guaranteed. Towards checking the robustness of the method with respect to occlusions, we removed some hands parts for several examples, and despite the important missing structure, the results were quite satisfactory as shown in Fig. 3.8(b). The weight of the prior in these cases was increased, and enforced the correct segmentation, as the data term was

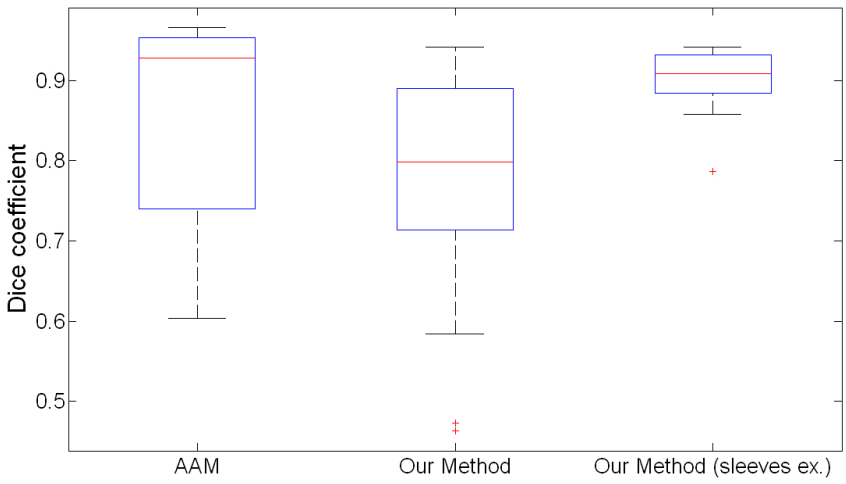
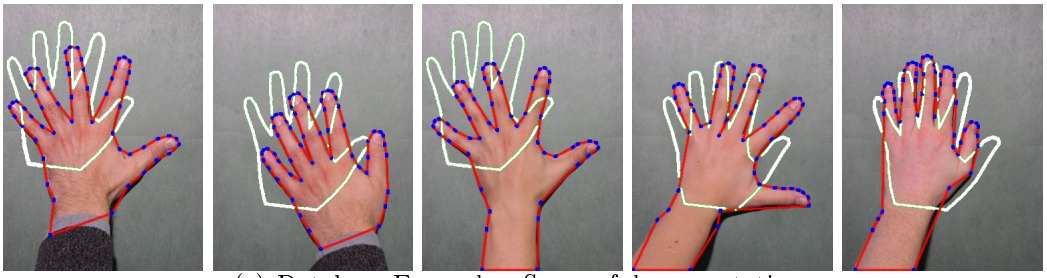


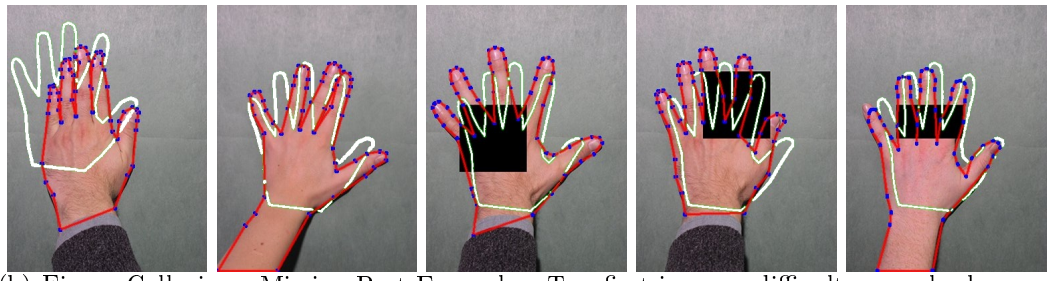
Figure 3.7: Boxplots of dice overlap coefficients comparing our method to AAM.

less reliable. To further assess the robustness of our method, we added severe Gaussian noise to the database images. The segmentations obtained in Fig. 3.8(a) are completely or almost recovered, thanks to the prior knowledge, as shown in Fig. 3.8(c). Eventually, we used our segmentation method in a real setting, on hand video frames, with a cluttered background and partial occlusion cases. Fig. 3.8(d) gives some examples of the obtained segmentations. We could reproduce the result we obtained on the noisy images using AAM segmentation [Cootes *et al.* 2001], but this algorithm could not reproduce our results for the occlusion cases. [Gross *et al.* 2006] developed an AAM algorithm that addresses the occlusion problem. Note that this is a dedicated development of the standard AAM, whereas our method naturally performs well for the segmentation of a partially occluded object.

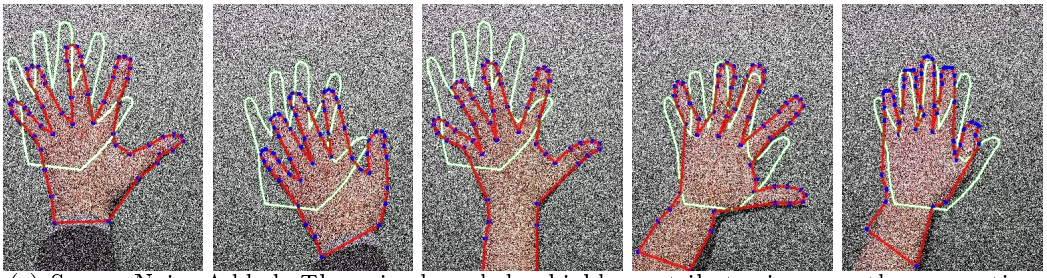
Left Ventricle segmentation In this experiment, we did not use our k-fan graph model. Instead, we considered the complete graph to enforce shape prior. In order to evaluate the performance of our method in the three-dimensional medical image analysis setting, we used a dataset of 28 computed tomography (CT) cardiac volumes, having an approximate mean size of $512 \times 512 \times 250$ voxels, where the voxel size is about $0.35 \times 0.35 \times 0.35\text{mm}^3$.



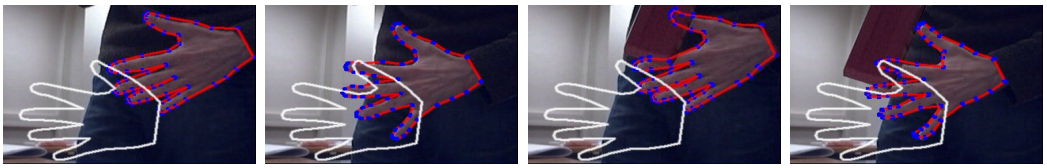
(a) Database Examples: Successful segmentations



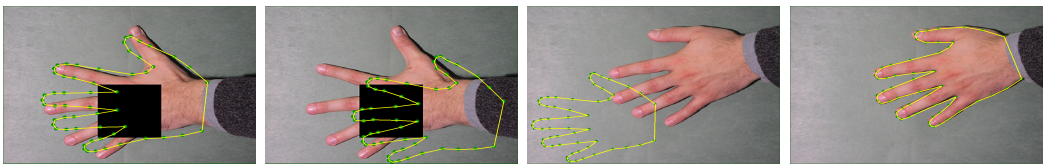
(b) Finger Collusion - Missing Part Examples. Two first images: difficult examples because of fingers collusions. Three last images: segmentation of hands with missing parts.



(c) Severe Noise Added: The prior knowledge highly contributes in correctly segmenting very noisy images.



(d) Video Frames - Partial Occlusions. Real video frames: cluttered background and occlusions.



(e) AAM results: succeeds with the learning examples but fails with occlusions. Initialization on the left - result on the right.

Figure 3.8: Model-based segmentation of the hand. Initialization is shown in white, segmentation in red, and the final control points positions in blue.

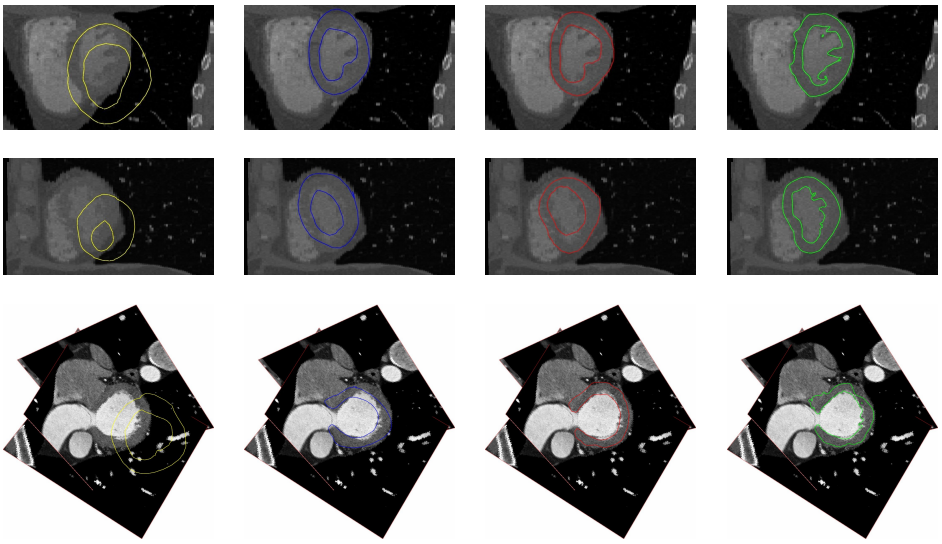
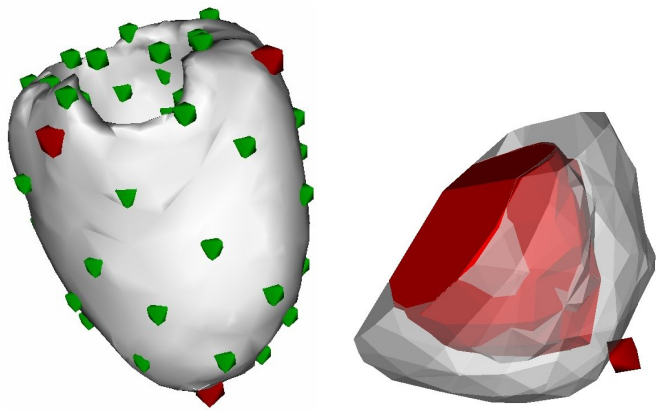


Figure 3.9: Segmentation results: 3 "unseen" examples (do not belong to the training set). Initialization in yellow, shape after affine transform in blue, final segmentation after TPS deformation in red. Random Walker result in green.

While no expert manual segmentation was available for this dataset, we could compare to the results given by the Random Walker algorithm described in [Grady 2006][3]. First, to build the model, we selected randomly 11 CT images as a training set. We placed manually on the surface of the left ventricle of one example from the training set 90 control points, that we will call P_{90}. The remaining 10 examples of the training set were then registered to the labeled example using the method described in [Taron *et al.* 2009] and correspondences between the control points instances were consequently deduced. After smoothing the segmentation mask obtained by [Grady 2006] on the labeled example, we generated a meshed surface $\mathcal{S}$ of the myocardium and the blood pool. We learned next the probability density distributions p_{ij} of the normalized chord lengths (section 2.2). The grey levels of the myocardium, the blood pool and the background were also learned from the training set to Gaussian probability distributions $p_{\mathcal{O}^{myo}}$, $p_{\mathcal{O}^{bp}}$ and $p_{\mathcal{B}}$ respectively (section. 3.2.1). By intersecting $\mathcal{S}$ with the Voronoi diagram of the set of control

[3]This dataset is courtesy of *Siemens Corporate Research*

points P_{90} we obtained the classes cells $\mathcal{O}_i^{myo}$ and $\mathcal{O}_i^{bp}$ corresponding to the myocardium and the blood pool respectively. In particular, 4 of the 90 control points are interesting as one is located in the apical area, and the others in the basal area (this set will be called P_4). Figure 3.10 shows the obtained surface $\mathcal{S}$ with the control points P_{90} and P_4, and the voronoi cells $\mathcal{O}_i^{myo}$ and $\mathcal{O}_i^{bp}$ of the apical control point. The set P_4 will be used in the initialization step, as explained in the implementation details in paragraph 3.2.3.



(a) (b)

Figure 3.10: Our model: a deformable shape associated with control points. (a) In red: control points P_4(in the apical and basal parts) used to define affine transforms of the shape. In green: the set of control points P_{90} that defines the TPS deformation. (b) The apical control point with the associated Voronoi cell, intersected with the blood pool and the myocardium.

As a first application, we tested the consistency of the learned prior by only minimizing the shape term of the designed energy, and canceling the data term, starting from randomly perturbed positions of the control points. The prior constraints made the initial shape converge in all cases to the "mean" shape. Next, the segmentation was performed following algorithm 2 by following two steps: (i) the surface $\mathcal{S}$ is deformed from its initial position by using the control points P_4, and by applying a similarity transformation to the mesh after each iteration. This is repeated till convergence. (ii) The control points P_{90} are introduced and the mesh is transformed using a Thin Plate Spline (TPS) [Bookstein 1989] deformation. For the segmentation experiments, the

Table 3.1: Comparison of our method with the Random Walker Algorithm

Correctly Segmented Voxels (True Positives)	Incorrectly Segmented Voxels (False Positives)	Segmentation Time (3 GHz, 3 GB)
85.12%±7.3	15.3%±10.2	90s±10

axial data slices were downsized to 128×128 pixels.

During implementation, we noticed that computing the Voronoi diagram in 3D, and performing the intersections of the resulting cells with the myocardial and blood pool surfaces was time consuming. Hence, in order to increase the computational efficiency, we introduced an additional approximation. In addition to the control points and the meshed surface, we computed for each landmark $\mathbf{x}_i$ the voronoi cells $\mathcal{O}_i^{myo}$ and $\mathcal{O}_i^{bp}$ offline. Then after each inner-iteration of the algorithm, these cells were updated similarly to the mesh using TPS. Thereby, complex geometric computations were no longer required.

We compared our results quantitatively for the whole dataset to [Grady 2006] and compiled them in Tab. 3.1.

3.3 Detector-based Segmentation

We presented in section 3.2 a region-based shape-driven segmentation algorithm and demonstrated its performance in a variety of settings. However, regional image cues are not always suitable in practice. An approach that uses such image features makes the implicit assumption that image regions are smooth or have homogeneous intensity distributions, at least locally. In applications where objects have a particularly textured appearance, this hypothesis does not hold, and region-based statistics would lead to a poor segmentation performance. For instance, Fig 3.11 shows posterior-anterior chest radiographs (PA X-Rays). This is an example of images that can be hardly segmented using region-based costs. Moreover, in addition to shape variations, object appearance can dramatically differ from one case to another. This is also the case in computer vision and in particular in video-based surveillance where segmenting and tracking people can be very difficult as human body appearance is almost arbitrary, depending on the clothes that are worn. As an

Figure 3.11: Example of posterior-anterior chest X-Rays with manual expert segmentations.

alternative to image intensity distributions, one can consider feature vectors extracted from the image. Such high dimensional descriptors benefit from the neighboring data around a particular image location, and can exhibit strong discriminative power. In our approach, we seek optimal landmark positioning. Therefore, a discriminative image support that isolates highly probable control point locations in the image is of a great interest. Indeed, combined with a shape prior, feature-based detectors can yield very powerful segmentation tools. [Seghers *et al.* 2007a] pursue this line of thought as they use a chain-structured shape prior to find the best landmark positions among candidates extracted from the image, based on locally orderless image (LOI) features. In the same spirit, [Donner *et al.* 2010] focus on symmetry interest points detected using GVF-based cues, and enforce a shape prior through a MRF formulation. Similarly, we propose to build image classifiers that discriminate the control points from the background. These classifiers provide landmark candidates, with associated confidence values. The segmentation task boils down to finding the trade-off between shape prior penalties, and candidate selection costs, in an integrated manner. In practice, we adopt a similar approach to the LOI for feature extraction. Then, we apply the Adaboost algorithm to the computed local descriptors in order to train the landmark classifiers. We observed in our experiments that in some difficult cases, the classifiers may fail to provide good candidates to the segmentation algorithm, which leads to a suboptimal result. We propose to account for the missing correspondence problem by incorporating an additional label in the segmentation formulation to account for outliers. Although this idea has already been mentioned in [Donner *et al.* 2007a] and [Seghers *et al.* 2008], the authors do not provide a clear method to post-process the outlier landmarks, and the impact of the procedure is not shown in the experiments. Our approach is detailed in the

next section, and the results that we obtain suggest that accounting for the outliers in the segmentation process can lead to improved results qualitatively and quantitatively in the case of the complete graph.

The remainder of this section is organized as follows: first, we describe the features that we use to build the control points descriptors and train the corresponding classifiers. Then, we explain how we embed this information in our segmentation framework. In the next subsection, we propose a method to account for outliers during segmentation, and "repair" their positions in a post-processing step. The last part is dedicated to the segmentation results we obtained on PA chest X-ray images. We show results for both cases: the case without handling outliers, and the case where outliers are accounted for.

3.3.1 Landmark-based Image support

We will explain in the following how we learn a classifier for each control point of the object, using features that are computed from the image. These trained classifiers are then used as detectors in our algorithm, to segment unseen images. Our method consists of extracting from the image a set of candidates for each landmark, and then choosing the best ones in terms of detection response on one hand, and global fidelity to the learned shape prior on the other hand.

Boosted Landmark Detectors

We present our method by describing the features and the classifiers that we use in practice. However, the framework is flexible, and different types of detectors can be plugged in, using for instance the gradient vector flow (GVF) descriptors that were considered in [Donner *et al.* 2007a], or the locally orderless image (LOI) features that were adopted in [Seghers *et al.* 2007a]. The recent boundary detection framework introduced in [Kokkinos 2010] can also represent an alternative choice to define the image features. The author considered the distribution of image gradient orientation in the neighborhood of the landmarks at multiple scales through Scale-Invariant Feature Transform (SIFT) descriptors [Lowe 2004], and then performed discriminative dimensionality reduction using the Sliced Average Variance Estimation algorithm

(SAVE) [Cook & Lee 1999].

In our work, we first process the image $\mathcal{I}$ using a filter bank of derivatives for different Gaussian smoothing values σ ($\sigma \in \{0.5, 1, 2, 4, 8\}$). The output of this process is a feature image $\mathcal{F}$ as showed in Fig. 3.12. Such a feature extraction was previously considered by [van Ginneken *et al.* 2002] and [Seghers *et al.* 2007a]. Then, they obtained locally orderless image (LOI) descriptors [Koenderink & Van Doorn 1999] by computing the first statistical moments of $\mathcal{F}$ in local neighborhoods. [Seghers *et al.* 2007a] modeled the feature vector probability function using a multivariate Gaussian distribution for each blurring level σ. Our derivation is slightly different as we build for each control point of the object a "boosted" classifier $f_T^{(i)}$, using Adaboost [Freund & Schapire 1996]. To this end, at each control point location $\mathbf{x}_i$, we extract a patch of size 17×17 from the feature image $\mathcal{F}$, amounting to a K-dimensional descriptor φ_i (with K =8959). Hence, given a set of N_p training images, where the positions of the landmarks are known, we obtain for each control point $\mathbf{x}_i$ a set of "positive" examples of size N_p. We also extract N_n feature vectors from the background to form a set of "negative" examples. Then, using the training set of $N = N_p + N_n$ examples $(\varphi_j^{(i)}, y_j^{(i)})_{1 \leq i \leq N}$, with the class label $y_j^{(i)} \in \{-1, 1\}$ referring to the background/object, we train the classifiers $f_T^{(i)}$ in order to discriminate the control points from their neighborhoods. Before explaining the relationship between these classifiers and the data-related energy term, we describe briefly the Boosting framework.

Boosting is a classification method that is based on machine learning. It uses labeled data to learn a decision function. Let our training data be a set of N pairs $(\varphi_i, y_i) \in \Phi \times \{-1, 1\}$, $1 \leq i \leq N$, where Φ is the feature vector space and y_i are the binary class labels. Boosting aims to find a highly accurate decision rule f_T as a linear combination of many "weak" or basic hypotheses h_t, or:

$$f_T(\varphi) = \sum_{i=1}^{T} \alpha_t h_t(\varphi) \quad . \tag{3.27}$$

These simple decisions are usually easy to train, but have quite poor individual performance. They are "boosted" by combining them, which yields a "strong" classifier f_T. Boosting works iteratively by trying to improve the function f_T at each step. During each iteration, the best weak classifier h_t is added with

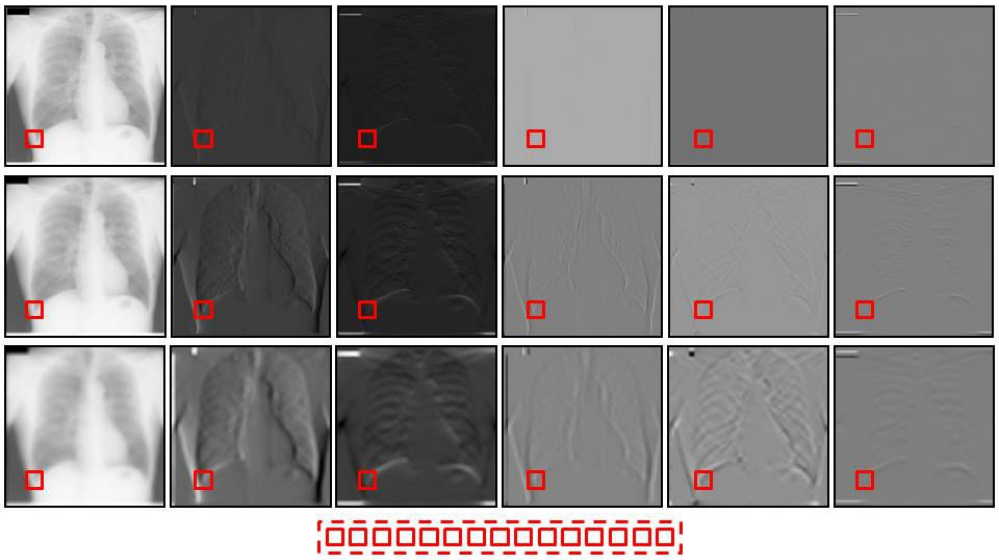


Figure 3.12: Features computation using a filter bank.

an appropriately chosen step size α_t, such that a cost function that compares the classifier's responses to the data labels is minimized. The cost function $L(f_T)$ is typically a sum of loss functions or:

$$L\left(f_T\right) = \sum_{i=1}^{N} \lambda\left(f_T\left(\varphi_i\right), y_i\right) \quad . \tag{3.28}$$

There is an extensive literature on Boosting, and several reviews on the subject exist. Among others we refer to [Schapire 2002] and [Meir & Rätsch 2003]. Different Boosting algorithms had been proposed. Their differences lie in the definition of the cost function, the selection of the weak classifiers and the computation of the step size. We restrict our description to the algorithm that we used in the experiments.

Adaboost A well-known Boosting algorithm is Adaboost which was introduced by [Freund & Schapire 1996]. This machine learning algorithm was then generalized by [Schapire & Singer 1998]. In this case an exponential loss function is used, or $\lambda\left(f_T\left(\varphi_i\right), y_i\right) = \exp\left(-y_i f_T\left(\varphi_i\right)\right)$. Essentially, at each Boosting step a weak classifier that best discriminates "positives" and "negatives" is learned from weighted training data, and a corresponding step size is

determined. Then, poorly classified training examples are identified. Subsequently, their weights are increased in order to emphasize their importance in the next step. A typical choice for the weak learner family is decision stumps. Algorithm 3 summarizes these different steps.

Algorithm 3 Adaboost

1: **Input:** N training examples $\{(\varphi_1, y_1), \ldots, (\varphi_N, y_N)\}$ where $\varphi_i \in \Phi$ and $y_i \in \{1, -1\}$.
2: **Initialize:** $D_1(i) = \frac{1}{N}$.
3: **for** $t = 1 \ldots T$ **do**
4: **Training:** Learn weak classifier $h_t : X \to \mathbb{R}$ using distribution D_t.
5: **Compute the weight:** $\alpha_t \in \mathbb{R}$.
6: **Update:**

$$D_{t+1}(i) = \frac{D_t(i) \exp\left(-\alpha_t y_i h_t(\varphi_i)\right)}{Z_t} \quad ,$$

where Z_t is a normalization factor.
7: **end for**
8: **Output:** The strong classifier:

$$\begin{aligned} H(\varphi) &= sign\left(f_T(\varphi)\right) \quad , \\ f_T(\varphi) &= \sum_{t=1}^{T} \alpha_t h_t(\varphi) \quad . \end{aligned}$$

Let us now summarize the learning phase. For each landmark $\mathbf{x}_i$, "positive" examples and "negative" examples are extracted from the training images to form a learning set $(\varphi_j^{(i)}, y_j^{(i)})_{1 \leq j \leq N}$. The latter is used to train an Adaboost classifier $f_T^{(i)}(\varphi) = \sum_{t=1}^{T} \alpha_t^{(i)} h_t^{(i)}(\varphi)$. The obtained classifier responses are incorporated into a prior-based segmentation algorithm as explained in the following.

Data Energy Term

We adopt the same Bayesian approach presented in section 3.2 to formulate the segmentation task. The equivalent MAP estimation problem is defined in equation (3.13) as minimizing:

$$E(\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathcal{I}) = E_{shape}(\mathbf{x}_1, \ldots, \mathbf{x}_n) + E_{image}(\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathcal{I}) \quad , \qquad (3.29)$$

with respect to $(\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \Omega^n$, Ω being the image domain. Recall that the prior term E_{shape} refers to our k-fan shape model and is expressed in terms of pairwise landmark interactions, similarly to equation(2.30):

$$E_{shape}\left(\mathbf{x}_1, \ldots, \mathbf{x}_n\right) = -\sum_{(i,j)\in\mathcal{E}_k} \gamma_{ij} \log\left(p\left(\frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{\widetilde{d}}\right)\right) \quad . \tag{3.30}$$

The energy term E_{image} is based on the image cues, and is defined thanks to the learned classifiers. Using a logistic function, we turn the strong classifier response into a probability measure as $p_i\left(f_T^{(i)}\left(\varphi\right)\right) = \frac{1}{1+\exp\left(-f_T^{(i)}(\varphi)\right)}$. Then, by assuming independence of the feature vectors at different image locations, we can write:

$$\begin{aligned} E_{image}\left(\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathcal{I}\right) &= \sum_{i=1}^{n} -\log\left(p_i\left(f_T^{(i)}\left(\varphi\left(\mathbf{x}_i\right)\right)\right)\right) \\ &= \sum_{i=1}^{n} -\log\left(\frac{1}{1+\exp\left(-f_T^{(i)}\left(\varphi\left(\mathbf{x}_i\right)\right)\right)}\right) \quad , \end{aligned} \tag{3.31}$$

where $\varphi\left(\mathbf{x}_i\right)$ is the feature vector at the landmark $\mathbf{x}_i$ location. Let us consider now that each position in the image domain Ω corresponds to a label $l \in \mathcal{L}$, $\mathcal{L}$ being the labeling set. We will note $\mathbf{x}_i(l)$ placing the landmark $\mathbf{x}_i$ at the location labeled by l. Then, minimizing the energy (3.29) is equivalent to solving the labeling problem:

$$\min_{\substack{l_i\in\mathcal{L} \\ 1\leq i\leq n}} E\left(l_1 \ldots, l_n\right) = \alpha \sum_{i=1}^{n} V_i(l_i) + \sum_{(i,j)\in\mathcal{E}_k} V_{ij}(l_i, l_j) \quad , \tag{3.32}$$

where the unary and pairwise potentials are defined as:

$$\begin{aligned} V_i(l_i) &= -\log\left(p_i\left(f_T^{(i)}\left(\varphi\left(\mathbf{x}_i(l_i)\right)\right)\right)\right) \\ V_{ij}(l_i, l_j) &= -\gamma_{ij} \log\left(p\left(\frac{\|\mathbf{x}_i\left(l_i\right) - \mathbf{x}_j(l_j)\|}{\widetilde{d}}\right)\right) \quad . \end{aligned} \tag{3.33}$$

However, searching in the whole image domain Ω increases the complexity of the problem. Therefore, we restrict the search in practice to the image locations where the classifiers provide their highest m responses (for each landmark), m being a parameter that is fixed by the user. Hence, the problem

boils down to finding the best configuration among the top m control points candidates. For efficiency, we used an approximate optimization methods to solve (3.29), namely the sequential tree-reweighted message passing algorithm (TRW-S) [Kolmogorov 2006]. In TRW-S, a lower bound on the optimal energy is maximized. We observed in our experiments the lower bound value was very often equal to the graph energy at the end of the optimization, meaning that a global optimum was found. The results we obtained are presented in the next section. Considering a relatively small number of candidate points allows solving the problem efficiently. The results can be however let down if there are no good landmarks in the candidate pool. We address this limitation in the following.

Segmentation Outliers

The above formulation can be in some difficult cases insufficient to obtain satisfactory segmentations because of the limitation of the detected candidates. The proposed algorithm relies heavily on the quality of the detectors and the provided costs, and hence a poor generalization performance of the classifier may affect the whole segmentation process. To account for the missing correspondences problem, we consider the idea of an additional artificial candidate per landmark, as suggested in [Donner *et al.* 2007a]. The $m + 1^{th}$ candidate will refer to finding an outlier. The critical aspect is then to define the corresponding unary and pairwise costs. [Donner *et al.* 2007a] define these costs as being proportional to the mean of the one computed for the real candidates. On one hand, this cost choice introduces an additional parameter (of proportionality) that needs to be adjusted. On the other hand, using the current test image to compute these costs may be unsuitable as these computations will be dependent on the number of candidates. A different number of candidates may result in different costs, and hence different outliers detection. Moreover, the generalization performance of the classifier will have a direct impact on the outlier cost in this case. An alternative way to define these costs relies on their statistical estimation from the training set, where the energy terms of the available solutions (segmentations) can be easily evaluated. Hence, we compute from the training set the mean and standard deviation of each unary and pairwise cost, and set the outliers penalties to the mean plus two standard

deviations.

The post-processing of the detected outlier landmarks is another important aspect when handling such difficult cases. Hence, one has to define a strategy to "repair" the outlier landmark positions, and obtain an allowable segmentation. It is not clear in [Donner *et al.* 2007a] how this step is accounted for. [Seghers *et al.* 2008] proposed to relax the positions of landmarks that have high shape energy terms to lower shape energy locations. However, the used method is not clearly explained. We provide now a way to solve for the missing outlier positions.

Augmenting the candidate set by an outlier label provides a partial segmentation of the object. In practice, after minimizing (3.32), most of the landmarks positions are determined, and only a few of them are set to be outliers. Then, the next step consists to locate these outliers according to the shape prior, and to the set of fixed (already solved) landmarks. Let $\mathbb{O}$ be the set of outliers, and let $\mathbb{I}$ be the set of fixed landmarks. Then iteratively, we search the optimal set of displacements $\mathbf{d}_i^*,\ i \in \mathbb{O}$ of the outlier landmarks such that a shape energy is minimized. More formally, we consider the approach presented in section 2.2.6, and more precisely the equation (2.34). Let $\mathcal{D}_i = \{\mathbf{d}_{i1}, \ldots, \mathbf{d}_{ir}\},\ i \in \mathbb{O}$ be a quantization of the possible displacements $\mathbf{d}_{ii'}$. Hence, we iteratively look for:

$$\{\mathbf{d}_{kk'}^*\}_{i\in\mathbb{O}} = \underset{\mathbf{d}_{kk'}\in\mathcal{D}_k}{\operatorname{argmin}}\ E_{outliers}(\mathbf{d}_{ij})\ , \tag{3.34}$$

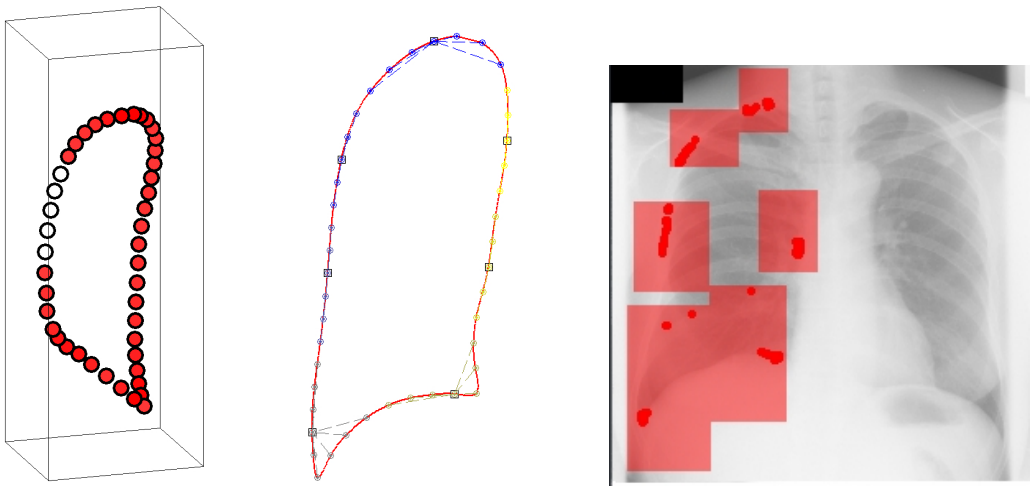
where:

$$\begin{aligned} E_{outliers}(\mathbf{d}_{kk'}) = &\sum_{\substack{(i,j)\in\mathbb{O}\times\mathbb{O}\\(i,j)\in\mathcal{E}}} -\log\Big(p_{ij}\left(\mathbf{x}_i+\mathbf{d}_{ii'},\mathbf{x}_j+\mathbf{d}_{jj'}\right)\Big) \\ &+ \sum_{\substack{(i,j)\in\mathbb{O}\times\mathbb{I}\\(i,j)\in\mathcal{E}}} -\log\Big(p_{ij}\left(\mathbf{x}_i+\mathbf{d}_{ii'},\mathbf{x}_j\right)\Big)\ . \end{aligned} \tag{3.35}$$

This final step is hence also expressed as a labeling problem. As in section 2.2.6, we used the FastPD algorithm [Komodakis *et al.* 2008b] to solve the energy (3.34).

3.3.2 Experimental Validation

We applied our method to the segmentation of the right lung in 2D radiographs. We used an available public dataset [van Ginneken *et al.* 2006][Shiraishi *et al.* 2000] of 247 images of healthy and non healthy subjects (presenting nodules). The database contains gold standard segmentations from radiologists (Fig. 3.11), that provided a delineation of the organ. Gold standard segmentation masks are hence available as well as corresponding landmark positions lying on the contour. We tested our algorithm on rescaled images of size 256×256. The database was split in two subsets (containing both healthy and non healthy cases). The first subset (corresponding to the odd indexed images) was used for training, and the second subset was used for testing. As explained previously in section 2.2, we used the annotated data in the training set to learn the structure of the graph and the prior model. We used the 44 available landmarks, and the clustering lead to the result presented in Fig. 3.13(b) which shows the 7 automatically determined clusters. The clustering was used to define the k-fan structure of



(a) (b) (c)

Figure 3.13: Modeling the right lung using the training set. (a) The obtained shape maps embedding. (b) The unsupervised linear programming-based clustering lead to 7 clusters. (c) A region of interest is used for each landmark to speed up the segmentation algorithm.

Table 3.2: Overlap coefficient and Mean Distance Error between the manual gold standard and the obtained right lung segmentation. Mean value over the testing set of 123 images and standard deviation are reported.

	Γ (%)	MDE (pixels)
Complete Graph	94.77($\pm$2.37)	1.39($\pm$0.64)
k-fan Graph	94.74($\pm$2.31)	1.40($\pm$0.62)
Complete Graph + Outliers	94.95($\pm$2.48)	1.34($\pm$0.65)
k-fan Graph + Outliers	93.87($\pm$3.74)	1.62($\pm$0.87)
MISCP algorithm [Seghers *et al.* 2007a]	93.9($\pm$3.1)	1.49($\pm$0.63)

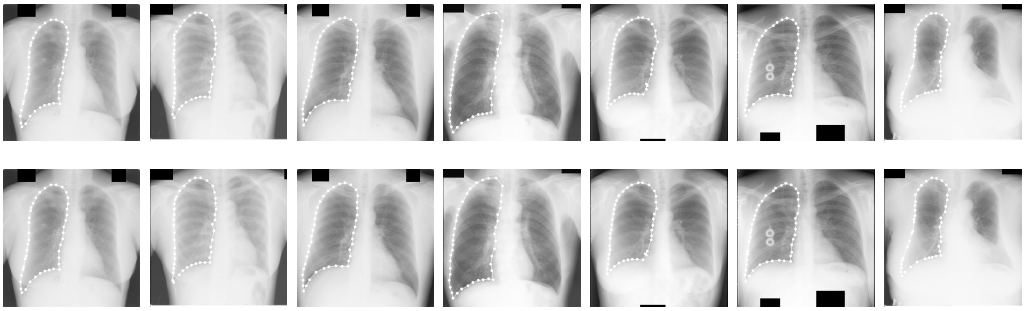
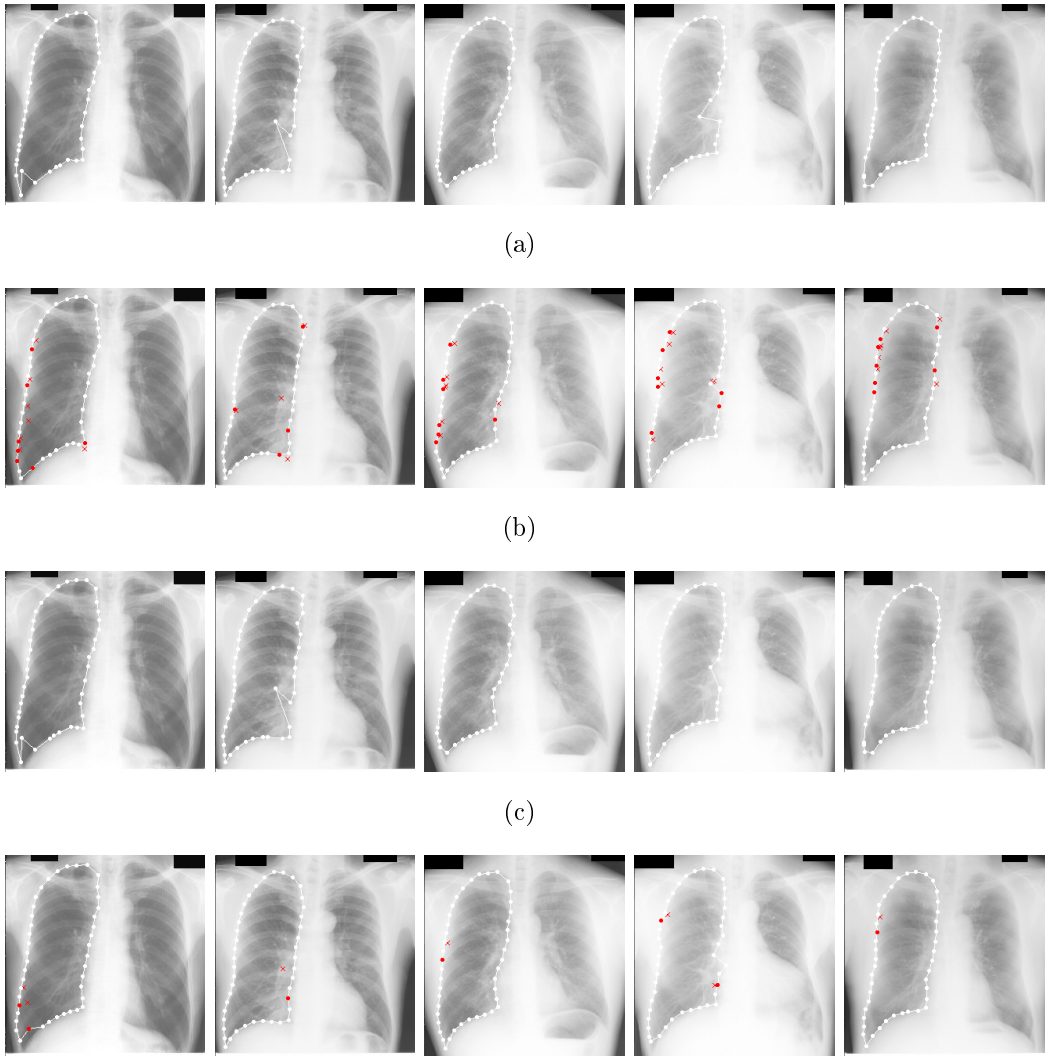


Figure 3.14: Segmentation results of the right lung, using the complete graph (top row) and the k-fan graph (bottom row).

the graph, and the pairwise normalized distance distributions were learned using Gaussian kernel density estimation. The boosted classifiers $f_T^{(i)}$ were also learned using the 124 images of the training set according to the described scheme in section 3.3.1. The Fig. 3.13(c) shows also that we used for each node regions of interests (ROIs) to speed up the candidate search step. The limits of these ROIs were learned from the training data, as done in [Seghers *et al.* 2007a]. In order to evaluate the performance of our method, we compared the segmentation results that we obtained using a complete graph against using the k-fan learned graph. In both cases, the maximum number of candidates m was set to 20. As in [van Ginneken *et al.* 2006] and [Seghers *et al.* 2007a], we used the overlap coefficient $\Gamma = \frac{TP}{TP+FP+FN}$ (where TP stands for true positive, FP for false positive, and FN for false negative) and the mean curve distance error (MDE) to assess the results

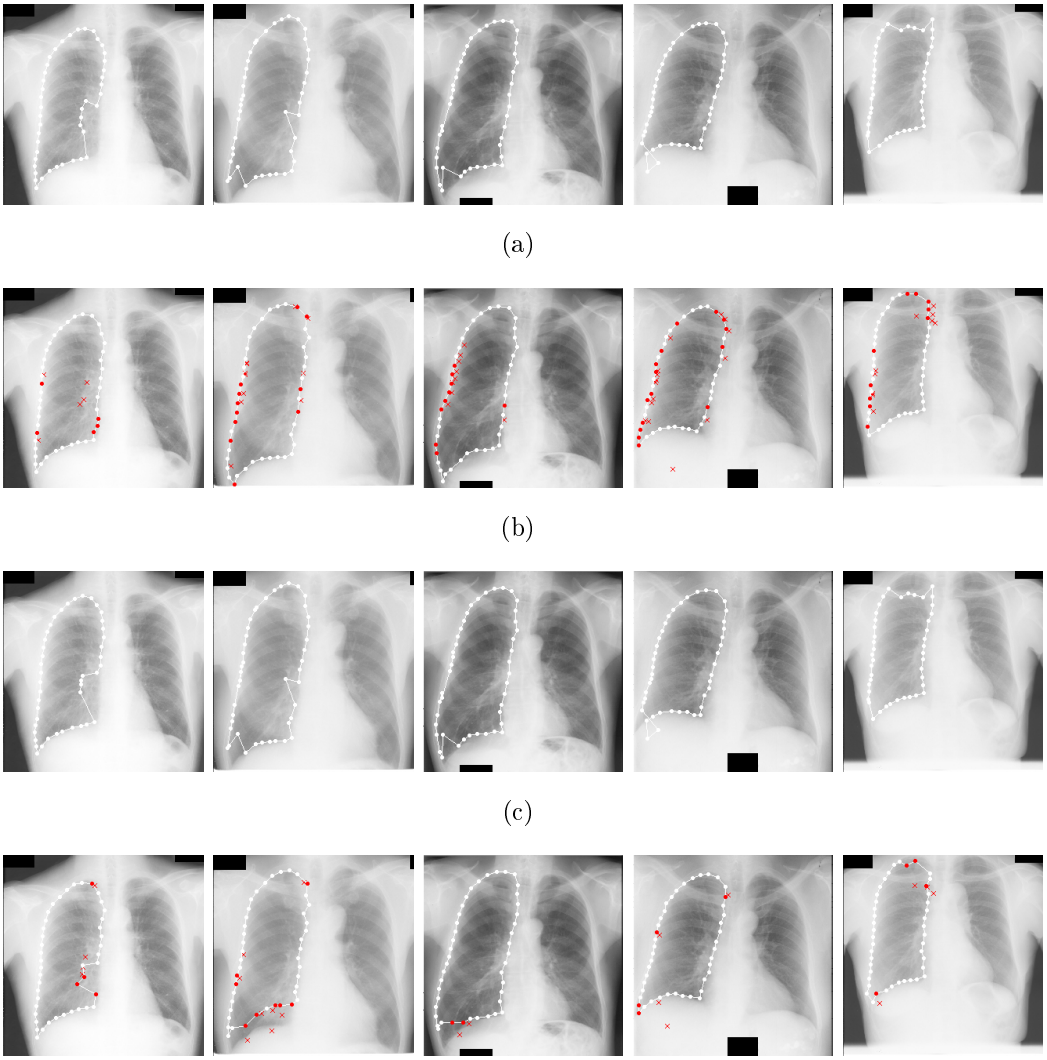
quantitatively. They are summarized in Table 3.2. We also report the results in [Seghers *et al.* 2007a], where 40 control points were used. The distributions of these quantities in the testing set are also represented by the boxplots in Fig. 3.17. Examples of the obtained segmentations are shown in



(a) (b) (c) (d)

Figure 3.15: Effect of the outlier label on the segmentation results. Outlier landmarks are in red. The crosses correspond to the top candidate position that is used as initialization for the perturbation algorithm. (a) Complete Graph (b) Complete Graph + Outliers (c) k-fan Graph (d) k-fan Graph + Outliers.

Fig. 3.14. The quantitative and qualitative results back our claims regarding the learned graph structure: we obtain a sparse model, and we maintain a performance that is equivalent to the complete graph. Moreover, as described earlier, we learned from the training set unary and pairwise costs for the



(a)

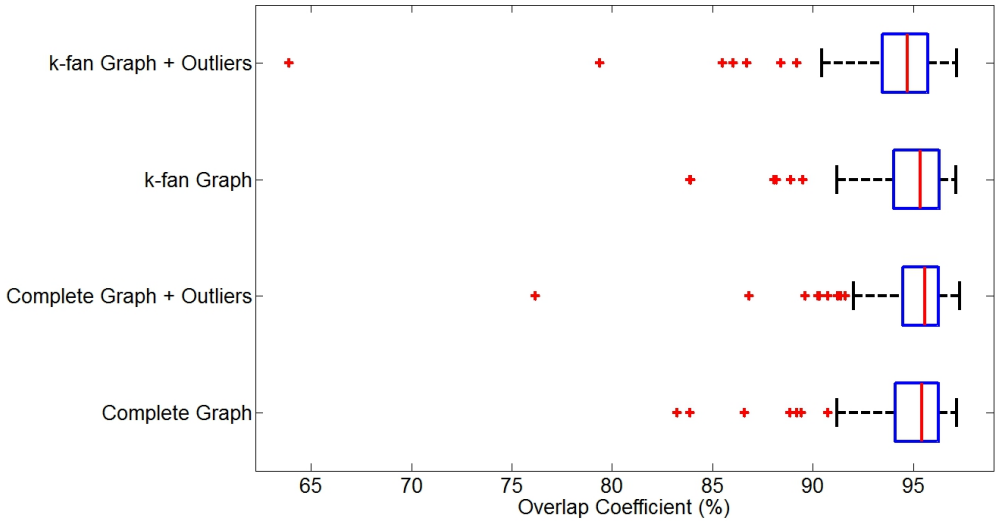
(b)

(c)

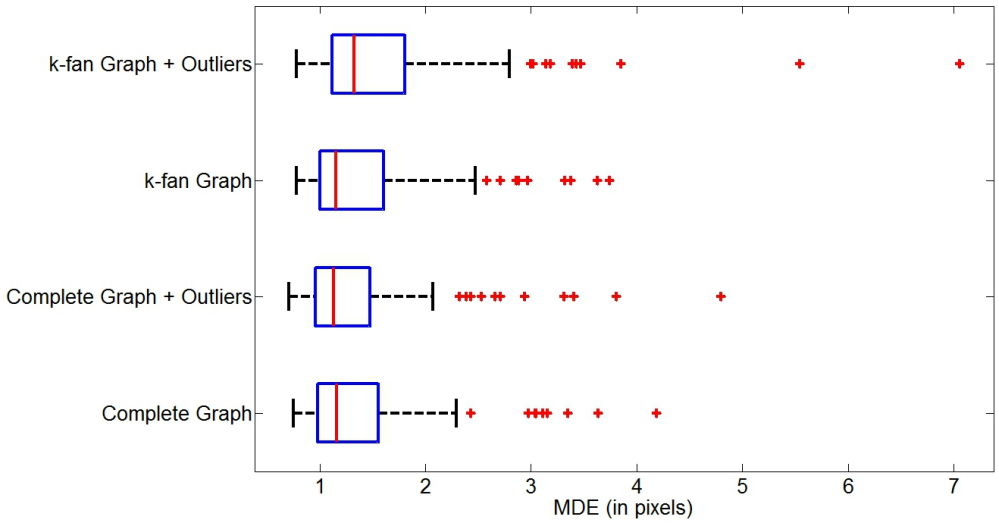
(d)

Figure 3.16: Effect of the outlier label on the segmentation results. Outlier landmarks are in red. The crosses correspond to the top candidate position that is used as initialization for the perturbation algorithm. (a) Complete Graph (b) Complete Graph + Outliers (c) k-fan Graph (d) k-fan Graph + Outliers.

outlier label, and evaluated its impact on the segmentation results. Table 3.2 and the boxplots in Fig. 3.17 show that the quantitative performance of the segmentation algorithm is improved when using the complete graph and handling outliers. We show in Fig. 3.15 and Fig. 3.16 examples of images where outliers were detected, and where the segmentation using the complete graph was improved. The examples in Fig. 3.15 also suggest that the performance of the k-fan graph can be improved by accounting for the outliers. However, as we can see in Fig. 3.16, the method performs better in the case of the complete graph. This observation is confirmed by the quantitative results in Table 3.2 and the boxplots in Fig. 3.17. They show that the computed performance indicators are slightly impaired when using outliers with the k-fan graph. This difference in the performance of the k-fan graph and the compete graph may be explained as follows: selecting a landmark as being an outlier can be thought as eliminating it from the graph, and destroying all its connections. Suppressing these geometric constraints is clearly more critical for the k-fan graph than for the complete graph, especially if landmarks of the base k-cliques are labeled as outliers.



(a)



(b)

Figure 3.17: Quantitative evaluation of the segmentation of the right lung: boxplot representation of the distributions of (a) the overlap coefficient Γ and (b) the mean curve distance error (MDE). The box lines represent the lower quartile, the median, and upper the quartile values. The whiskers extend to a range of 1.5 times the interquartile range. The values beyond this range are drawn with a "+" sign.

Chapter 4

Discussion

Introduction Statistical shape modeling is a critical task towards incorporating prior knowledge in segmentation algorithms. Not only the shape model has to represent faithfully the data, but it is also required to have a suitable representation that allows efficient inference during segmentation.

Two major trends can be distinguished from the literature on segmentation and shape priors. On the one hand, global linear models represent a very popular approach that was widely used to model implicit and explicit shapes. On the other hand, Markov random fields (MRFs) are now established as a solid framework for image segmentation, where the image is generally accounted for in a pixel-wise manner. The increasing attractiveness of this framework is also due to the big research efforts that have been made to design fast and efficient discrete optimization algorithms that provide optimality or suboptimality guarantees.

For these reasons, several research directions pointed towards combining the power of global shape models, and the flexibility and the efficiency of MRFs. However, this problem is not straightforward and it leads in practice to limiting approximations. The difficulty of integrating statistical shape priors in the MRF formulation is generally due to the incompatibility between the local graph connectivity and the global shape parameters.

Our Contributions Motivated by this challenging and difficult problem, we introduced in this thesis a novel statistical shape model and used it to enforce prior knowledge in segmentation algorithms, that were formulated in a Markov random fields (MRF) framework. In our model, the shape is represented by a graph where the vertices refer to control points lying on the shape boundary, and inter-landmark constraints are encoded by the graph edges. Inspired by the Euclidean Distance Matrix (EDM) theory, we defined these constraints as the probability distribution function of the nor-

malized chord lengths in the graph, leading to a similarity-invariant representation. Although the proposed shape model encodes local properties between pairs of control points, it is able to enforce the global shape constraints thanks to the graph structure and connectivity. Because our model uses inter-point normalized distances as shape representation, we only need to estimate one-dimensional probability density functions for the statistical shape model. Thereby, we alleviate the need of an extensive pool of training data, which is needed in higher dimension estimation.

Then, observing the computational burden induced by a complete graph structure, and the redundant information that is carried by the complete set of edges, we tackled the problem of graph structure learning. Seeking efficiency and model compactness, we approximated the complete graph by a k-fan graph which structure is determined from a training set of shapes using manifold learning and unsupervised clustering techniques. We showed that this approximation is tight in the case of the exact matching problem, by using concept of graph rigidity. We qualitatively verified this strong geometric property in experiments, where we recovered 2D and 3D shapes from random points thanks to the shape prior constraints.

The obtained k-fan graph belongs to the more general class of chordal graphs, and hence inherits the perfect elimination order property. Theoretically, this property allows the use of dynamic programming to exactly solve optimization problem defined on the graph in polynomial time. Unfortunately, the dimensionality of the applications that we considered made this approach intractable in practice.

In the next step, we proposed different methods to combine the shape prior with various image cues in a maximum a posteriori MRF optimization framework towards image segmentation. The obtained energy minimizations are cast into labeling problems that are efficiently solved using recent approximate optimization techniques. In the experiments, we noticed that we were able to recover the global optima in some cases.

More specifically, we integrated statistical regional image information in the objective energy as unary potentials using a Voronoi decomposition of the image domain. We proposed an iterative coarse-to-fine procedure to solve the problem, in a deformable model fashion. We also considered the use of

landmark-based cues through boosted classifiers in a Bayesian framework. To account for poor generalization performance of the classifiers, we proposed a method that detects outliers during inference, and repairs their positions according to the shape prior. We also proposed a tracking algorithm that uses edge-based image information and combines the static shape model with dynamic priors. In addition, we applied graph matching techniques in the image registration setting to select optimal mutually-salient corresponding landmarks. Although we used these different image cues seperately in the applications that we presented, our framework is flexible and allows the integration of cues of different nature at the same time.

Despite the encouraging results that we obtained, we think that our approach has some limitations. Our shape model handles naturally partial object occlusions. However, it does not explicitly account for self-occlusions. Hence, our region-based segmentation algorithm may be unsuitable to process objects with overlapping parts. Moreover, the scale-invariance of our model is enforced through normalization, which supposes knowing the object scale. We resort in practice to approximations in order to estimate the unknown scale values. Computing the regional image information as unary potential is another approximation that we made.

Future Research Directions In our future research work, we will endeavor to address the above mentioned downsides. The use of pairwise terms to model the regional support will reduce the shortcomings introduced from the Voronoi approximation. In particular using the Stokes' theorem, regional integrals could be expressed as inner products on the local 2D segments forming the 2D contour. Then, an explicit and direct estimation of the regional properties would be feasible. Furthermore, the combination of image-based (edges or regional support) and landmark based segmentation will be also feasible. The idea will be to have a two layer graph with interconnected variables, one acting on the landmarks and the second acting on the image. The interconnections will guarantee that the solution to both problems will be unique.

The case of 3D will be more complex since it will require higher order MRFs. The Green's theorem will decompose the regional image term on the boundary elements (triangles in 3D) in an exact manner. However, such a

formalism will inherit higher order MRFs. This is also a natural extension of our approach since it will introduce the notion of complete shape invariance to the process. The use of third order potentials inherits invariance with respect to scale changes, translation and rotation, while being able to cope with arbitrary image support. The main challenge in this context will be to minimize the energy but either the use of methods that convert higher order MRFs into pairwise MRFs or the ones based on dual decomposition and master-slave approaches will be considered.

As far as optimization is concerned, we have already mentioned that benefiting from the structure of our model through the use of dynamic programming is cursed by the dimensionality of the considered problems. Developing optimization algorithms that take into account the particular structure of the k-fan graph would represent an interesting research direction.

The case of spatio-temporal shape modeling is also of great interest in particular for medical image analysis when studying organs with periodic motion. Inferring the temporal connectivity of the graph can be extremely challenging and beneficial as well since the notion of k-fan model is not applicable. Re-formulating the inference process as an unsupervised clustering problem with terms accounting for spatial and others for temporal connectivity might be the right direction to go.

Last but not least in all above approaches the selection of the model was based on purely geometric properties and characteristics. Then, during inference, the model was associated with image cues. A promising research direction would bridge the gap between these two aspects. An interesting modeling scheme would not only learn the structure of the graph from the data, but it would also benefit from training image information to define the control points as salient image points. Combining shape and appearance learning is a possible extension of our work.

Appendix A

Extensions and Other Applications

Contents

- Chapter A.1 represents an independent part of the thesis and is based on a joint work with Yangming Ou.
- Chapter A.2 is based on a joint work with Yun Zeng.

A.1 Graph Matching for Registration

A.1.1 Introduction

Deformable image registration is one of the most challenging problems in medical imaging. We distinguish two classes of prior work: iconic methods and geometric methods. Iconic (or voxel-based) methods optimize a similarity criterion (for instance mutual information) that is defined on the images in order to find an optimal transformation with respect to this criterion. These methods generally perform well for intra-modal registration. However, when images belong to different modalities, such approaches can lead to failure. In the geometric methods, landmarks are extracted from the images and the transformation that maps the source landmarks to the target landmarks is estimated.

Establishing landmark correspondences is of a great importance in many landmark-based non-rigid registration algorithm. This task is often performed in two successive steps, namely, landmark detection (e.g., [Kadir & Brady 2001]) and landmark matching (e.g., [Joshi & Miller 2000]). One can formulate two criticisms to these separate two-step methods.

First, landmarks are often separately detected in two images. Although having individual saliency property, they might be not discriminative for matching. As pointed out in [Ou & Davatzikos 2009], salient points in one image are not necessarily present, or uniquely present in the other image. Therefore, an alternative approach that detects pairs of landmarks with a preferably unique matching score across images would yield better registration performance.

Second, individually optimal correspondence for each landmark is not necessarily globally optimal for the deformation field. For instance, such an approach might result in displacement vectors at two nearby landmarks that point to completely opposite directions, leading to a dense deformation field with self-intersections. This undesired behavior can be avoided by seeking a trade-off between the global smoothness of the resulting transformation and the landmarks similarities when establishing correspondences.

In this section, we tackle the above-mentioned drawbacks. First, instead of extracting individual interest points, we simultaneously detect and match

a number of mutually-salient candidate landmark pairs, i.e., pairs that are uniquely corresponding to each other across images. This pairs finding step relies on the mutual-saliency measure introduced in [Ou & Davatzikos 2009], which assesses the matching uniqueness of a pair of voxels. In order to alleviate the second limitation and seeking global optimality of the correspondences, we formulate the problem in the Markov-random-field (MRF) framework. We combine similarity measures with geometric constraints in an energy functional. Its minimization leads to consistent landmark displacements, thereby avoiding potential self-intersections in the resultant deformation field. Radiologists evaluated the results that we obtained, suggesting the advantage of the two components in our framework.

The remainder of this section is organized as follows: we present our framework in section A.1.2 and provide our results in section A.1.3.

A.1.2 Detecting Landmark Pairs

In this section, we first give a brief definition of mutual-saliency measure [Ou & Davatzikos 2009]. Then, we describe the two components of our framework: the detection of mutually-salient landmark pairs, and the MRF formulation to find globally optimal correspondences.

Definition of Mutual-Saliency Measure

Given two images $\mathcal{I}_1 : \Omega_1 \mapsto \mathbb{R}$ and $\mathcal{I}_2 : \Omega_2 \mapsto \mathbb{R}$ in the 3D image domains $\Omega_i (i = 1, 2) \subset \mathbb{R}^3$, a pair of voxels $\mathbf{u} \in \Omega_1$ and $\mathbf{v} \in \Omega_2$ is mutually-salient if they are similar to each other and only to each others in their neighborhoods. In the perfect case that is shown in Fig. A.1, the similarity map between $\mathbf{u}$ and all voxels in the neighborhood of $\mathbf{v}$ should exhibit a delta-shaped function centered at $\mathbf{v}$. We denote the mutual-saliency between $\mathbf{u}$ and $\mathbf{v}$ as $\mu(\mathbf{u}, \mathbf{v})$. We define this quantity as the ratio of the mean similarity in the central part of the neighborhood of $\mathbf{v}$ (denoted as $CN(\mathbf{v})$), by the mean similarity in the peripheral neighborhood of $\mathbf{v}$ (denoted as $PN(\mathbf{v})$), or:

$$\mu(\mathbf{u}, \mathbf{v}) \stackrel{def}{=} \frac{\frac{1}{|CN(\mathbf{v})|} \sum_{\mathbf{w} \in CN(\mathbf{v})} \rho(\mathbf{u}, \mathbf{w})}{\frac{1}{|PN(\mathbf{v})|} \sum_{\mathbf{w} \in PN(\mathbf{v})} \rho(\mathbf{u}, \mathbf{w})} \quad , \tag{A.1}$$

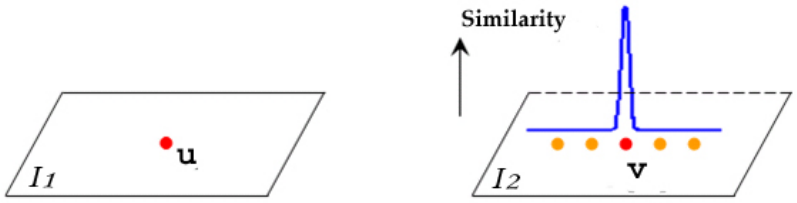


Figure A.1: The idea of mutual-saliency measure.

where $\rho(\mathbf{u}, \mathbf{w})$ is a similarity measure between the voxels $\mathbf{u}$ and $\mathbf{w}$. It is defined on the d-dimensional attribute vectors $A_i(\cdot)$ characterizing each voxel ($i = 1, 2$ for image $\mathcal{I}_i$), or:

$$\rho\left(\mathbf{u}, \mathbf{w}\right) \stackrel{def}{=} \frac{1}{1 + \frac{1}{d}\|A_1(\mathbf{u}) - A_2(\mathbf{w})\|^2} \quad . \tag{A.2}$$

Hence, the smaller their attribute vector Euclidean distance, the higher their similarity will be. The central neighborhood $CN(\cdot)$ and peripheral neighborhood $PN(\cdot)$ are defined as concentric rings around the voxel of interest, and their radii are determined according to the scales of the descriptors A_i [Ou & Davatzikos 2009]. More specifically, we construct the attribute vector $A_i(\cdot)$ by incorporating the multi-scale and multi-orientation Gabor attributes. We opted for Gabor attributes to characterize each voxel because different application, including registration (e.g., [Liu *et al.* 2002] [Ou & Davatzikos 2009]), proved their success in practice. Moreover, their ability to capture multi-scale and multi-orientation cues provides them with a great potential to describe each voxel more distinctively than the intensity attribute or other texture features [Kadir & Brady 2001]. Fig. A.2 shows examples of different voxel pairs having different mutual-saliency values. Here, similarity maps (c-e) are generated by calculating the attribute-based similarity between a specific voxel (denoted as red $+$, blue $\times$ and orange $\star$) in the subject image (a) and all voxels in the template image (b). It is observed that, the red point and its correspondence has the highest mutual-saliency (indicating most unique correspondence), followed by the blue point and its correspondence, and lastly the orange point and its correspondence. Note that the mutual-saliency is defined on a pair of voxels. Therefore, it can be

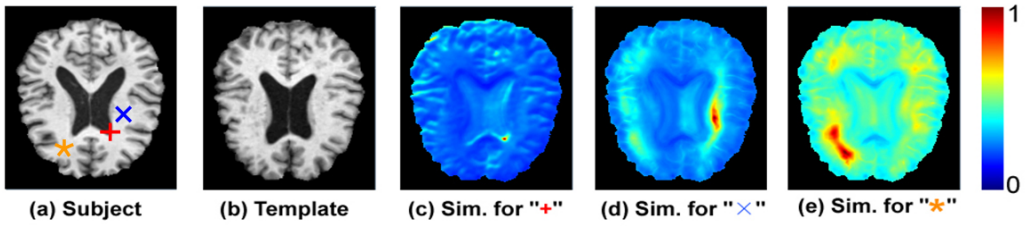


Figure A.2: Examples of different voxel pairs having different mutual-saliency values.

used to detect pairs of landmarks, as those having high similarity uniqueness across images. Thereby, one can naturally unify landmark detection and landmark matching in the same procedure. Then, detecting a landmark also determines automatically a corresponding interest point which is guaranteed to be relatively unique.

Based on this property, we describe now how we extract mutually-salient landmark pairs between two images.

Detecting Mutually-Salient Landmark Pairs

Based on the aforementioned advantage of mutual-saliency measure, this subsection describes the extraction of mutually-salient landmark pairs. In order to favor uniform registration accuracy in the image, it is preferable that the detected mutually-salient landmark pairs are scattered within the image space. Therefore, as sketched in Fig. A.3, the source image space Ω_1 is regularly partitioned into J regions. In each region, the K most mutually-salient landmark pairs across images are selected. Formally, from the j^{th} $(j = 1, 2, \ldots, J)$ region, we select K pairs $(\mathbf{p}_j^k \in \Omega_1, \mathbf{q}_j^k \in \Omega_2)_{k=1}^K$ that are ranked by their similarity measure weighted by the mutual-saliency value,or: $\rho(\cdot,\cdot) \times \mu(\cdot,\cdot)$. This criterion encourages the detected pair to be similar through $\rho(\cdot,\cdot)$, while favoring uniquely matched landmarks thanks to the term $\mu(\cdot,\cdot)$. Note in Fig. A.3 that, the template image space Ω_2 is not partitioned, because at this stage, no transformation is conducted, therefore no corresponding regions can be assumed.

As we observed earlier, the top ranked pair is a locally optimal pair, that might be not globally optimal with respect to a smooth deformation field be-

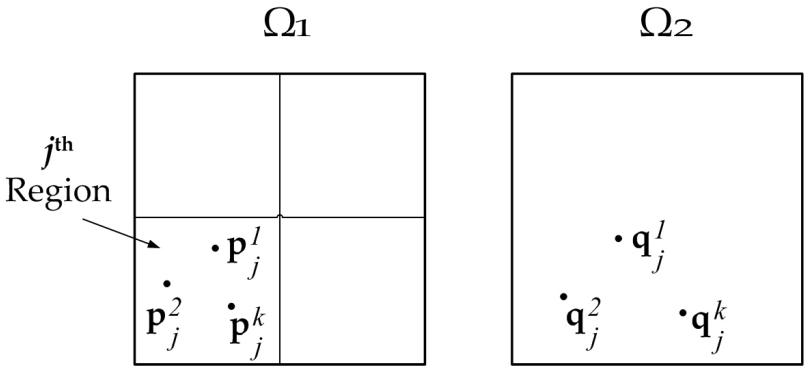


Figure A.3: Sketch of the detection of landmark pairs.

tween the images. Therefore, we keep the K top-ranked pairs and seek the globally optimal ones using a Markov-random-field (MRF)-based optimization. This step is explained in the next sub-section.

Finding Globally Optimal Landmark Pairs by MRF Formulation

Let us denote all the $J \times K$ landmark pairs detected in the previous step as a set $\mathcal{P} = \{(\mathbf{p}_j^k, \mathbf{q}_j^k) | j = 1, 2, \ldots, J; k = 1, 2, \ldots, K\}$. One should note that K is not necessarily the same number for different regions. In practice, we keep $K = 10$ top-ranked candidate pairs in each region.

We aim in this section to select one pair (out of K pairs) from each region, i.e., $\mathcal{P}^\star = \{(\mathbf{p}_j^{l_j}, \mathbf{q}_j^{l_j}) | j = 1, 2, \ldots; l_j \in \{1, 2, \ldots, K\}\}$, such that the obtained set of pairs is globally optimal, in that they altogether maintain high mutual-saliency as well as the smoothness of the resultant deformation field.

We can formulate this task as a Markov-random-field (MRF)-based labeling problem. We seek a set of optimal labels $\mathbf{l}^\star = (l_1, \ldots, l_J)$, where $l_j \in \{1, 2, \ldots, K\}$ is the label (or index) of the globally optimal pair out of K candidate pairs in the j^{th} region. Therefore, we construct a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where each node in $\mathcal{V}$ represents a region and each edge in $\mathcal{E}$ expresses the constraint of a local geometric compatibility between the landmark pairs.

In this formulation, our goal can be achieved by minimizing the following

labeling energy on the graph,

$$\mathbf{l}^{\star} = \arg\min_{\mathbf{l}} E(l_1, \ldots, l_J), \tag{A.3}$$

where the energy $E(l_1, \ldots, l_J)$ consists of two terms,

$$E(l_1, \ldots, l_J) = E_{data}(l_1, \ldots, l_J) + \alpha\, E_{reg}(l_1, \ldots, l_J). \tag{A.4}$$

Those two terms express the two criteria for selecting the globally optimal landmark pairs. The first term, E_{data}, is related to the quality of the landmark pair, as defined in the pair extraction step, or:

$$E_{data}(\mathbf{l}) = \sum_{j=1}^{J} \exp\left(-\frac{\mu(\mathbf{p}_j^{l_j}, \mathbf{q}_j^{l_j}) \cdot \rho(\mathbf{p}_j^{l_j}, \mathbf{q}_j^{l_j})}{2\sigma^2}\right), \tag{A.5}$$

where σ is a scaling factor, estimated as the standard deviation of $\mu(.,.)\times\rho(.,.)$ values of all the candidate pairs.

The second term, E_{reg} is a regularization term that favors the smoothness of the resultant deformation field. The regularization imposes constraints on both the spatial positions and the displacement directions on the selected landmark pairs. Specifically, spatial position constraint encourages those selected pairs to scatter in the image space other than being close to each other. Displacement direction constraint encourages displacement vectors on adjacent pairs to be consistently oriented, therefore avoiding self-intersection of the resultant deformation field. Those two constraints are similar to the distortion terms in [Leordeanu & Hebert 2005] and [Torresani *et al.* 2008]. Mathematically, the regularization term is expressed as:

$$E_{reg}(\mathbf{l}) = \sum_{(m,n)\in\mathcal{E}} \left\| (\mathbf{p}_m^{l_m} - \mathbf{p}_n^{l_n}) - (\mathbf{q}_m^{l_m} - \mathbf{q}_n^{l_n}) \right\|. \tag{A.6}$$

One should note that such a framework is not invariant to scale changes. However, since we have assumed that a rigid registration step was performed beforehand, this is not a main concern.

The defined labeling problem is NP-hard in general. However, recent approximate solvers perform well in practice. The total energy in Eqn. (A.3) is minimized using the sequential tree-reweighted message passing (TRW-S) algorithm [Kolmogorov 2006].

A.1.3 Experimental Validation

Our framework is applied to finding mutually-salient landmark pairs in brain and cardiac images. Results in the following sub-sections aim to demonstrate the advantage of unifying landmark extraction and landmark matching in our framework.

Results for Detecting Mutually-Salient Landmark Pairs Fig. A.5 shows a landmark pair detected across subjects based on the mutual-saliency measure. Alternative methods that consider only image intensity to describe voxels and that use edge/corner detectors to extract landmarks separately in the images would probably fail to detect this pair of voxels: these voxels are simply not edges or corners or surface boundaries. We succeed in detecting this pair for two reasons. On the one hand, Gabor attributes are powerful descriptors that provide a distinctive characterization of voxels, as shown in similarity maps (c,d). On the other hand, the mutual-saliency measure effectively quantifies the matching uniqueness, as shown in similarity maps (e,f). Overall, this example shows the advantage of detecting landmarks in pairs other than one-by-one and separately from the two images.

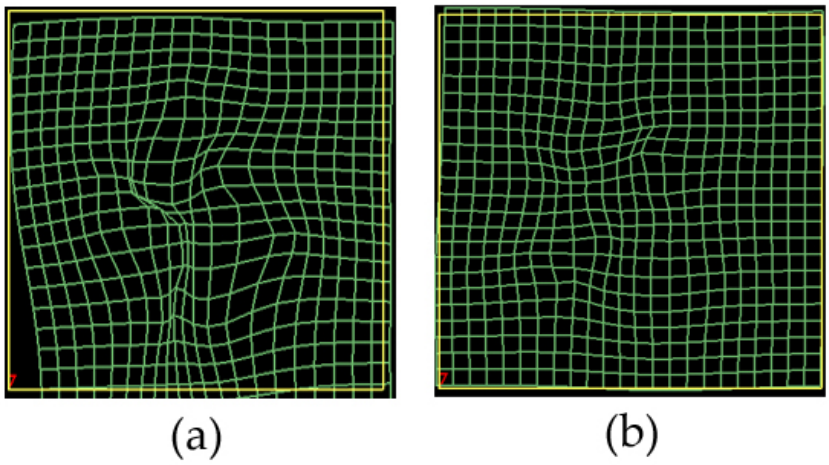


Figure A.4: Dense deformation fields generated by (a) M1 – no MRF regularization and (b) M2 – with MRF regularization.

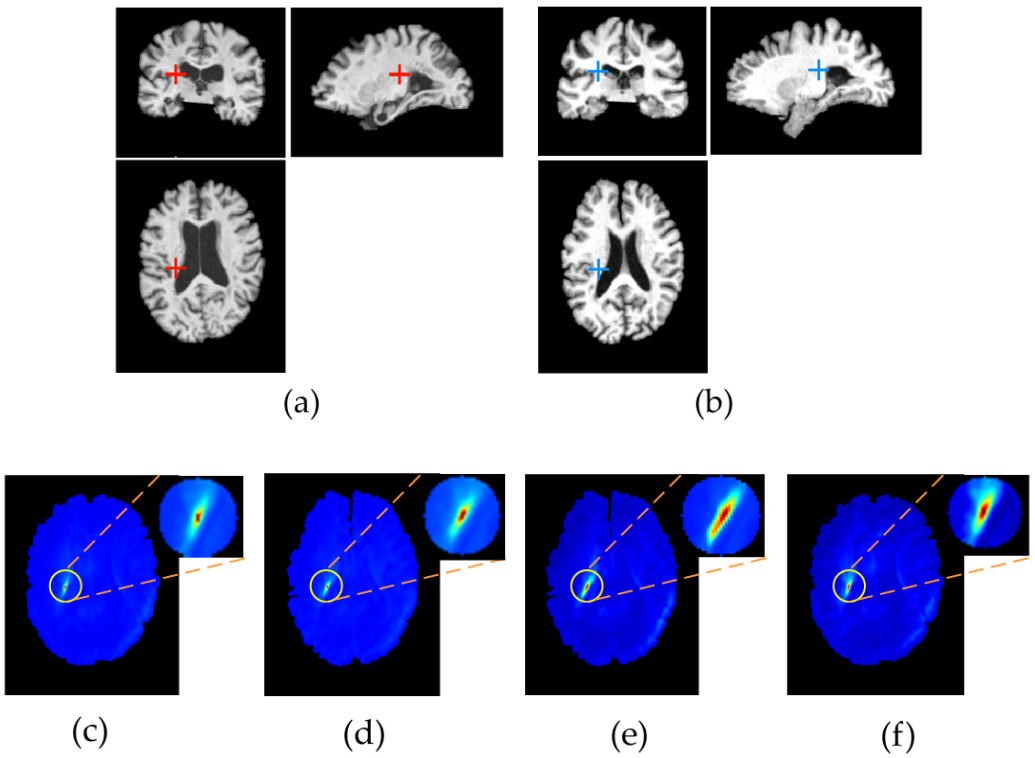


Figure A.5: An example landmark pair (denoted by red and blue crosses) detected based on mutual-saliency measure. (a) Source image $\mathcal{I}_1$ (b) Target image $\mathcal{I}_2$. Similarity maps are generated (c) between the red cross point and all voxels in Ω_1 (d) between the blue cross point and all voxels in Ω_2 (e) between the red point and all voxels in Ω_2 and (f) between the blue cross point and all voxels in Ω_1.

Results for Finding Globally Optimal Landmark Pairs To demonstrate the advantage of MRF-based regularization, landmark pair detections without and with the MRF regularization are compared. They are respectively denoted as Method 1 (M1) and Method 2 (M2) in the following comparisons. The first comparison is in terms of the smoothness of the resultant dense deformation field (based on thin-plate-spline interpolation). The avoidance of self-intersection of the deformation field in Fig. A.4(b) illustrate the benefit of a global MRF-based optimization as we described in Section A.1.2. The second comparison involves the radiologist's evaluation on the landmark correspondences finally determined without and with MRF regularization (M1

Table A.1: Radiologist's evaluation on results generated by two methods (M1 and M2 for without and with MRF optimization). $\gg$ (or $\ll$) means "obviously better (or worse)", $>$ (or $<$) means "slightly better (or worse)", and $\approx$ means "almost equivalent".

	M1 $\gg$ M2	M1 $>$ M2	M1 $\approx$ M2	M1 $<$ M2	M1 $\ll$ M2
Dataset1			68%	12%	20%
Dataset2		4%	80%	16%	
Dataset3	4%	4%	68%	20%	4%
Dataset4		4%	60%	24%	12%

v.s. M2). Four datasets are used: *Dataset1* (size 256 $\times$ 256 $\times$ 171) and *Dataset2* (size 192 $\times$ 236 $\times$ 171) for two different sets of intra-modality brain MR images across-subjects; *Dataset3* (size 192 $\times$ 236 $\times$ 171) for a pair of multi-modality brain MR images across-subjects; and *Dataset4* (size 150 $\times$ 150 $\times$ 49) for a pair of intra-modality cardiac MR images across-subjects. For each dataset, 25 random pairs out of all the finally selected pairs (typically hundreds) are evaluated in terms of uniqueness and accuracy of correspondences. Evaluation results in Table A.1 show the advantage of incorporating MRF regularization (M2) in all datasets.

A.2 Dynamic Shape Prior and Tracking

A.2.1 Introduction

Object tracking is one of the fundamental problems of motion analysis in computer vision. Despite enormous work in the field, the case of highly deformable objects is still an open problem. Prior work on tracking involves blob-based appearance methods, dynamical systems and part-based models. We briefly mention some examples of these lines of work.

In the first case, one can cite for instance methods like the Kanade-Lucas-Tomasi tracker [Tomasi & Kanade 1991], correlation-based methods [Shi & Tomasi 1994], the mean-shift algorithm [Comaniciu & Meer 2002] and its numerous variants [Collins 2003]. These methods use a similarity criterion to compare the previous appearance of the object with possible candidates in the new images towards recovering the most probable position. These methods are a good compromise in terms of computational complexity but may fail with severe deformations.

Dynamical systems are a promising alternative to encode motion dynamics. One can cite numerous examples using such model: Kalman filtering [Kalman 1960], condensation [Isard & Blake 1998], or multiple hypotheses testing and particle filter tracking [Arulampalam *et al.* 2002]. These methods benefit from an explicit model that represents the motion of the object, and can thereby perform better than blob-based methods.

One should also mention methods that aim to track articulate models. In such a context, objects are represented with parts, and then constraints between the relative positions of these parts are introduced [Felzenszwalb & Huttenlocher 2005] [Black & Jepson 1998] [Urtasun *et al.* 2006] [Wang *et al.* 2008]. These methods can be very efficient but assume an explicit hierarchical representation of the model, and impose constraints on their dynamics. Therefore, their use in the context of tracking arbitrary highly deformable objects is not straightforward.

The tracking task has been formulated in different optimization frameworks. For instance, [Paragios & Deriche 1999] proposed a solution to the problem in the level set framework. [Cremers 2008] adopted a variational approach as well incorporating dynamical shape priors. Other approaches

are based on Markov Random Fields (MRFs). [Kohli *et al.* 2008] introduced an algorithm that performs human pose estimation and segmentation. Their method estimates in each frame the pose parameters of the stickman shape model, which is subsequently used in a binary graph cut segmentation as a distance map prior. An alternative approach is presented in [Malcolm *et al.* 2007b], where the solutions in the previous frames are used to predict the poses of the tracked objects. Then, binary shape maps that are centered at the predicted positions are used to enforce shape prior in a multi-label graph cut. Another method of multiple object tracking was recently proposed in [Wang *et al.* 2009]. The authors define a generative image model that naturally handles occlusions in the MRF framework. Minimizing the objective energy allows them to simultaneously segment and track objects, and determine their depths. In another recent work, [Bugeau & Pérez 2009] address the problem of tracking in dynamic scenes with moving camera. The authors proposed an algorithm that is threefold. They determine first a set of points which motion is different from the camera's (the camera motion is assumed to be the predominant one in the scene). A descriptor vector that combines multiple cues such as motion and photometric features is then associated with each extracted point. In the second step, these points are clustered based on their descriptors using a variable-bandwidth mean-shift algorithm. Then, the joint segmentation and tracking are achieved by performing a graph cut on the pixel level, where the cluster information is used as a constraint.

In this section, we present an extension of our segmentation framework to the temporal domain. We describe a tracking algorithm that is based on image boundary information, and that benefits from static and dynamic shape priors. The static shape prior refers to the normalized chord length shape model introduced in section 2.2. Hence, the spatial consistency is encoded through probability densities on the relative positions of pairs of control points. Similarly, we encode temporal priors through relative deformations of different control points in time once motion has been implicitly accounted for. Tracking is then reformulated as an energy minimization problem, where the positions of the control points are deformed in the current frame in order to solve a MAP-MRF problem.

The remainder of this section is organized as follows: first, we define the

dynamic shape prior, and then, we formulate the edge-based energy term. In the last subsection, we show the tracking results that we obtained.

A.2.2 Weak Edges - Static and Dynamic Shape Priors

Let $\left(\mathcal{I}_1, \ldots, \mathcal{I}_N\right)$ be a sequence of image frames where an object is sought. Let $\mathbf{x}_i(t)$ be the position of the i^{th} control point in the image frame $\mathcal{I}_t$. We extend the MAP estimation problem defined in section 3.2 to the temporal domain by considering an energy term that accounts for the interactions between the control points in the current frame, and those in the previous ones. Then, we rewrite the equation (3.13) as minimizing:

$$\begin{aligned} E\left(\mathbf{x}_1(t), \ldots, \mathbf{x}_n(t), \mathcal{I}(t)\right) = & \beta E_{dyn}\left(\tau, \mathbf{x}_1(t), \ldots, \mathbf{x}_n(t)\right) \\ & + E_{shape}\left(\mathbf{x}_1(t), \ldots, \mathbf{x}_n(t)\right) \\ & + \alpha E_{image}\left(\mathbf{x}_1(t), \ldots, \mathbf{x}_n(t), \mathcal{I}(t)\right) \quad , \end{aligned} \tag{A.7}$$

where $E_{dyn}\left(\tau, \mathbf{x}_1(t), \ldots, \mathbf{x}_n(t)\right)$ is a dynamic shape prior of order τ, as explained in the following, and α and β are weighting parameters that modulate the contribution of each energy term. Recall that the static shape prior energy E_{shape} is related to our k-fan shape model and expressed as pairwise interaction between the landmarks (2.30). We consider such pairwise interactions in the temporal domain as well.

Dynamic Prior

Let us also consider the inter-point normalized distance between two landmarks in different frames separated by an index τ. We suppose that the global motion is accounted for through translation of the gravity center of the object:

$$d_{ij}(t,\tau) = \frac{\left\| \left(\mathbf{x}_i(t) - \bar{\mathbf{x}}(t)\right) - \left(\mathbf{x}_j(t-\tau) - \bar{\mathbf{x}}(t-\tau)\right) \right\|}{\widetilde{d}(t,\tau)} \quad , \tag{A.8}$$

where $\widetilde{d}(t,\tau) = \frac{1}{n^2}\sum_{k=1}^{n}\sum_{l=1}^{n}\left\| \left(\mathbf{x}_k(t) - \bar{\mathbf{x}}(t)\right) - \left(\mathbf{x}_l(t-\tau) - \bar{\mathbf{x}}(t-\tau)\right) \right\|$, and $\bar{\mathbf{x}}(t)$ (resp. $\bar{\mathbf{x}}(t-\tau)$) being the gravity center of the object at the frame t (resp. at the frame $t-\tau$). Given a training set of image frames where the object of interest is segmented, and correspondences between the control

points are guaranteed, one can learn a probability distribution $p\big(d_{ij}(t,\tau)\big)$ for each of the "temporal" distances $d_{ij}(t,\tau)$, and for different dynamic orders τ. We consider such interactions not only for the the same landmark (simple case $d_{ii}(t,\tau)$), but for the relative dynamics of the different landmarks. Then, given the position of the object at the previous frames $(t-\tau,\ldots,t-1)$, one can define dynamic constraints that account for the correlation between the landmark positions at the current frame and their motion history. More formally, the dynamic shape energy term can be written as:

$$\begin{aligned} E_{dyn}\big(\tau,\mathbf{x}_1(t),\ldots,\mathbf{x}_n(t)\big) &= \sum_{\theta=1}^{\tau}\sum_{i=1}^{n}\sum_{j=1}^{n} -\log\Big(p\big(d_{ij}(t,\theta)\big)\Big) \\ &= \sum_{i=1}^{n} V_i(\mathbf{x}_i(t),\tau) \ , \end{aligned} \tag{A.9}$$

where $V_i(\mathbf{x}_i(t),\tau) = \sum_{\theta=1}^{\tau}\sum_{j=1}^{n} -\log\Big(p\big(d_{ij}(t,\theta)\big)\Big)$. Hence the dynamic shape prior is expressed as a sum of unary potentials. Now that we have defined the static and dynamic prior model, the next task consists in defining relationships between the graph and the image sequence $(\mathcal{I}_1,\ldots,\mathcal{I}_N)$, which enables us to infer the optimal positions of the control points.

Weak Edge Support

We define the data-term in an image $\mathcal{I}_t$ of the sequence using its edges. Independently from the previous frames, the optimal positions of the control points should superimpose the boundary of the object to its edges in the image. We suppose that the boundary of the object can be simply approximated by connecting the successive control points in 2D (the 3D counterpart is a mesh of the control points, which would lead to high order MRF terms). Let $\mathcal{M}_t$ be the distance map to the edges of the considered image (we computed in practice the chamfer distance to the canny edges). Then, since the object contour can be approximated by the landmark segments, its distance to the edges can be computed using the same approximation. Supposing that the control points are indexed according to their order on the boundary, this

(a) frame 0 (b) frame 66 (c) frame 94

Figure A.6: Frames extracted from a tracking sequence using our method with the static prior. Note that the tracking quality is better when using the dynamic prior for the frame 66 for example (see Fig. A.7(b)): here one leg is missed.

yields the following data energy term:

$$\begin{aligned} E_{image}\left(\mathbf{x}_1(t),\ldots,\mathbf{x}_n(t),\mathcal{I}(t)\right) &= \sum_{i=1}^{n} \int_{[\mathbf{x}_i(t),\mathbf{x}_{i+1}(t)]} \mathcal{M}_t(u)du \\ &= \sum_{i=1}^{n} V_{i,i+1}(\mathbf{x}_i(t),\mathbf{x}_{i+1}(t),\mathcal{I}) \ , \end{aligned} \tag{A.10}$$

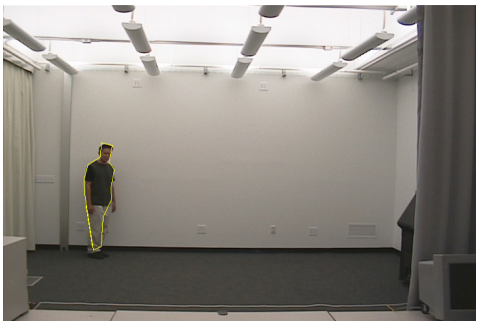
where we use the notation $\mathbf{x}_{n+1}(t) = \mathbf{x}_1(t)$. Although this low-level feature is sensitive to noise, and would often fail, its association with the static and dynamic prior overcomes its usual limitations. To summarize, the objective energy (A.7) is the sum of unary terms that account for the dynamic prior, and two types of pairwise terms: an edge-based one and a static prior-related one.

To minimize the objective energy, we consider the same approach described in section 3.2.3. At each image frame $\mathcal{I}_t$, we initialize the shape model using the result of the previous frame. We consider a quantization of the search space, and we solve iteratively several labeling problems, where each label refers to a possible control point displacement. The obtained MRF minimization is an NP-hard problem. We consider approximate and computationally-efficient solutions given by the FastPD algorithm [Komodakis *et al.* 2008b]. We also use a coarse-to-fine search to speed up computations.

A.2.3 Experimental Validation

To test the applicability of our model and algorithm, we considered a commonly-used example in tracking: walking people. The problem of tracking walking people provides a deformable object with interesting dynamics. We used in our experiments video sequences from the Georgia Institute of Technology database (http://www.cc.gatech.edu/cpl/projects/hid/Description.html).

We first selected a total of 455 frames of walking persons from the database with different gaits. We labeled manually these frames placing the landmarks at the corresponding positions. Then we used these labeled images to learn a static prior and a dynamic prior. Next, for the testing, we applied our trained priors to three video sequences. The model points were initialized close to the walking target. For the qualitative evaluation, we compare the results that we obtain using the static prior, and then by using the static and the dynamic prior. From our experiment we observe that the results with dynamic priors are more robust with respect to the noisy edges, and have a better quality than those from the static prior tests. Figures A.6 and A.7 show examples of the obtained results. Although the features we use are weak, our algorithm is able to track the object thanks to the learned prior. Additional tracking results on other sequences and using the dynamic shape prior are presented in Fig. A.8 A.9 A.10.

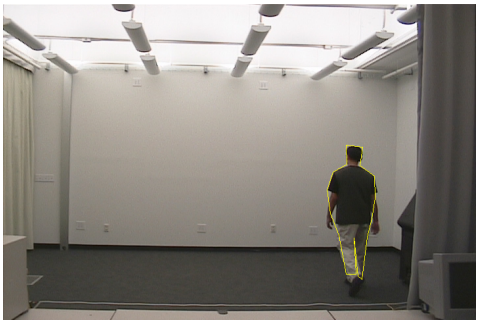


(a) frame 1

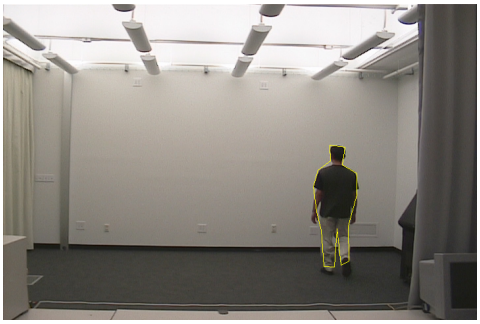
(b) frame 66

(c) frame 94

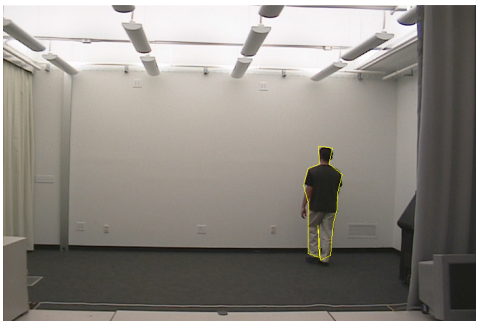
(d) frame 120



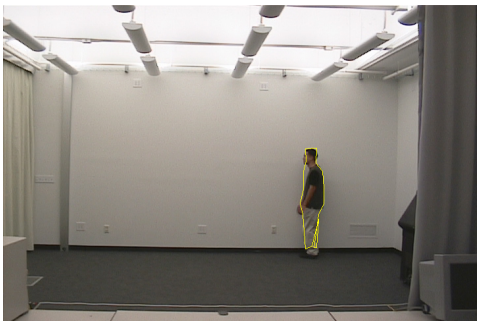
(e) frame 163



(f) frame 183



(g) frame 207



(h) frame 237

Figure A.7: Frames extracted from a tracking sequence using our method with the dynamic prior. Note that the tracking quality is better than in the static prior case for the frame 66 for example (see Fig. A.6(b)).

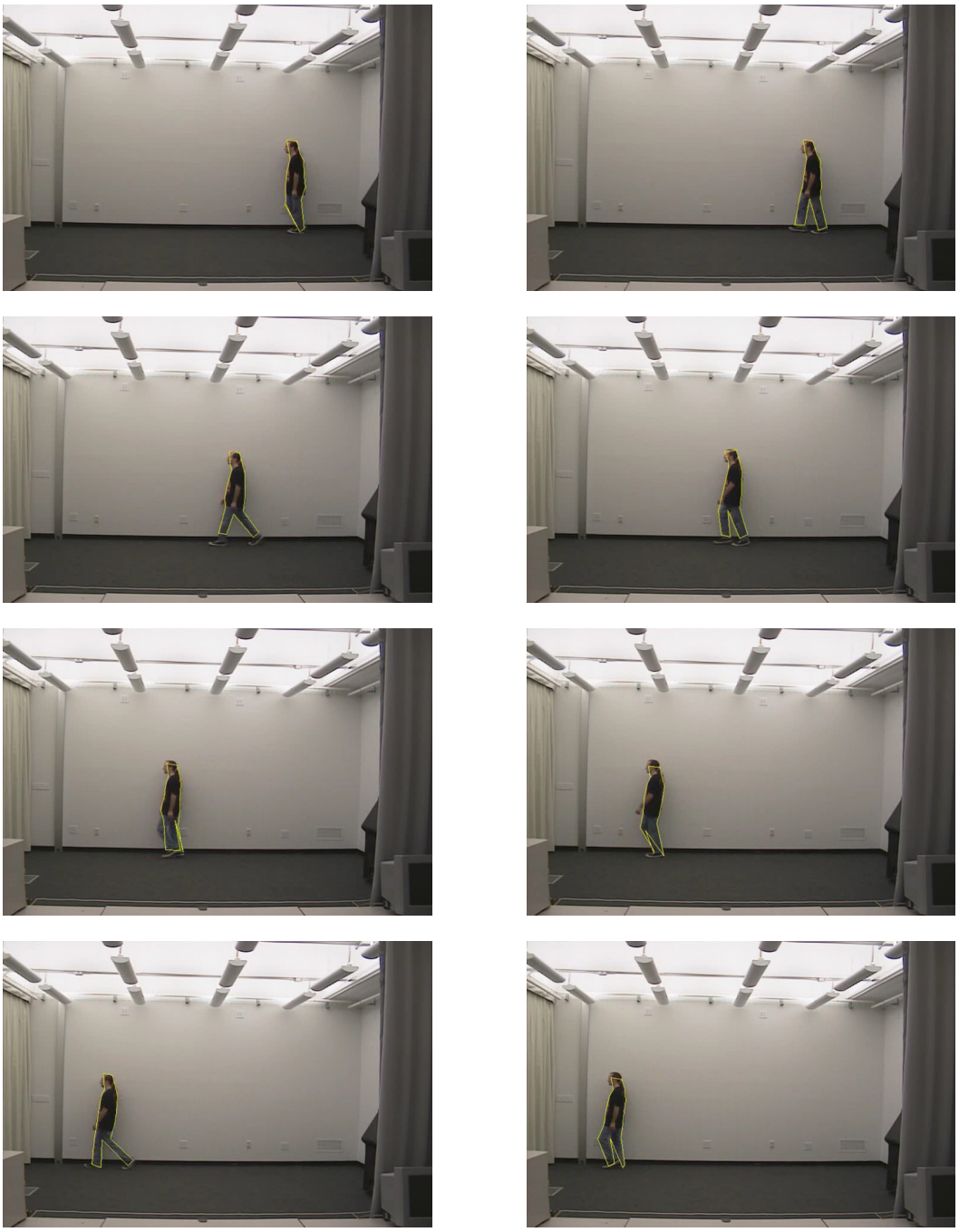


Figure A.8: Frames extracted from a tracking sequence using our method with the dynamic prior.

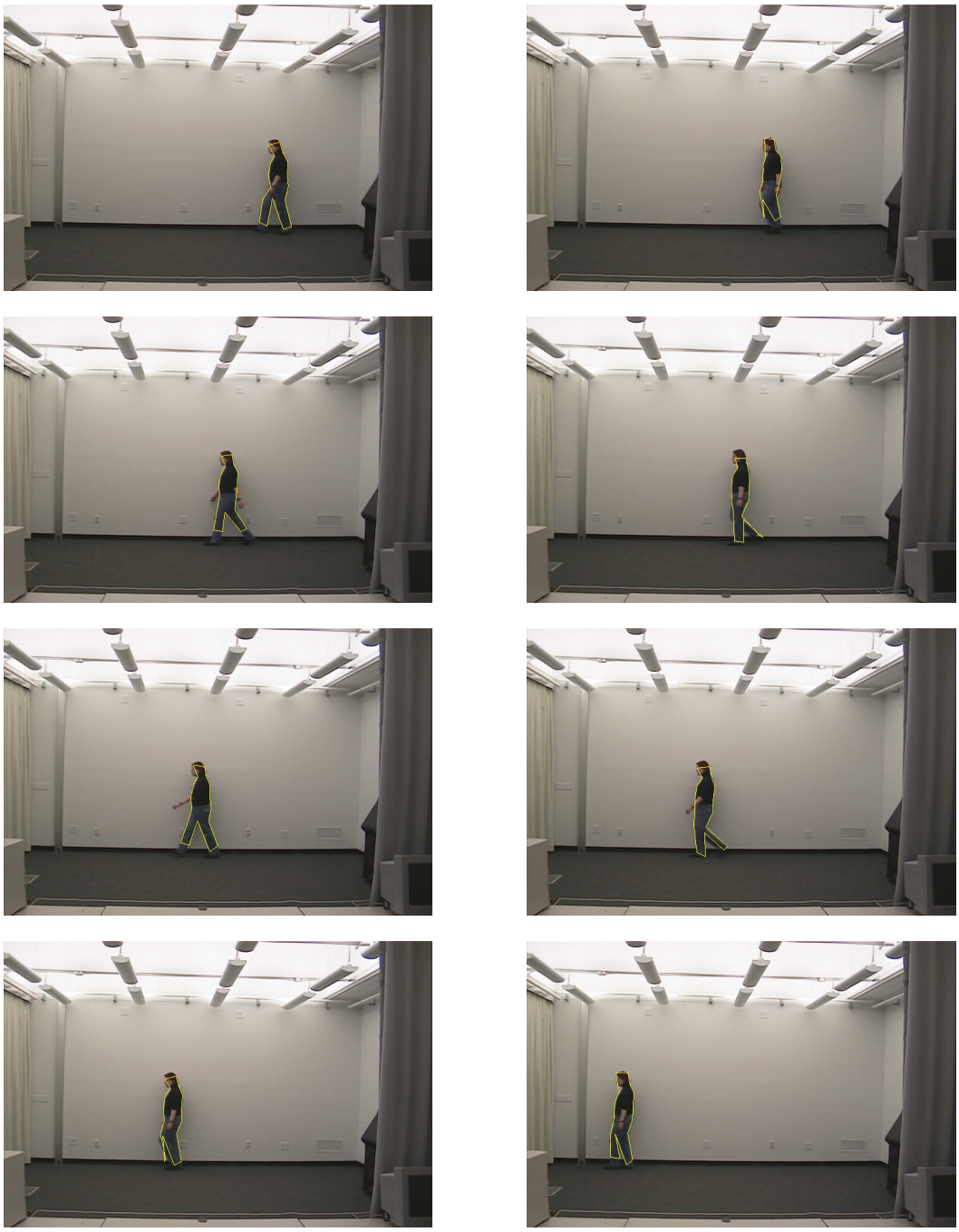


Figure A.9: Frames extracted from a tracking sequence using our method with the dynamic prior.

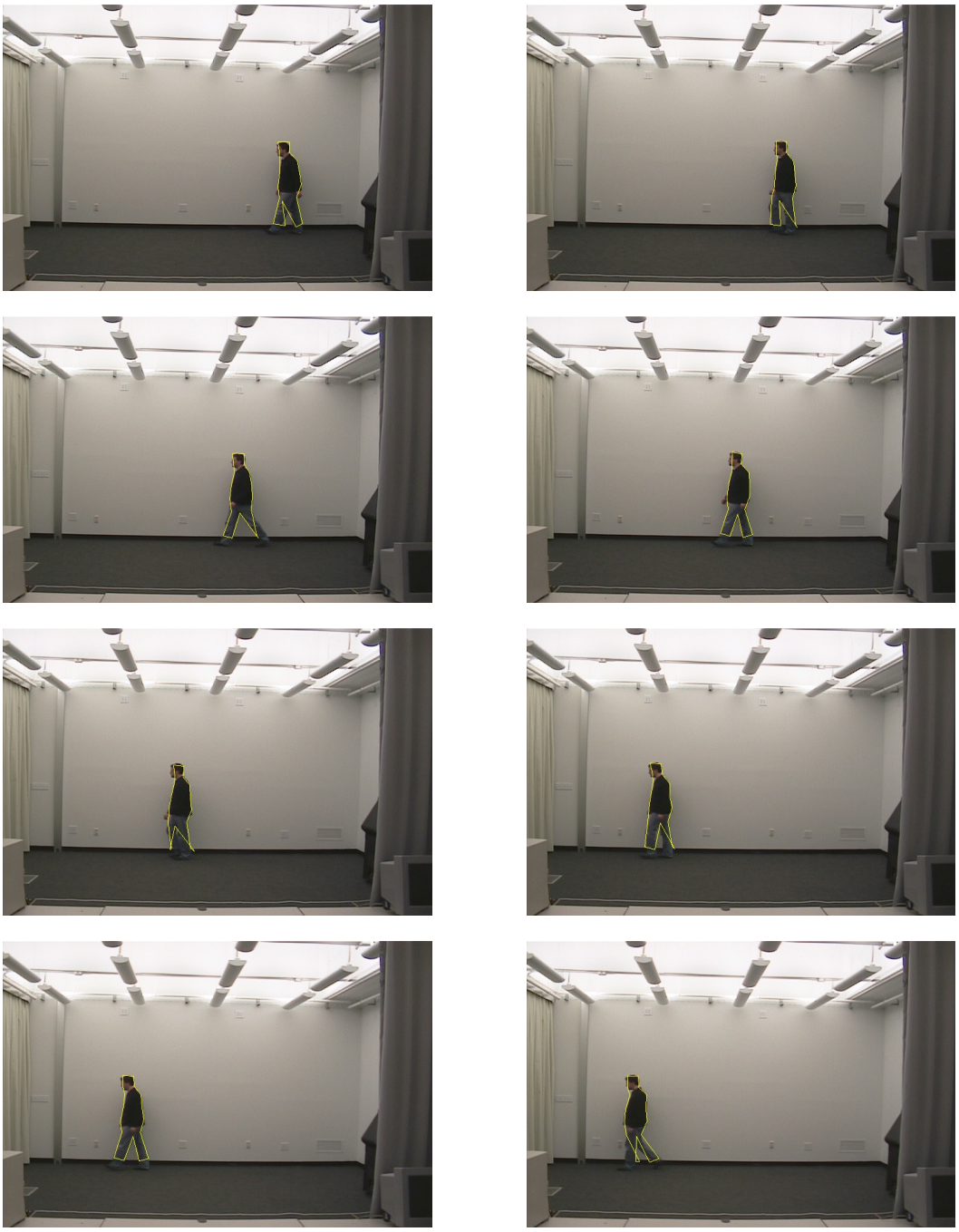


Figure A.10: Frames extracted from a tracking sequence using our method with the dynamic prior.

Appendix B

EDM Realization in $\mathbb{R}^n$

Proof of Theorem 2.1. Suppose that there exists N points $(x_1, \ldots, x_N)$ in $\mathbb{R}^n$ such that $d_{ij} = \|x_i - x_j\|_2$ for $1 \leq i, j \leq N$, and that $span(x_1, \ldots, x_N) = \mathbb{R}^n$. Without loss of generality, we may assume that $x_1 = 0$. Upon writing $x_i - x_j = (x_i - x_1) + (x_1 - x_j)$, we see that:

$$\|x_i - x_j\|^2 = \|x_i - x_1\|^2 + \|x_j - x_1\|^2 - 2x_i^T x_j$$

or equivalently,

$$x_i^T x_j = \frac{1}{2}(\|x_i - x_1\|^2 + \|x_j - x_1\|^2 - \|x_i - x_j\|^2) = \frac{1}{2}(d_{1i}^2 + d_{1j}^2 - d_{ij}^2) = a_{(i-1)(j-1)} \ .$$

Subsequently, for any $y \in \mathbb{R}^{N-1}$, we have:

$$y^T A y = \sum_{i,j=1}^{N-1} a_{ij} y_i y_j = \left\| \sum_{i=1}^{N-1} x_{i+1} y_i \right\|^2 \geq 0$$

meaning that A is positive semidefinite. Moreover, if we let V to be the $n \times (N-1)$ matrix whose i^{th} column is the vector x_{i+1} (where $1 \leq i \leq N-1$), then we see that A is the Gram matrix of V, or $A = V^T V$. Since $span(x_1, \ldots, x_N) = \mathbb{R}^n$, we have $rank(A) = n$.

Conversely, suppose that A is positive semidefinite and has rank $n \leq N-1$. Then we can write $A = U \Lambda U^T$, where U is a $(N-1) \times (N-1)$ orthogonal matrix and $\Lambda = diag(\lambda_1, \ldots, \lambda_n)$, where $\lambda_1 \geq \cdots \geq \lambda_n > \lambda_{n+1} = \ldots \lambda_{N-1} = 0$ are the eigenvalues of A. Now, let $(e_1, \ldots, e_N - 1)$ be the standard basis vectors of $\mathbb{R}^{N-1}$, and set $x_{i+1} = P_n \Lambda^{1/2} U^T e_i \in \mathbb{R}^n$ for $1 \leq i \leq N-1$, where P_n is the orthogonal projection of $\mathbb{R}^N$ onto the first n coordinates. Then for $1 \leq i, j \leq N-1$ we have:

$$d_{1(i+1)}^2 = a_{ii} = e_i^T A e_i = e_i^T U \Lambda U^T e_i = \left\| \Lambda^{1/2} U^T e_i \right\|^2 = \|x_{i+1}\|^2$$

$$d_{(i+1)(j+1)}^2 = a_{ii} + a_{jj} - 2a_{ij} = (e_i - e_j)^T A (e_i - e_j) = \|x_{i+1} - x_{j+1}\|^2 \ .$$

Thus, $(0, x_2, \ldots, x_N)$ defines the desired realization, which completes the proof. □

Appendix C

Spanning k-tree of a k'-tree

Proof of Lemma 2.1. Without loss of generality, we may assume that $k' > k$. Let $\mathcal{T}'$ be a k-tree with t vertices $v_1, \ldots, v_t$, and let v be the last vertex in a given SEO of $\mathcal{T}'$. The proof of the lemma is given by induction on t.

If $t = k'$, then $\mathcal{T}'$ is a clique. Let $\mathcal{T}$ be the k-tree that is built as follows (according to definition 2.2): start from the clique $\{v_1, \ldots, v_k\}$, and for $k+1 \leq i \leq t$, connect v_i to $\{v_{i-k}, \ldots, v_{i-1}\}$. Then $\mathcal{T}$ solves the problem and (P) holds clearly.

If $t > k'$, we apply the induction hypothesis to $\mathcal{T}' - v$. Then there exists a spanning k-tree $\mathcal{T}''$ of $\mathcal{T}' - v$, such that every k'-clique of $\mathcal{T}' - v$ contains a k-clique of $\mathcal{T}''$. In particular, the neighborhood of v in $\mathcal{T}' - v$ is a k'-clique Q', and so it contains a k-clique Q of $\mathcal{T}''$. Let $\mathcal{T}$ be the graph obtained by adding to $\mathcal{T}''$ the vertex v and all the edges between v and Q. Clearly $\mathcal{T}$ is a spanning k-tree of $\mathcal{T}'$. Moreover, $\mathcal{T}'$ satisfies (P). Indeed, we already know that every k'-clique of $\mathcal{T}'$ that does not contain v contains a k-clique of $\mathcal{T}''$, and thus of $\mathcal{T}$. Furthermore, every k'-clique $\mathcal{C}'$ of $\mathcal{T}'$ that contains v excludes exactly one vertex u in the neighborhood of v in $\mathcal{T}'$ (because v is simplicial and hence has exactly k' neighbors in $\mathcal{T}'$). Thus $\mathcal{C}' = Q' + v - u$.If u is not a vertex of Q then $Q \subset (Q' - u) \subset \mathcal{C}'$, and (P) is proved. Else, the remaining $k - 1$ vertices of Q together with v form a k-clique of $\mathcal{T}$ contained in $\mathcal{C}'$. □

Publications of the Author

- **Peer-reviewed Conference Publications:**

  - Ahmed Besbes, Nikos Komodakis, Ben Glocker, Georgios Tziritas and Nikos Paragios. *4D Ventricular Segmentation and Wall Motion Estimation Using Efficient Discrete Optimization.* In ISVC '07: Advances in Visual Computing, pp. 189–198, 2007.

  - Ahmed Besbes, Nikos Komodakis, Georg Langs and Nikos Paragios. *Shape priors and discrete MRFs for knowledge-based segmentation.* In CVPR '09: Proceedings of the 2009 Conference on Computer Vision and Pattern Recognition, pp. 1295–1302, 2009.

  - Ahmed Besbes, Nikos Komodakis and Nikos Paragios. *Graph-Based Knowledge-Driven Discrete Segmentation of the Left Ventricle.* In ISBI '09: Proceedings of the 6th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 49–52, 2009.

  - Ahmed Besbes and Nikos Paragios. *Landmark-based Segmentation of Lungs while Handling Partial Correspondences using Sparse Graph-based Priors.* In ISBI '11: Proceedings of the 8th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2011.

  - Radhouène Neji, Ahmed Besbes, Nikos Komodakis, Jean-François Deux, Mezri Maatouk, Alain Rahmouni, Guillaume Bassez, Gilles Fleury and Nikos Paragios. *Clustering of the Human Skeletal Muscle Fibers Using Linear Programming and Angular Hilbertian Metrics.* In IPMI '09: Proceedings of the 21th International Conference on Information Processing in Medical Imaging, pp. 14–25, 2009.

  - Yangming Ou, Ahmed Besbes, Michel Bilello, Mohamed Mansour, Christos Davatzikos and Nikos Paragios. *Detecting Landmark Pairs by Mutual-Saliency and MRF Regularization.* In ISBI '10: Proceedings of the 7th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2010.

- **Invited Publication:**

  - Nikos Komodakis, Ahmed Besbes, Ben Glocker and Nikos Paragios. *Biomedical Image Analysis Using Markov Random Fields & Efficient Linear Programming.*In EMBC'09: International Conference of the IEEE Engineering in Medicine and Biology Society, 2009.invited publication.

- **Technical Report:**

  - Ahmed Besbes, Nikos Paragios and Nikos Komodakis. *Cue Integration and Discrete MRFs towards Knowledge-based Segmentation and Tracking.*Research report 6831, Equipe GALEN, INRIA Saclay - Ile-de-France, 2009.

Bibliography

[Abi Nahed *et al.* 2006] Julien Abi Nahed, Marie-Pierre Jolly and Guang-Zhong Yang. *Robust Active Shape Models: A Robust, Generic and Simple Automatic Segmentation Tool.* In MICCAI '06: Proceedings of the 9th International Conference on Medical Image Computing and Computer Assisted Intervention, 2006. 71

[Alcantara *et al.* 2009] Dan A. Alcantara, Owen Carmichael, Will Harcourt-Smith, Kirstin Sterner, Stephen R. Frost, Rebecca Dutton, Paul Thompson, Eric Delson and Nina Amenta. *Exploration of Shape Variation Using Localized Components Analysis.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 8, pp. 1510–1516, 2009. 9

[Ali *et al.* 2007] Asem M. Ali, Aly A. Farag and Ayman S. El-Baz. *Graph Cuts Framework for Kidney Segmentation with Prior Shape Constraints.* In MICCAI '07: Proceedings of the 10th International Conference on Medical Image Computing and Computer Assisted Intervention, pp. 384–392, 2007. 79, 82

[Andreopoulos & Tsotsos 2008] Alexander Andreopoulos and John K. Tsotsos. *Efficient and Generalizable Statistical Models of Shape and Appearance for Analysis of Cardiac MRI.* Medical Image Analysis, vol. 12, no. 3, pp. 335–357, 2008. 73

[Arulampalam *et al.* 2002] M. Sanjeev Arulampalam, Simon Maskell, Neil Gordon and Tim Clapp. *A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking.* IEEE Transactions on Signal Processing, vol. 50, no. 2, pp. 174–188, 2002. 133

[Aurenhammer 1991] Franz Aurenhammer. *Voronoi Diagrams - A Survey of a Fundamental Geometric Data Structure.* ACM Computing Surveys, vol. 23, no. 3, pp. 345–405, 1991. 88

[Bardinet *et al.* 1995] Eric Bardinet, Laurent D. Cohen and Nicholas Ayache. *Superquadrics and Free-Form Deformations: A Global Model to Fit*

and Track 3D Medical Data. In CVRMed '95: Proceedings of the First International Conference on Computer Vision, Virtual Reality and Robotics in Medicine, pp. 319–326, 1995. 14

[Bardinet *et al.* 1998] Eric Bardinet, Laurent D. Cohen and Nicholas Ayache. *A Parametric Deformable Model to Fit Unstructured 3D Data.* Computer Vision and Image Understanding, vol. 71, no. 1, pp. 39–54, 1998. 14

[Behiels *et al.* 1999] Gert Behiels, Dirk Vandermeulen, Frederik Maes, Paul Suetens and Piet Dewaele. *Active Shape Model-Based Segmentation of Digital X-ray Images.* In MICCAI '99: Proceedings of the 2nd International Conference on Medical Image Computing and Computer Assisted Intervention, pp. 128–137, 1999. 71

[Belongie *et al.* 2002] Serge Belongie, Jitendra Malik and Jan Puzicha. *Shape Matching and Object Recognition Using Shape Contexts.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 24, pp. 509–522, 2002. 16

[Ben Ayed *et al.* 2009] Ismail Ben Ayed, Kumaradevan Punithakumar, Shuo Li, Ali Islam and Jaron Chong. *Left Ventricle Segmentation via Graph Cut Distribution Matching.* In MICCAI '09: Proceedings of the 12th International Conference on Medical Image Computing and Computer Assisted Intervention, 2009. 81, 82

[Bertrand 1994] Gilles Bertrand. *Simple Points, Topological Numbers and Geodesic Neighborhoods in Cubic Grids.* Pattern Recognition Letters, vol. 15, no. 10, pp. 1003–1011, 1994. 78

[Bertsekas 1999] Dimitri P. Bertsekas. Nonlinear Programming. Athena Scientific, 1999. 64

[Besag 1986] Julian Besag. *On the Statistical Analysis of Dirty Pictures.* Journal of the Royal Statistical Society, vol. 48, no. 3, pp. 259–302, 1986. 55

[Bilbro *et al.* 1989] Griff Bilbro, Reinhold Mann, Thomas K. Miller, Wesley E. Snyder, David E. Van den Bout and Mark White. *Optimization by Mean Field Annealing.* In NIPS '89: Advances in Neural Information Processing Systems 2, pp. 91–98, 1989. 56

[Black & Jepson 1998] Michael J. Black and Allan D. Jepson. *EigenTracking: Robust Matching and Tracking of Articulated Objects Using a View-Based Representation.* International Journal of Computer Vision, vol. 26, no. 1, pp. 63–84, 1998. 133

[Blake *et al.* 2004] Andrew Blake, Carsten Rother, Matthew Brown, Patrick Pérez and Philip Torr. *Interactive Image Segmentation Using an Adaptive GMMRF Model.* In ECCV '04: Proceedings of the 8th European Conference on Computer Vision, pp. 428–441, 2004. 74

[Blum 1973] Harry Blum. *Biological Shape and Visual Science.* Theoretical Biology, vol. 38, no. 2, pp. 205–287, 1973. 14

[Bookstein 1989] Fred L. Bookstein. *Principal Warps: Thin-Plate Splines and the Decomposition of Deformations.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 11, no. 6, pp. 567–585, 1989. 102

[Boykov & Funka-Lea 2006] Yuri Boykov and Gareth Funka-Lea. *Graph Cuts and Efficient N-D Image Segmentation.* International Journal of Computer Vision, vol. 70, no. 2, pp. 109–131, 2006. 55, 66, 74

[Boykov & Jolly 2001] Yuri Boykov and Marie-Pierre Jolly. *Interactive Graph Cuts for Optimal Boundary and Region Segmentation of Objects in N-D Images.* In ICCV '01: Proceedings of the 8th IEEE International Conference on Computer Vision, pp. 105–112, 2001. 57, 73, 74, 86

[Boykov & Kolmogorov 2004] Yuri Boykov and Vladimir Kolmogorov. *An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 9, pp. 1124–1137, 2004. 58

[Boykov *et al.* 1998] Yuri Boykov, Olga Veksler and Ramin Zabih. *Markov Random Fields with Efficient Approximations.* In CVPR '98: Proceedings of the 1998 Conference on Computer Vision and Pattern Recognition, pp. 648–655, 1998. 59

[Boykov *et al.* 1999] Yuri Boykov, Olga Veksler and Ramin Zabih. *Fast Approximate Energy Minimization via Graph Cuts.* In ICCV '99: Proceedings of the 7th IEEE International Conference on Computer Vision, pp. 377–384, 1999. 57

[Boykov *et al.* 2001] Yuri Boykov, Olga Veksler and Ramin Zabih. *Fast Approximate Energy Minimization via Graph Cuts.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 11, pp. 1222–1239, 2001. 59, 93

[Bugeau & Pérez 2009] Aurélie Bugeau and Patrick Pérez. *Detection and segmentation of moving objects in complex scenes.* Computer Vision and Image Understanding, vol. 113, no. 4, pp. 459–476, 2009. 134

[Caetano *et al.* 2006] Tibério S. Caetano, Terry Caelli, Dale Schuurmans and Dante A.C. Barone. *Graphical Models and Point Pattern Matching.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 10, pp. 1646–1663, 2006. 21, 22, 23, 32, 33, 34

[Cai & Maffray 1993] Leizhen Cai and Frédéric Maffray. *On the SPANNING k-TREE problem.* Discrete Applied Mathematics, vol. 44, no. 1–3, pp. 139–156, 1993. 33

[Caselles *et al.* 1997] Vicent Caselles, Ron Kimmel and Guillermo Sapiro. *Geodesic Active Contours.* International Journal of Computer Vision, vol. 22, no. 1, pp. 61–79, 1997. 67

[Chakraborty *et al.* 1996] Amit Chakraborty, Lawrence H. Staib and James S. Duncan. *Deformable Boundary Finding in Medical Images by Integrating Gradient and Region Information.* IEEE Transactions on Medical Imaging, vol. 15, no. 6, pp. 859–870, 1996. 15, 67, 89

[Chan & Vese 2001] Tony F. Chan and Luminita A. Vese. *Active Contours without Edges*. IEEE Transactions on Image Processing, vol. 10, no. 2, pp. 266–277, 2001. 67

[Chan & Zhu 2005] Tony Chan and Wei Zhu. *Level Set Based Shape Prior Segmentation*. In CVPR '05: Proceedings of the 2005 Conference on Computer Vision and Pattern Recognition, pp. 1164–1170, 2005. 12, 80

[Charpiat *et al.* 2007] Guillaume Charpiat, Olivier Faugeras and Renaud Keriven. *Shape Statistics for Image Segmentation with Prior*. In CVPR '07: Proceedings of the 2007 Conference on Computer Vision and Pattern Recognition, pp. 1–6, 2007. 11

[Chen *et al.* 2002] Yunmei Chen, Hemant Tagare, Sheshadri Thiruvenkadam, Feng Huang, David Clifford Wilson, Kaundinya S. Gopinath, Richard W. Briggs and Edward A. Geiser. *Using Prior Shapes in Geometric Active Contours in a Variational Framework*. International Journal of Computer Vision, vol. 50, no. 3, pp. 315–328, 2002. 17

[Chennubhotla & Jepson 2001] Chakra Chennubhotla and Allan Jepson. *Sparse PCA Extracting Multi-scale Structure from Data*. In ICCV '01: Proceedings of the 8th IEEE International Conference on Computer Vision, pp. 641–647, 2001. 9

[Chou & Brown 1990] Paul B. Chou and Christopher M. Brown. *The Theory and Practice of Bayesian Image Labeling*. International Journal of Computer Vision, vol. 4, no. 3, pp. 185–210, 1990. 56

[Chui & Rangarajan 2000] Haili Chui and Anand Rangarajan. *A New Algorithm for Non-Rigid Point Matching*. In CVPR '00: Proceedings of the 2000 Conference on Computer Vision and Pattern Recognition, pp. 44–51, 2000. 71

[Cohen 1991] Laurent D. Cohen. *On Active Contour Models and Balloons*. CVGIP: Image Understanding, vol. 53, no. 2, pp. 211–218, 1991. 67

[Coifman & Lafon 2006] Ronald R. Coifman and Stéphane Lafon. *Diffusion Maps*. Applied and Computational Harmonic Analysis, vol. 21, no. 1, pp. 5–30, 2006. 13, 19, 34

[Coifman & Maggioni 2006] Ronald R. Coifman and Mauro Maggioni. *Diffusion Wavelets*. Applied and Computational Harmonic Analysis, vol. 21, pp. 53–94, 2006. 15

[Collins 2003] Robert T. Collins. *Mean-shift Blob Tracking through Scale Space*. In CVPR '03: Proceedings of the 2003 Conference on Computer Vision and Pattern Recognition, volume 2, pp. 234–240, 2003. 133

[Comaniciu & Meer 2002] Dorin Comaniciu and Peter Meer. *Mean Shift: A Robust Approach Toward Feature Space Analysis*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 5, pp. 603–619, 2002. 70, 133

[Cook & Lee 1999] R. Dennis Cook and Hakbae Lee. *Dimension-Reduction in Binary Response Regression*. Journal of the American Statistical Association, vol. 94, no. 448, pp. 1187–1200, 1999. 106

[Cooper *et al.* 1991] David H. Cooper, Christopher J. Taylor, Jim Graham and Tim F. Cootes. *Locating Overlapping Flexible Shapes Using Geometrical Constraints*. In BMVC '91: Proceedings of the British Machine Vision Conference, pp. 185–194, 1991. 20

[Cootes & Taylor 1992] Tim F. Cootes and Christopher J. Taylor. *Active Shape Models: Smart Snakes*. In BMVC '92: Proceedings of the British Machine Vision Conference, 1992. 68

[Cootes & Taylor 1996] Tim F. Cootes and Christopher J. Taylor. *Data Driven Refinement of Active Shape Model Search*. In BMVC '96: Proceedings of the British Machine Vision Conference, 1996. 10

[Cootes & Taylor 1999] Tim F. Cootes and Christopher J. Taylor. *A Mixture Model for Representing Shape Variation*. Image and Vision Computing, vol. 17, no. 8, pp. 567–574, 1999. 11

[Cootes *et al.* 1992] Tim F. Cootes, David H. Cooper, Christopher J. Taylor and Jim Graham. *Trainable Method of Parametric Shape Description.* Image and Vision Computing, vol. 10, no. 5, pp. 289–294, 1992. 20, 21

[Cootes *et al.* 1995] Tim F. Cootes, Christopher J. Taylor, David H. Cooper and Jim Graham. *Active Shape Models - Their Training and Application.* Computer Vision and Image Understanding, vol. 61, no. 1, pp. 38–59, 1995. 8, 10, 20

[Cootes *et al.* 2001] Tim F. Cootes, Gareth J. Edwards and Christopher J. Taylor. *Active Appearance Models.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 6, pp. 681–685, 2001. 9, 72, 99

[Coughlan & Ferreira 2002] James M. Coughlan and Sabino J. Ferreira. *Finding Deformable Shapes Using Loopy Belief Propagation.* In ECCV '02: Proceedings of the 7th European Conference on Computer Vision, 2002. 16

[Coughlan *et al.* 2000] James Coughlan, Alan Yuille, Camper English and Dan Snow. *Efficient Deformable Template Detection and Localization without User Initialization.* Computer Vision and Image Understanding, vol. 78, no. 3, pp. 303–319, 2000. 15

[Crandall *et al.* 2005] David Crandall, Pedro Felzenszwalb and Daniel Huttenlocher. *Spatial Priors for Part-Based Recognition Using Statistical Models.* In CVPR '05: Proceedings of the 2005 Conference on Computer Vision and Pattern Recognition, pp. 10–17, 2005. 25, 40

[Cremers & Grady 2006] Daniel Cremers and Leo Grady. *Statistical Priors for Efficient Combinatorial Optimization Via Graph Cuts.* In ECCV '06: Proceedings of the 9th European Conference on Computer Vision, volume 3, pp. 263–274, 2006. 29

[Cremers *et al.* 2001] Daniel Cremers, Christoph Schnorr and Joachim Weickert. *Diffusion-Snakes: Combining Statistical Shape Knowledge and Image Information in a Variational Framework.* In Proceedings of the

1st International Workshop on Variational, Geometric, and Level Set Methods in Computer Vision, pp. 137–144, 2001. 10, 11

[Cremers *et al.* 2002a] Daniel Cremers, Timo Kohlberger and Christoph Schnörr. *Nonlinear Shape Statistics in Mumford–Shah Based Segmentation.* In ECCV '02: Proceedings of the 7th European Conference on Computer Vision, pp. 93–108, 2002. 11

[Cremers *et al.* 2002b] Daniel Cremers, Florian Tischhäuser, Joachim Weickert and Christoph Schnörr. *Diffusion Snakes: Introducing Statistical Shape Knowledge into the Mumford-Shah Functional.* International Journal of Computer Vision, vol. 50, no. 3, pp. 295–313, 2002. 10

[Cremers *et al.* 2007] Daniel Cremers, Mikaël Rousson and Rachid Deriche. *A Review of Statistical Approaches to Level Set Segmentation: Integrating Color, Texture, Motion and Shape.* International Journal of Computer Vision, vol. 72, no. 2, pp. 195–215, 2007. 7, 12, 66

[Cremers 2008] Daniel Cremers. *Nonlinear Dynamical Shape Priors for Level Set Segmentation.* Journal of Scientific Computing, vol. 35, no. 2-3, pp. 132–143, 2008. 133

[Das *et al.* 2009] Piali Das, Olga Veksler, Vyacheslav Zavadsky and Yuri Boykov. *Semiautomatic Segmentation with Compact Shape Prior.* Image and Vision Computing, vol. 27, no. 1-2, pp. 206–219, 2009. 77

[Dattorro 2005] Jon Dattorro. Convex Optimization & Euclidean Distance Geometry. Meboo Publishing USA, 2005. 20, 31

[Davatzikos *et al.* 2003] Christos Davatzikos, Xiaodong Tao and Dinggang Shen. *Hierarchical Active Shape Models, Using the Wavelet Transform.* IEEE Transactions on Medical Imaging, vol. 22, no. 3, pp. 414–423, March 2003. 15

[Davies *et al.* 2002] Rhodri H. Davies, Carole J. Twining, Tim F. Cootes, John C. Waterton and Christopher J. Taylor. *A Minimum Description Length Approach to Statistical Shape Modeling.* IEEE Transactions on Medical Imaging, vol. 21, no. 5, pp. 525–537, 2002. 8

[de Bruijne & Nielsen 2004] Marleen de Bruijne and Mads Nielsen. *Shape Particle Filtering for Image Segmentation.* In MICCAI '04: Proceedings of the 7th International Conference on Medical Image Computing and Computer Assisted Intervention, pp. 168–175, 2004. 70

[de Bruijne & Nielsen 2005] Marleen de Bruijne and Mads Nielsen. *Multi-object Segmentation Using Shape Particles.* In IPMI '05: Proceedings of the 19th International Conference on Information Processing in Medical Imaging, pp. 762–773, 2005. 70

[de Bruijne *et al.* 2003] Marleen de Bruijne, Bram van Ginneken, Wiro J. Niessen, Marco Loog and Max A. Viergever. *Model-based Segmentation of Abdominal Aortic Aneurysms in CTA Images.* In CTA images. Proceedings of SPIE Medical Imaging, Image Processing, 2003. 69, 70

[Delingette 1999] Hervé Delingette. *General Object Reconstruction Based on Simplex Meshes.* International Journal of Computer Vision, vol. 32, no. 2, pp. 111–146, September 1999. 67

[Donner *et al.* 2006] René Donner, Michael Reiter, Georg Langs, Philipp Peloschek and Horst Bischof. *Fast Active Appearance Model Search Using Canonical Correlation Analysis.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 10, pp. 1690–1694, 2006. 73

[Donner *et al.* 2007a] René Donner, Branislav Micušík, Georg Langs and Horst Bischof. *Sparse MRF Appearance Models for Fast Anatomical Structure Localisation.* In BMVC '07: Proceedings of the British Machine Vision Conference, 2007. 104, 105, 110, 111

[Donner *et al.* 2007b] René Donner, Branislav Micušík, Georg Langs, Lech Szumilas, Philipp Peloschek, Klaus Friedrich and Horst Bischof. *Object Localization Based on Markov Random Fields and Symmetry Interest Points.* In MICCAI '07: Proceedings of the 10th International Conference on Medical Image Computing and Computer Assisted Intervention, pp. 460–468, 2007. 84

[Donner *et al.* 2009] René Donner, Horst Wildenauer, Horst Bischof and Georg Langs. *Weakly Supervised Group-Wise Model Learning Based on Discrete Optimization.* In International Conference on Medical Image Computing and Computer Assisted Intervention, 2009. 84

[Donner *et al.* 2010] René Donner, Georg Langs, Branislav Micušík and Horst Bischof. *Generalized Sparse MRF Appearance Models.* Image and Vision Computing, vol. 28, no. 6, pp. 1031–1038, 2010. 26, 84, 85, 86, 104

[Essafi *et al.* 2009] Salma Essafi, Georg Langs and Nikos Paragios. *Hierarchical 3D Diffusion Wavelet Shape Priors.* In ICCV '09: Proceedings of the 12th IEEE International Conference on Computer Vision, pp. 1717–1724, 2009. 15

[Etyngier *et al.* 2007] Patrick Etyngier, Florent Ségonne and Renaud Keriven. *Shape Priors using Manifold Learning Techniques.* In ICCV '07: Proceedings of the 11th IEEE International Conference on Computer Vision, pp. 1–8, 2007. 13

[Felzenszwalb & Huttenlocher 2000] Pedro F. Felzenszwalb and Daniel P. Huttenlocher. *Efficient Matching of Pictorial Structures.* In CVPR '00: Proceedings of the 2000 Conference on Computer Vision and Pattern Recognition, pp. 66–73, 2000. 17

[Felzenszwalb & Huttenlocher 2005] Pedro F. Felzenszwalb and Daniel P. Huttenlocher. *Pictorial Structures for Object Recognition.* International Journal of Computer Vision, vol. 61, no. 1, pp. 55–79, 2005. 17, 24, 25, 43, 83, 133

[Felzenszwalb 2005] Pedro F. Felzenszwalb. *Representation and Detection of Deformable Shapes.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 2, pp. 208–220, 2005. 23, 24, 83, 85

[Fischler & Bolles 1981] Martin A. Fischler and Robert C. Bolles. *Random Sample Consensus: a Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography.* Commumications of the ACM, vol. 24, no. 6, pp. 381–395, 1981. 71

[Fischler & Elschlager 1973] Martin A. Fischler and Robert A. Elschlager. *The Representation and Matching of Pictorial Structures.* IEEE Transactions on Computers, vol. 22, no. 1, pp. 67–92, 1973. 16, 17

[Fletcher *et al.* 2004] P. Thomas Fletcher, Conglin Lu, Stephen M. Pizer and Sarang Joshi. *Principal Geodesic Analysis for the Study of Nonlinear Statistics of Shape.* IEEE Transactions on Medical Imaging, vol. 23, no. 8, pp. 995–1005, 2004. 14

[Ford & Fulkerson 1962] Lester R. Ford and Delbert R. Fulkerson. Flows in Networks. Princeton University Press, 1962. 58

[Freedman & Zhang 2005] Daniel Freedman and Tao Zhang. *Interactive Graph Cut Based Segmentation with Shape Priors.* In CVPR '05: Proceedings of the 2005 Conference on Computer Vision and Pattern Recognition, pp. 755–762, 2005. 78, 82

[Freeman & Adelson 1991] William T. Freeman and Edward H. Adelson. *The Design and Use of Steerable Filters.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 13, no. 9, pp. 891–906, 1991. 69

[Freund & Schapire 1996] Yoav Freund and Robert E. Schapire. *Experiments with a New Boosting Algorithm.* In Proceedings of the 13th International Conference on Machine Learning, 1996. 106, 107

[Freund & Schapire 1997] Yoav Freund and Robert E. Schapire. *A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting.* Journal of Computer and System Sciences, vol. 55, no. 1, pp. 119–139, 1997. 70

[Funka-Lea *et al.* 2006] Gareth Funka-Lea, Yuri Boykov, Charles Florin, Marie-Pierre Jolly, Romain Moreau-Gobard, Rana Ramaraj and Daniel Rinck. *Automatic Heart Isolation for CT Coronary Visualization Using Graph-Cuts.* In ISBI '06: Proceedings of the 2006 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 614–617, 2006. 77

[Geman & Geman 1984] Stuart Geman and Donald Geman. *Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 6, no. 6, pp. 721–741, 1984. 55

[Glocker *et al.* 2007] Ben Glocker, Nikos Komodakis, Nikos Paragios, Georgios Tziritas and Nassir Navab. *Inter and Intra-Modal Deformable Registration: Continuous Deformations Meet Efficient Optimal Linear Programming*. In IPMI '07: Proceedings of the 20th International Conference on Information Processing in Medical Imaging, 2007. 86, 91

[Glocker *et al.* 2008] Ben Glocker, Nikos Komodakis, Georgios Tziritas, Nassir Navab and Nikos Paragios. *Dense Image Registration through MRFs and Efficient Linear Programming*. Medical Image Analysis, vol. 12, no. 6, pp. 731–741, 2008. 86

[Glocker *et al.* 2009] Ben Glocker, Nikos Komodakis, Nassir Navab, Georgios Tziritas and Nikos Paragios. *Dense Registration with Deformation Priors*. In IPMI '09: Proceedings of the 21th International Conference on Information Processing in Medical Imaging, 2009. 26, 28

[Goldberg & Tarjan 1988] Andrew V. Goldberg and Robert E. Tarjan. *A New Approach to the Maximum-Flow Problem*. Journal of the ACM, vol. 35, no. 4, pp. 921–940, 1988. 58

[Grady *et al.* 2005] Leo Grady, Yiyong Sun and James Williams. Handbook of Mathematical Models of Computer Vision, chapter: Interactive Graph-Based Segmentation Methods in Cardiovascular Imaging, pp. 453–469. Springer Verlag, 2005. 48

[Grady 2005] Leo Grady. *Multilabel Random Walker Image Segmentation Using Prior Models*. In CVPR '05: Proceedings of the 2005 Conference on Computer Vision and Pattern Recognition, pp. 763–770, 2005. 75

[Grady 2006] Leo Grady. *Random Walks for Image Segmentation*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 11, pp. 1768–1783, 2006. 49, 75, 101, 103

[Grady 2010] Leo Grady. *Minimal Surfaces Extend Shortest Path Segmentation Methods to 3D.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 2, pp. 321–334, 2010. 83

[Greig *et al.* 1989] D. Greig, B. Porteous and A. Seheult. *Exact Maximum a Posterori Estimation for Binary Images.* Journal of the Royal Statistical Society, vol. 51, no. 2, pp. 271–279, 1989. 56, 57

[Gross *et al.* 2006] Ralph Gross, Iain Matthews and Simon Baker. *Active Appearance Models with Occlusion.* Image and Vision Computing, vol. 24, no. 1, pp. 593–604, 2006. 73, 99

[Gu *et al.* 2007] Lie Gu, Eric P. Xing and Takeo Kanade. *Learning GMRF Structures for Spatial Priors.* In CVPR '07: Proceedings of the 2007 Conference on Computer Vision and Pattern Recognition, pp. 1–6, 2007. 26

[Hammer *et al.* 1984] P. L. Hammer, P. Hansen and B. Simeone. *Roof Duality, Complementation and Persistency in Quadratic 0-1 Optimization.* Mathematical Programming, vol. 28, pp. 121–155, 1984. 58

[Heap & Hogg 1997] Tony Heap and David Hogg. *Improving Specificity in PDMs using a Hierarchical Approach.* In BMVC '97: Proceedings of the British Machine Vision Conference, 1997. 10

[Heimann & Meinzer 2009] Tobias Heimann and Hans-Peter Meinzer. *Statistical Shape Models for 3D Medical Image Segmentation: A Review.* Medical Image Analysis, vol. 13, no. 4, pp. 543–563, 2009. 7, 66

[Hong *et al.* 2006] Byung-Woo Hong, Emmanuel Prados, Stefano Soatto and Luminita Vese. *Shape Representation based on Integral Kernels: Application to Image Matching and Segmentation.* In CVPR '06: Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition, pp. I: 833–840, 2006. 16

[Hyvärinen *et al.* 2001] Aapo Hyvärinen, Juha Karhunen and Erkki Oja. Independent component analysis. John Wiley & Sons, Inc., 2001. 9

[Isard & Blake 1998] Michael Isard and Andrew Blake. *Condensation – Conditional Density Propagation for Visual Tracking.* International Journal of Computer Vision, vol. 29, no. 1, pp. 5–28, 1998. 70, 133

[Ishikawa & Jermyn 2001] Hiroshi Ishikawa and Ian H. Jermyn. *Region Extraction from Multiple Images.* In CVPR '01: Proceedings of the 2001 Conference on Computer Vision and Pattern Recognition, 2001. 81

[Ishikawa 2003] Hiroshi Ishikawa. *Exact Optimization for Markov Random Fields with Convex Priors.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 10, pp. 1333–1336, 2003. 59

[Jain *et al.* 1998] Anil K. Jain, Yu Zhong and Marie-Pierre Dubuisson-Jolly. *Deformable Template Models: A Review.* Signal Processing, vol. 71, no. 2, pp. 109–129, 1998. 66

[Jermyn & Ishikawa 2001] Ian H. Jermyn and Hiroshi Ishikawa. *Globally Optimal Regions and Boundaries as Minimum Ratio Weight Cycles.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 10, pp. 1075–1088, 2001. 81

[Jiao *et al.* 2003] Feng Jiao, Stan Li, Heung-Yeung Shum and Dale Schuurmans. *Face Alignment Using Statistical Models and Wavelet Features.* In CVPR '03: Proceedings of the 2003 Conference on Computer Vision and Pattern Recognition, pp. 321–327, 2003. 69

[Joshi & Miller 2000] Sarang C. Joshi and Michael I. Miller. *Landmark Matching via Large Deformation Diffeomorphisms.* IEEE Transactions on Image Processing, vol. 9, no. 8, pp. 1357–1370, 2000. 124

[Kadir & Brady 2001] Timor Kadir and Michael Brady. *Saliency, Scale and Image Description.* International Journal of Computer Vision, vol. 45, no. 2, pp. 83–105, 2001. 124, 126

[Kalman 1960] Rudolph E. Kalman. *A New Approach to Linear Filtering and Prediction Problems.* Transactions of the ASME – Journal of Basic Engineering, vol. 82, no. Series D, pp. 35–45, 1960. 133

[Kass *et al.* 1988] Michael Kass, Andrew Witkin and Demetri Terzopoulos. *Snakes: Active contour models.* International Journal of Computer Vision, vol. 1, no. 4, pp. 321–331, 1988. 66

[Kelemen *et al.* 1999] András Kelemen, Gábor Székely and Guido Gerig. *Elastic Model-Based Segmentation of 3-D Neuroradiological Data Sets.* IEEE Transactions on Medical Imaging, vol. 18, no. 10, pp. 828–839, 1999. 15

[Kichenassamy *et al.* 1995] Satyanad Kichenassamy, Arun Kumar, Peter Olver, Allen Tannenbaum and Anthony Yezzi. *Gradient Flows and Geometric Active Contour Models.* In ICCV '95: Proceedings of the 5th IEEE International Conference on Computer Vision, 1995. 67

[Kirkpatrick *et al.* 1983] Scott Kirkpatrick, C. Daniel Gelatt Jr. and Mario P. Vecchi. *Optimization by Simulated Annealing.* Science, vol. 220, no. 4598, pp. 671–680, 1983. 56

[Koenderink & Van Doorn 1999] Jan J. Koenderink and Andrea J. Van Doorn. *The Structure of Locally Orderless Images.* International Journal of Computer Vision, vol. 31, no. 2-3, pp. 159–168, 1999. 69, 106

[Kohli *et al.* 2008] Pushmeet Kohli, Jonathan Rihan, Matthieu Bray and Philip H.S. Torr. *Simultaneous Segmentation and Pose Estimation of Humans Using Dynamic Graph Cuts.* International Journal of Computer Vision, vol. 79, pp. 285–298, 2008. 134

[Kokkinos 2010] Iasonas Kokkinos. *Highly Accurate Boundary Detection and Grouping.* In CVPR '10: Proceedings of the 2010 Conference on Computer Vision and Pattern Recognition, 2010. 105

[Kolmogorov & Rother 2007] Vladimir Kolmogorov and Carsten Rother. *Minimizing Nonsubmodular Functions with Graph Cuts–A Review.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 7, pp. 1274–1279, 2007. 58

[Kolmogorov & Wainwright 2005] Vladimir Kolmogorov and Martin J. Wainwright. *On the Optimality of Tree-reweighted Max-product Message-passing*. In UAI '05: Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence, pp. 316–323, 2005. 64

[Kolmogorov & Zabih 2004] Vladimir Kolmogorov and Ramin Zabih. *What Energy Functions Can Be Minimized via Graph Cuts?* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 2, pp. 147–159, 2004. 58

[Kolmogorov 2006] Vladimir Kolmogorov. *Convergent Tree-Reweighted Message Passing for Energy Minimization*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 10, pp. 1568–1583, 2006. 25, 45, 64, 93, 110, 129

[Komodakis & Paragios 2008] Nikos Komodakis and Nikos Paragios. *Beyond Loose LP-Relaxations: Optimizing MRFs by Repairing Cycles*. In ECCV '08: Proceedings of the 10th European Conference on Computer Vision, volume 3, pp. 806–820, 2008. 65

[Komodakis & Tziritas 2007] Nikos Komodakis and Georgios Tziritas. *Approximate Labeling via Graph Cuts Based on Linear Programming*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 8, pp. 1436–1435, 2007. 62, 63

[Komodakis *et al.* 2007] Nikos Komodakis, Nikos Paragios and Georgios Tziritas. *MRF Optimization via Dual Decomposition: Message-Passing Revisited*. In ICCV '07: Proceedings of the 11th IEEE International Conference on Computer Vision, pp. 1–8, 2007. 64

[Komodakis *et al.* 2008a] Nikos Komodakis, Nikos Paragios and Georgios Tziritas. *Clustering via LP-based Stabilities*. In NIPS '08: Advances in Neural Information Processing Systems 21, 2008. 19, 26, 37, 39

[Komodakis *et al.* 2008b] Nikos Komodakis, Georgios Tziritas and Nikos Paragios. *Performance vs Computational Efficiency for Optimizing Single and Dynamic MRFs: Setting the State of the Art with Primal*

Dual Strategies. Computer Vision and Image Understanding, vol. 112, no. 1, pp. 14–29, 2008. 25, 45, 63, 94, 111, 137

[Kschischang *et al.* 2001] Frank R. Kschischang, Brendan J. Frey and Hans-Andrea Loeliger. *Factor Graphs and the Sum-Product Algorithm*. IEEE Transactions on Information Theory, vol. 47, no. 2, pp. 498–519, 2001. 60

[Kumar *et al.* 2006] M. Pawan Kumar, Philip H. S. Torr and Andrew Zisserman. *Solving Markov Random Fields using Second Order Cone Programming Relaxations*. In CVPR '06: Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition, pp. 1045–1052, 2006. 64

[Kumar *et al.* 2009] M. Pawan Kumar, Vladimir Kolmogorov and Philip H. S. Torr. *An Analysis of Convex Relaxations for MAP Estimation of Discrete MRFs*. Journal of Machine Learning Research, vol. 10, pp. 71–106, 2009. 64

[Kumar *et al.* 2010] M. Pawan Kumar, Philip H.S. Torr and Andrew Zisserman. *OBJCUT: Efficient Segmentation Using Top-Down and Bottom-Up Cues*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 3, pp. 530–545, 2010. 84, 85, 96

[Langs & Paragios 2008] Georg Langs and Nikos Paragios. *Modeling the Structure of Multivariate Manifolds: Shape Maps*. In CVPR '08: Proceedings of the 2008 Conference on Computer Vision and Pattern Recognition, 2008. 19, 28, 34, 35, 36, 97

[Langs *et al.* 2006] Georg Langs, René Donner, Michael Reiter and Horst Bischof. *Active Feature Models*. In 18th International Conference on Pattern Recognition, pp. 417–420, 2006. 69

[Leordeanu & Hebert 2005] Marius Leordeanu and Martial Hebert. *A Spectral Technique for Correspondence Problems Using Pairwise Constraints*. In ICCV '05: Proceedings of the 10th IEEE International Conference on Computer Vision, pp. 1482–1489, 2005. 129

[Leventon *et al.* 2000] Michael E. Leventon, W. Eric L. Grimson and Olivier Faugeras. *Statistical Shape Influence in Geodesic Active Contours.* In CVPR '00: Proceedings of the 2000 Conference on Computer Vision and Pattern Recognition, pp. I: 316–323, 2000. 12

[Li & Ito 2005] Yuanzhong Li and Wataru Ito. *Shape Parameter Optimization for AdaBoosted Active Shape Model.* In ICCV '05: Proceedings of the 10th IEEE International Conference on Computer Vision, pp. 251–258, 2005. 70

[Li *et al.* 2004a] Shuyu Li, Litao Zhu and Tianzi Jiang. *Active Shape Model Segmentation Using Local Edge Structures and AdaBoost.* In Medical Imaging and Augmented Reality, 2004. 70

[Li *et al.* 2004b] Yin Li, Jian Sun, Chi-Keung Tang and Heung-Yeung Shum. *Lazy Snapping.* ACM Transactions on Graphics, vol. 23, no. 3, pp. 303–308, 2004. 74

[Li *et al.* 2006] Kang Li, Xiaodong Wu, D. Z. Chen and M. Sonka. *Optimal Surface Segmentation in Volumetric Images - A Graph-Theoretic Approach.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 1, pp. 119–134, 2006. 74

[Lin *et al.* 2005] Xiang Lin, Brett Cowan and Alistair Young. *Model-based Graph Cut Method for Segmentation of the Left Ventricle.* In EMBS '05: Proceedings of the 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 3059–3062, 2005. 78

[Liu *et al.* 2002] Jundong Liu, Baba C. Vemuri and José L. Marroquín. *Local frequency representations for robust multimodal image registration.* IEEE Transactions on Medical Imaging, vol. 21, no. 5, pp. 462–469, 2002. 126

[Lowe 2004] David G. Lowe. *Distinctive Image Features from Scale-Invariant Keypoints.* International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004. 105

[Malcolm *et al.* 2007a] James Malcolm, Yogesh Rathi Allen and Tannenbaum. *Graph Cut Segmentation with Nonlinear Shape Priors.* In ICIP '07: Proceedings of the 14th IEEE International Conference on Image Processing, pp. IV: 365–368, 2007. 79, 82

[Malcolm *et al.* 2007b] James Malcolm, Yogesh Rathi and Allen Tannenbaum. *Multi-object tracking through clutter using graph cuts.* In ICCV'07 Workshop on Non-rigid Registration and Tracking Through Learning, 2007. 134

[Malladi *et al.* 1995] Ravikanth Malladi, James A. Sethian and Baba C. Vemuri. *Shape Modeling with Front Propagation: A Level Set Approach.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 17, no. 2, pp. 158–175, 1995. 67

[Mallows & Clark 1970] C. L. Mallows and J. M. C. Clark. *Linear Intercept Distributions Do Not Characterize Plane Sets.* Journal of Applied Probability, vol. 7, pp. 240–244, 1970. 20

[Manay *et al.* 2006] Siddharth Manay, Daniel Cremers, Byung-Woo Hong, Anthony J. Yezzi and Stefano Soatto. *Integral Invariants for Shape Matching.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 10, pp. 1602–1618, 2006. 16

[Matthews & Baker 2004] Iain Matthews and Simon Baker. *Active Appearance Models Revisited.* International Journal of Computer Vision, vol. 60, no. 2, pp. 135–164, 2004. 73

[McInerney & Terzopoulos 1996] Tim McInerney and Demetri Terzopoulos. *Deformable Models in Medical Images Analysis: a Survey.* Medical Image Analysis, vol. 1, no. 2, pp. 91–108, 1996. 66

[McInerney & Terzopoulos 1999] Tim McInerney and Demetri Terzopoulos. *Topology Adaptive Deformable Surfaces for Medical Image Volume Segmentation.* IEEE Transactions on Medical Imaging, vol. 18, no. 10, pp. 840–850, 1999. 67

[Meinshausen & Bühlmann 2006] Nicolai Meinshausen and Peter Bühlmann. *High dimensional graphs and variable selection with the Lasso.* Annals of Statistics, vol. 34, pp. 1436–1462, 2006. 26

[Meir & Rätsch 2003] Ron Meir and Gunnar Rätsch. *An Introduction to Boosting and Leveraging.* In Advanced lectures on machine learning, pp. 118–183, 2003. 107

[Metaxas & Terzopoulos 1993] Dimitri Metaxas and Demetri Terzopoulos. *Shape and Nonrigid Motion Estimation Through Physics-Based Synthesis.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, no. 6, pp. 580–591, 1993. 13

[Montagnat *et al.* 1999] Johan Montagnat, Hervé Delingette and Grégoire Malandain. *Cylindrical Echocardiographic Image Segmentation Based on 3D Deformable Models.* In MICCAI '99: Proceedings of the 2nd International Conference on Medical Image Computing and Computer Assisted Intervention, pp. 168–175, 1999. 67

[Montagnat *et al.* 2001] Johan Montagnat, Hervé Delingette and Nicholas Ayache. *A Review of Deformable Surfaces: Topology. Geometry and Deformation.* Image and Vision Computing, vol. 19, no. 14, pp. 1023–1040, 2001. 66

[Montagnat *et al.* 2003] Johan Montagnat, Maxime Sermesant, Hervé Delingette, Grégoire Malandain and Nicholas Ayache. *Anisotropic filtering for model based segmentation of 4D cylindrical echocardiographic images.* Pattern Recognition Letters, vol. 24, pp. 815–828, 2003. 67

[Mumford & Shah 1989] David Mumford and Jayant Shah. *Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems.* Communications on Pure and Applied Mathematics, vol. 42, no. 5, pp. 577–685, 1989. 67

[Muramatsu & Suzuki 2003] Masakazu Muramatsu and Tsunehiro Suzuki. *A New Second-order Cone Programming Relaxation for MAX-CUT Problems.* Journal of Operations Research of Japan, vol. 46, no. 2, pp. 164–177, 2003. 64

[Nain *et al.* 2007] Delphine Nain, Steven Haker, Aaron Bobick and Allen Tannenbaum. *Multiscale 3-D Shape Representation and Segmentation Using Spherical Wavelets.* IEEE Transactions on Medical Imaging, vol. 26, no. 4, pp. 598–618, 2007. 15

[Osher & Sethian 1988] Stanley Osher and James A. Sethian. *Fronts Propagating with Curvature-Dependent Speed: Algorithms Based on Hamilton-Jacobi Formulations.* Journal of Computational Physics, vol. 79, pp. 12–49, 1988. 12, 67

[Ou & Davatzikos 2009] Yangming Ou and Christos Davatzikos. DRAMMS*: Deformable Registration via Attribute Matching and Mutual-Saliency Weighting.* In IPMI '09: Proceedings of the 21th International Conference on Information Processing in Medical Imaging, pp. 50–62, 2009. 124, 125, 126

[Paragios & Deriche 1999] Nikos Paragios and Rachid Deriche. *Unifying Boundary and Region-Based Information for Geodesic Active Tracking.* In CVPR '99: Proceedings of the 1999 Conference on Computer Vision and Pattern Recognition, pp. 2300–2305, 1999. 133

[Paragios & Deriche 2000] Nikos Paragios and Rachid Deriche. *Coupled Geodesic Active Regions for Image Segmentation: A Level Set Approach.* In ECCV '00: Proceedings of the 6th European Conference on Computer Vision, pp. 224–240, 2000. 67

[Pearl 1988] Judea Pearl. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, 1988. 59

[Pizer *et al.* 1999] Stephen M. Pizer, Daniel S. Fritsch, Paul A. Yushkevich, Valen E. Johnson and Edward L. Chaney. *Segmentation, Registration, and Measurement of Shape Variation via Image Object Shape.* IEEE Transactions on Medical Imaging, vol. 18, no. 10, pp. 851–865, 1999. 14

[Pizer *et al.* 2003] Stephen M. Pizer, Guido Gerig, Sarang Joshi and Stephen R. Aylward. *Multiscale Medial Shape-Based Analysis of Image*

Objects. Proceedings of the IEEE, vol. 91, no. 10, pp. 1670–1679, 2003. 14

[Pohl *et al.* 2006] Kilian M. Pohl, John Fisher, Martha Shenton, Robert W. McCarley, W. Eric L. Grimson, Ron Kikinis and William M. Wells. *Logarithm Odds Maps for Shape Representation*. In MICCAI '06: Proceedings of the 9th International Conference on Medical Image Computing and Computer Assisted Intervention, pp. 955–963, 2006. 13

[Pohl *et al.* 2007] Kilian M. Pohl, John Fisher, Sylvain Bouix, Martha Shenton, Robert W. McCarley, W. Eric L. Grimson, Ron Kikinis and William M. Wells. *Using the Logarithm of Odds to Define a Vector Space on Probabilistic Atlases*. Medical Image Analysis, vol. 11, no. 6, pp. 465–477, 2007. 13

[Qu *et al.* 2008] Wei Qu, Xiaolei Huang and Yuanyuan Jia. *Segmentation in Noisy Medical Images Using PCA Model Based Particle Filtering*. In SPIE Conference on Medical Imaging, 2008. 70

[Ravikumar & Lafferty 2006] Pradeep Ravikumar and John Lafferty. *Quadratic programming relaxations for metric labeling and Markov random field MAP estimation*. In ICML '06: Proceedings of the 23rd International Conference on Machine Learning, pp. 737–744, 2006. 64

[Rogers & Graham 2002] Mike Rogers and Jim Graham. *Robust Active Shape Model Search*. In ECCV '02: Proceedings of the 7th European Conference on Computer Vision, pp. 517–530, 2002. 71

[Rohlfing *et al.* 2005] Torsten Rohlfing, Robert Brandt, Randolf Menzel, Daniel B. Russakoff and Calvin R. Maurer Jr. The Handbook of Medical Image Analysis: Segmentation and Registration Models, chapter: Quo Vadis, Atlas-Based Segmentation?, pp. 435–486. Kluwer Academic / Plenum Publishers, 2005. 66

[Romdhani *et al.* 1999] Sami Romdhani, Shaogang Gong and Alexandra Psarrou. *Multi-View Nonlinear Active Shape Model Using Kernel PCA*. In BMVC '99: Proceedings of the British Machine Vision Conference, 1999. 11

[Rother *et al.* 2004] Carsten Rother, Vladimir Kolmogorov and Andrew Blake. *"GrabCut": Interactive Foreground Extraction using Iterated Graph Cuts.* ACM Transactions on Graphics, vol. 23, pp. 309–314, 2004. 74

[Rother *et al.* 2005] Carsten Rother, Sanjiv Kumar, Vladimir Kolmogorov and Andrew Blake. *Digital Tapestry.* In CVPR '05: Proceedings of the 2005 Conference on Computer Vision and Pattern Recognition, pp. 589–596, 2005. 59

[Rousson & Paragios 2002] Mikaël Rousson and Nikos Paragios. *Shape Priors for Level Set Representations.* In ECCV '02: Proceedings of the 7th European Conference on Computer Vision, pp. 78–92, 2002. 12

[Rousson & Paragios 2008] Mikaël Rousson and Nikos Paragios. *Prior Knowledge, Level Set Representations & Visual Grouping.* International Journal of Computer Vision, vol. 76, no. 3, pp. 231–243, 2008. 12

[Roy & Cox 1998] Sébastien Roy and Ingemar J. Cox. *A Maximum-Flow Formulation of the N-Camera Stereo Correspondence Problem.* In ICCV '98: Proceedings of the 6th IEEE International Conference on Computer Vision, p. 492, 1998. 59

[Schapire & Singer 1998] Robert E. Schapire and Yoram Singer. *Improved Boosting Algorithms Using Confidence-Rated Predictions.* In COLT'98: Proceedings of the 11th Annual Conference on Computational Learning Theory, pp. 80–91, 1998. 107

[Schapire 2002] Robert E. Schapire. *The Boosting Approach to Machine Learning: an Overview.* In Workshop on Nonlinear Estimation and Classification, 2002. 107

[Schlesinger 1976] Mikhail I. Schlesinger. *Sintaksicheskiy analiz dvumernykh zritelnikh signalov v usloviyakh pomekh (Syntactic Analysis of Two-Dimensional Visual Signals in Noisy Conditions).* Kibernetika, vol. 4, pp. 113–130, 1976. in Russian. 61, 84

[Schoenberg 1935] Isaac J. Schoenberg. *Remarks to Maurice Fréchet's Article "Sur la Définition Axiomatique d'une Classe d'Espace Distanciés Vectoriellement Applicable sur l'Espace de Hilbert"*. Annals of Mathematics, vol. 36, no. 3, pp. 724–732, 1935. 21

[Schoenemann & Cremers 2007] Thomas Schoenemann and Daniel Cremers. *Globally Optimal Image Segmentation with an Elastic Shape Prior*. In ICCV '07: Proceedings of the 11th IEEE International Conference on Computer Vision, pp. 1–6, 2007. 17, 80, 82

[Schölkopf *et al.* 1998] Bernhard Schölkopf, Alexander Smola and Klaus-Robert Müller. *Nonlinear Component Analysis as a Kernel Eigenvalue Problem*. Neural Computation, vol. 10, no. 5, pp. 1299–1319, 1998. 11

[Seghers *et al.* 2007a] Dieter Seghers, Dirk Loeckx, Frederik Maes, Dirk Vandermeulen and Paul Suetens. *Minimal Shape and Intensity Cost Path Segmentation*. IEEE Transactions on Medical Imaging, vol. 26, no. 8, pp. 1115–1129, 2007. 25, 83, 85, 86, 104, 105, 106, 113, 114

[Seghers *et al.* 2007b] Dieter Seghers, Pieter Slagmolen, Yves Lambelin, Jeroen Hermans, Dirk Loeckx, Frederik Maes and Paul Suetens. *Landmark Based Liver Segmentation Using Local Shape and Local Intensity Models*. In 3D Segmentation in The Clinic: A Grand Challenge, 2007. 25, 83

[Seghers *et al.* 2008] Dieter Seghers, Jeroen Hermans, Dirk Loeckx, Frederik Maes, Dirk Vandermeulen and Paul Suetens. *Model-Based Segmentation Using Graph Representations*. In MICCAI '08: Proceedings of the 11th International Conference on Medical Image Computing and Computer Assisted Intervention, pp. 393–400, 2008. 29, 84, 104, 111

[Shen & Davatzikos 2000] Dinggang Shen and Christos Davatzikos. *An Adaptive-Focus Deformable Model Using Statistical and Geometric Information*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 8, pp. 906–913, 2000. 16

[Shen *et al.* 2001] Dinggang Shen, Edward H. Herskovits and Christos Davatzikos. *An Adaptive-Focus Statistical Shape Model for Segmentation*

and Shape Modeling of 3-D Brain Structures. IEEE Transactions on Medical Imaging, vol. 20, no. 4, pp. 257–270, 2001. 16

[Shi & Malik 2000] Jianbo Shi and Jitendra Malik. *Normalized Cuts and Image Segmentation.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 8, pp. 888–905, 2000. 75

[Shi & Tomasi 1994] Jianbo Shi and Carlo Tomasi. *Good features to track.* In CVPR '94: Proceedings of the 1994 Conference on Computer Vision and Pattern Recognition, 1994. 133

[Shiraishi *et al.* 2000] Junji Shiraishi, Shigehiko Katsuragawa, Junpei Ikezoe, Tsuneo Matsumoto, Takeshi Kobayashi, Ken-ichi Komatsu, Mitate Matsui, Hiroshi Fujita, Yoshie Kodera and Kunio Doi. *Development of a Digital Image Database for Chest Radiographs With and Without a Lung Nodule: Receiver Operating Characteristic Analysis of Radiologists' Detection of Pulmonary Nodules.* American Journal of Roentgenology, vol. 174, pp. 71–74, 2000. 48, 112

[Singaraju *et al.* 2008] Dheeraj Singaraju, Leo Grady and René Vidal. *Interactive Image Segmentation of Quadratic Energies on Directed Graphs.* In CVPR '08: Proceedings of the 2008 Conference on Computer Vision and Pattern Recognition, 2008. 76

[Singaraju *et al.* 2009] Dheeraj Singaraju, Leo Grady and René Vidal. *P-Brush: Continuous Valued MRFs with Normed Pairwise Distributions for Image Segmentation.* In CVPR '09: Proceedings of the 2009 Conference on Computer Vision and Pattern Recognition, 2009. 76

[Sinop & Grady 2007] Ali Kemal Sinop and Leo Grady. *A Seeded Image Segmentation Framework Unifying Graph Cuts and Random Walker Which Yields a New Algorithm.* In ICCV '07: Proceedings of the 11th IEEE International Conference on Computer Vision. IEEE Computer Society, IEEE, Oct. 2007. 76

[Slabaugh & Unal 2005] Greg Slabaugh and Gozde Unal. *Graph Cuts Segmentation Using an Elliptical Shape Prior.* In ICIP '05: Proceedings

of the 12th IEEE International Conference on Image Processing, pp. 1222–1225, 2005. 78

[Smith & Jain 1982] Stephen P. Smith and Anil K. Jain. *Chord Distributions for Shape Matching.* Computer Graphics and Image Processing, vol. 20, no. 3, pp. 259 – 271, 1982. 20

[So 2007] Anthony Man-Cho So. *A Semidefinite Programming Approach To The Graph Realization Problem: Theory, Applications And Extensions.* PhD thesis, Stanford University, 2007. 21

[Staib & Duncan 1992] Lawrence H. Staib and James S. Duncan. *Boundary Finding with Parametrically Deformable Models.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, no. 11, pp. 1061–1075, 1992. 15

[Stegmann & Gomez 2002] Mikkel B. Stegmann and David Delgado Gomez. *A Brief Introduction to Statistical Shape Analysis*, 2002. 46, 96

[Stegmann *et al.* 2006] Mikkel B. Stegmann, Karl Sjöstrand and Rasmus Larsen. *Sparse Modeling of Landmark and Texture Variability using the Orthomax Criterion.* In SPIE International Symposium on Medical Imaging 2006, volume 6144, 2006. 9

[Sullivan 1990] John M. Sullivan. *A Crystalline Approximation Theorem for Hypersurfaces.* PhD thesis, Princeton University, 1990. 83

[Szeliski *et al.* 2008] Richard Szeliski, Ramin Zabih, Daniel Scharstein, Olga Veksler, Member Vladimir Kolmogorov, Aseem Agarwala, Marshall Tappen and Carsten Rother. *A Comparative Study of Energy Minimization Methods for Markov Random Fields with Smoothness-Based Priors.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 6, pp. 1068–1079, 2008. 65, 66

[Székely *et al.* 1996] Gábor Székely, András Kelemen, Christian Brechbühler and Guido Gerig. *Segmentation of 2-D and 3-D Objects from MRI Volume Data Using Constrained Elastic Deformations of Flexible Fourier Contour And Surface Models.* Medical Image Analysis, vol. 1, no. 1, pp. 19–34, 1996. 15

[Tagare 1997] Hemant D. Tagare. *Deformable 2-D Template Matching Using Orthogonal Curves.* IEEE Transactions on Medical Imaging, vol. 16, no. 1, pp. 108–117, 1997. 17

[Tang & Chung 2007] Tommy W. H. Tang and Albert C. S. Chung. *Non-rigid Image Registration Using Graph-cuts.* In MICCAI '07: Proceedings of the 10th International Conference on Medical Image Computing and Computer Assisted Intervention, 2007. 86, 91

[Taron *et al.* 2005] Maxime Taron, Nikos Paragios and Marie-Pierre Jolly. *Uncertainty-Driven Non-parametric Knowledge-Based Segmentation: The Corpus Callosum Case.* In Proceedings of the 3rd International Workshop on Variational, Geometric, and Level Set Methods in Computer Vision, pp. 198–207, 2005. 12

[Taron *et al.* 2009] Maxime Taron, Nikos Paragios and Marie-Pierre Jolly. *Registration with Uncertainties and Statistical Modeling of Shapes with Variable Metric Kernels.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 1, pp. 99–113, 2009. 49, 101

[Taylor & Cooper 1990] Christopher J. Taylor and David H. Cooper. *Shape Verification Using Belief Updating.* In BMVC '90: Proceedings of the British Machine Vision Conference, 1990. 20

[Terzopoulos *et al.* 1988] Demetri Terzopoulos, Andrew Witkin and Michael Kass. *Constraints on deformable models: recovering 3D shape and nongrid motion.* Artificial Intelligence, vol. 36, no. 1, pp. 91–123, 1988. 66

[Tomasi & Kanade 1991] Carlo Tomasi and Takeo Kanade. *Detection and Tracking of Point Features.* Technical Report CMU-CS-91-132, Carnegie Mellon University, 1991. 133

[Torresani *et al.* 2008] Lorenzo Torresani, Vladimir Kolmogorov and Carsten Rother. *Feature Correspondence Via Graph Matching: Models and Global Optimization.* In ECCV '08: Proceedings of the 10th European Conference on Computer Vision, pp. 596–609, 2008. 129

[Tresadern *et al.* 2009] Philip A. Tresadern, Harish Bhaskar, Steve A. Adeshina, Christopher J. Taylor and Tim F. Cootes. *Combining Local and Global Shape Models for Deformable Object Matching.* In BMVC '09: Proceedings of the British Machine Vision Conference, 2009. 71

[Tsai *et al.* 2001] Andy Tsai, Anthony Yezzi Jr., William Wells III, Clare Tempany, Dewey Tucker, Ayres Fan, W. Eric Grimson and Alan Willsky. *Model-Based Curve Evolution Technique for Image Segmentation.* In CVPR '01: Proceedings of the 2001 Conference on Computer Vision and Pattern Recognition, pp. 463–468, 2001. 12

[Tu *et al.* 2004] Jilin Tu, Zhenqiu Zhang, Zhihong Zeng and Thomas Huang. *Face Localization via Hierarchical CONDENSATION with Fisher Boosting Feature Selection.* In CVPR '04: Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition, volume 2, pp. 719–724, 2004. 70

[Twining & Taylor 2001] Carole Twining and Christopher Taylor. *Kernel Principal Component Analysis and the Construction of Non-Linear Active Shape Models.* In BMVC '01: Proceedings of the British Machine Vision Conference, 2001. 11

[Unal *et al.* 2004] Gozde B. Unal, Anthony J. Yezzi and Hamid Krim. *Information-Theoretic Active Polygons for Unsupervised Texture Segmentation.* International Journal of Computer Vision, vol. 62, no. 3, pp. 199–220, 2004. 68, 89

[Urtasun *et al.* 2006] Raquel Urtasun, David J. Fleet and Pascal Fua. *Temporal motion models for monocular and multiview 3D human body tracking.* Computer Vision and Image Understanding, vol. 2-3, no. 104, pp. 157–177, 2006. 133

[Üzümcü *et al.* 2003] Mehmet Üzümcü, Alejandro Frangi, Milan Sonka, Johan H.C. Reiber and Boudewijn P.F. Lelieveldt. *ICA vs. PCA Active Appearance Models: Application to Cardiac MR Segmentation.* In MICCAI '03: Proceedings of the 6th International Conference on

Medical Image Computing and Computer Assisted Intervention, pp. 451–458, 2003. 9

[van Ginneken *et al.* 2002] Bram van Ginneken, Alejandro F. Frangi, Joes J. Staal, Bart M. ter Haar Romeny and Max A. Viergever. *Active Shape Model Segmentation with Optimal Features.* IEEE Transactions on Medical Imaging, vol. 21, no. 8, pp. 924–933, 2002. 69, 70, 83, 106

[van Ginneken *et al.* 2006] Bram van Ginneken, Mikkel B. Stegmann and Marco Loog. *Segmentation of Anatomical Structures in Chest Radiographs using Supervised Methods: A Comparative Study on a Public Database.* Medical Image Analysis, vol. 10, no. 1, pp. 19–40, 2006. 48, 112, 113

[Černý 1985] Vlado Černý. *A Thermodynamical Approach to the Travelling Salesman Problem: an Efficient Simulation Algorithm.* Journal of Optimization Theory and Applications, vol. 45, pp. 41–51, 1985. 56

[Veksler 2008] Olga Veksler. *Star Shape Prior for Graph-Cut Image Segmentation.* In ECCV '08: Proceedings of the 10th European Conference on Computer Vision, 2008. 77

[Vemuri *et al.* 1993] Baba C. Vemuri, A. Radisavljevic and Christiana M. Leonard. *Multi-Resolution Stochastic 3D Shape Models for Image Segmentation.* In IPMI '93: Proceedings of the 13th International Conference on Information Processing in Medical Imaging, pp. 62–76, 1993. 14

[Vicente *et al.* 2008] Sara Vicente, Vladimir Kolmogorov and Carsten Rother. *Graph Cut Based Image Segmentation with Connectivity Priors.* In CVPR '08: Proceedings of the 2008 Conference on Computer Vision and Pattern Recognition, 2008. 76

[Vu & Manjunath 2008] Nhat Vu and B.S. Manjunath. *Shape Prior Segmentation of Multiple Objects with Graph Cuts.* In CVPR '08: Proceedings of the 2008 Conference on Computer Vision and Pattern Recognition, 2008. 79, 80, 82

[Wainwright *et al.* 2005] Martin J. Wainwright, Tommi S. Jaakkola and Alan S. Willsky. *MAP Estimation Via Agreement on Trees: Message-Passing and Linear Programming.* IEEE Transactions on Information Theory, vol. 51, no. 11, pp. 3697–3717, 2005. 63

[Wang & Staib 1998] Yongmei Wang and Lawrence H. Staib. *Boundary Finding with Correspondence Using Statistical Shape Models.* In CVPR '98: Proceedings of the 1998 Conference on Computer Vision and Pattern Recognition, pp. 338–345, 1998. 10

[Wang *et al.* 2008] Jack M. Wang, David J. Fleet and Aaron Hertzmann. *Gaussian process dynamical models for Human Motion.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 2, no. 30, pp. 283–298, 2008. 133

[Wang *et al.* 2009] Chaohui Wang, Martin de La Gorce and Nikos Paragios. *Segmentation, Ordering and Multi-object Tracking Using Graphical Models.* In ICCV '09: Proceedings of the 12th IEEE International Conference on Computer Vision, pp. 747–754, 2009. 134

[Werner 2007] Tomáš Werner. *A Linear Programming Approach to Max-Sum Problem: A Review.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 7, pp. 1165–1179, 2007. 61, 65, 84

[Xu & Prince 1998] Chenyang Xu and Jerry L. Prince. *Snakes, Shapes, and Gradient Vector Flow.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 7, no. 3, pp. 359–369, 1998. 84

[Yedidia *et al.* 2003] Jonathan S. Yedidia, William T. Freeman and Yair Weiss. Exploring Artificial Intelligence in The New Millennium, chapter: Understanding Belief Propagation And Its Generalizations, pp. 239–269. Morgan Kaufmann Publishers Inc., 2003. 60

[Yezzi Jr. *et al.* 2002] Anthony Yezzi Jr., Andy Tsai and Alan Willsky. *A Fully Global Approach to Image Segmentation via Coupled Curve Evolution Equations.* Journal of Visual Communication and Image Representation, vol. 13, no. 1/2, pp. 195–216, March 2002. 67

[You & Jain 1984] Zhisheng You and Anil K. Jain. *Performance Evaluation of Shape Matching via Chord Length Distribution.* Computer Vision, Graphics, and Image Processing, vol. 28, no. 3, pp. 185–198, 1984. 20

[Young & Householder 1938] Gale Young and A. S Householder. *Discussion of a Set of Points in terms of Their Mutual Distances.* Psychometrika, vol. 3, no. 1, pp. 19–22, 1938. 21

[Yu *et al.* 2007] Peng Yu, P. Ellen Grant, Yuan Qi, Xiao Han, Florent Ségonne, Rudolph Pienaar, Evelina Busa, Jenni Pacheco, Nikos Makris, Randy L. Buckner, Polina Golland and Bruce Fischl. *Cortical Surface Shape Analysis Based on Spherical Wavelets.* IEEE Transactions on Medical Imaging, vol. 26, no. 4, pp. 582–597, 2007. 15

[Yuille *et al.* 1989] Alan L. Yuille, David S. Cohen and Peter W. Hallinan. *Feature Extraction from Faces Using Deformable Templates.* In CVPR '89: Proceedings of the 1989 Conference on Computer Vision and Pattern Recognition, 1989. 17

[Yushkevich *et al.* 2006] Paul A. Yushkevich, Hui Zhang and James C. Gee. *Continuous Medial Representation for Anatomical Structures.* IEEE Transactions on Medical Imaging, vol. 25, no. 12, pp. 1547–1564, 2006. 14

[Zabih & Kolmogorov 2004] Ramin Zabih and Vladimir Kolmogorov. *Spatially Coherent Clustering Using Graph Cuts.* In CVPR '04: Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition, pp. II: 437–444, 2004. 74

[Zeng *et al.* 2008] Yun Zeng, Dimitris Samaras, Wei Chen and Qunsheng Peng. *Topology Cuts: a Novel Min-Cut/Max-Flow Algorithm for Topology Preserving Segmentation in N-D Images.* Computer Vision and Image Understanding, vol. 112, no. 1, pp. 81–90, 2008. 77

[Zhang *et al.* 2004] Jiayong Zhang, Robert Collins and Yanxi Liu. *Representation and Matching of Articulated Shapes.* In CVPR '04: Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition, pp. 342–349, 2004. 82

[Zhu & Yuille 1996] Song Chun Zhu and Alan L. Yuille. *Region Competition: Unifying Snakes, Region Growing, and Bayes/MDL for Multiband Image Segmentation.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 18, no. 9, pp. 884–900, 1996. 67, 88

[Zhu-Jacquot & Zabih 2007] Jie Zhu-Jacquot and Ramin Zabih. *Graph Cuts Segmentation with Statistical Shape Priors For Medical Images.* In 3rd International IEEE Conference on Signal-Image Technologies and Internet-Based System, 2007. 78, 79

[Zhu-Jacquot & Zabih 2008] Jie Zhu-Jacquot and Ramin Zabih. *Segmentation Of The Left Ventricle In Cardiac Mr Images Using Graph Cuts With Parametric Shape Priors.* In IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. 78, 79