DISSERTATION FOR THE DOCTORAL DEGREE OF THE
HUNGARIAN ACADEMY OF SCIENCES

From Regions to Shapes: The Extraction and Alignment of Visual Objects

Zoltan Kato
Department of Image Processing and Computer Graphics,
Institute of Informatics,
University of Szeged, Hungary

2013

I gratefully acknowledge the contributions of my MSc and PhD students from University of Szeged Hungary: Csaba Domokos, Péter Horváth, Zsolt Sánta, Csaba Molnár, József Németh; as well as Guo Qiang Song from National University of Singapore. My colleagues at various institutions also provided valuable contributions: Josiane Zerubia and Marc Berthod from INRIA Sophia Antipolis, France; Ian Jermyn from Department of Mathematical Sciences, Durham University, UK; Ting Chuen Pong and John Chung Mong Lee from Computer Science Department of the Hong Kong University of Science & Technology, Hong Kong; Attila Tanács from University of Szeged, Hungary; Joseph Francos from Ben Gurion University of the Negev, Israel; Jhimli Mitra, Soumya Ghose, and Fabrice Meriaudeau from Le2i-UMR CNRS 6306, Université de Bourgogne, France; Csaba Benedek and Tamás Szirányi from Distributed Events Analysis Research Laboratory of the Computer and Automation Institute of Hungarian Academy of Sciences; Natasa Sladoje from Faculty of Technical Sciences of the University of Novi Sad, Serbia; and Joakim Lindblad from Centre for Image Analysis, Uppsala University, Sweden.

The SPOT satellite images were provided by the French Space Agency (CNES). Lipid droplet microscopy images were obtained from László Vígh and Zsolt Török from Biological Research Centre, Szeged, Hungary; other microscopy images were provided by Péter Horváth from Light Microscopy Centre, ETH Zurich, Switzerland. The fractured bone CT images were obtained from the University of Szeged, Department of Trauma Surgery and were used with permission of Prof. Endre Varga, MD. Pelvic CT studies and hip prosthesis Xray images were provided by Endre Szabó, Ádám Perényi, Ágnes Séllei and András Palkó from the Radiology Department of the University of Szeged. Lung CT images and thoracic CT studies were provided by László Papp from Mediso Ltd., Hungary.

Contents

Figures

Tables

Introduction

The first step in almost every computer vision process, called *early vision*, involves a variety of digital image processing tasks dealing directly with massive amounts of pixel data. The goal is to transform the digitized image data into more meaningful tokens (edges, regions, objects, etc.) for higher level processing.

First, we deal with statistical approaches of image segmentation, where the final goal is to extract coherent regions corresponding to visual objects of a particular application (*e.g.* cells in a microscope image, or land coverage in satellite images). In real scenes, neighboring pixels usually have similar properties. In a probabilistic framework, such regularities are well expressed by Markov Random Fields. On the other hand, the local behavior of Markov Random Fields permits to develop highly efficient algorithms in the solution of the combinatorial optimization problem associated with such a model. We also discuss parameter estimation methods in order to develop completely data-driven algorithms.

Second, we will consider methods to recover the geometric relationship between a pair of visual objects extracted from images. This is a fundamental problem, also known as *registration* or *matching*, which occurs in many image analysis systems where views or different modalities of an object need to be compared or fused, *e.g.* multimodal medical imagery or the comparison of a template with the image of a manufactured part in an industrial inspection system.

An image processing system involves a sensing device (usually a camera) and computer algorithms to interpret the picture. The term *image* (more precisely, *monochrome image*) refers to a two dimensional light intensity function whose value at any point is proportional to the brightness (*grey-level*) of the image at that point [100]. A *digital image* is a discretized image both in spatial coordinates and in brightness. It is usually represented as a two dimensional matrix, the elements of such a digital array are called pixels. The digitized image is the starting point of any kind of computer analysis. In some applications, the sensing device may be more specific responding to other forms of light: infrared imaging, photon emission tomography, radar imaging [171], ultrasonic imaging, etc.

Extraction of coherent image regions

When human observers are interpreting images, they are not only taking into account direct observations like color or intensity, but also a priori knowledge about the world. However, such a complex, interacting method is rarely used in image processing systems. Most of the algorithms are bottom-up: they try to extract some useful information (basically a segmentation) solely from the observed image data and then the segmentation is interpreted. Obviously, image data alone cannot provide reliable information. Hence the use of higher level knowledge, in the form of shape priors, received more attention in the past few years. The dominating approach adopts a variational or level set framework where the segmentation criteria is summarized in an energy functional which takes its minimum at the desired segmentation of the input image. Previous work concentrated on foreground - background segmentation with a data model relying on image gradient and with template-like shape priors where the actual contour is matched to a reference shape and high deviations are penalized. However, handling of more than one, possibly different objects in a scene remains a challenge as well as the use of more elaborated data models. On the other hand, Markovian approaches are well suited to multi-object segmentation but little work has been done on embedding shape priors into such models.

The primary goal of any segmentation algorithm is to divide the domain R of the input image into the disjoint parts R_i such that they belong to distinct objects in the scene. The solution of this problem sometimes requires high level knowledge about the shape and appearance of the objects under investigation [80, 129, 172, 188]. In many applications, however, such information is not available or impractical to use. Hence low-level features of the surface patches are used for the segmentation process [57, 137, 206]. In either case, we have to summarize all relevant information in a model which is then adjusted to fit the image data.

One broadly used class of models is the so called *cartoon model*, which has been extensively studied from both probabilistic [32, 97] and variational [66, 158] viewpoints. The model assumes that the real world scene consists of a set of regions whose observed low-level features changes slowly, but across the boundary between them, these features change abruptly. What we want to infer is a *cartoon* ω (also called a *labeling*) consisting of a simplified, abstract version of the input image $\mathcal{I}$: regions R_i has a constant value (called a *label*

in our context) and the discontinuities between them form a curve Γ - the contour. The pair (ω, Γ) specifies a *segmentation*. Region based methods are mainly focusing on ω while edge based methods are trying to determine Γ directly. However, a good approach has to model both (either explicitly or implicitly).

Active Contours (snakes) are closed curves evolving toward the boundary of the object of interest. The curve evolution is governed by a boundary functional [127] which takes its minimum on the object contour. The main drawback of the parametric snake model is that it cannot handle topological changes easily. Nevertheless, they became quite popular because they make it relatively easy to enforce contour-smoothness; and starting from an appropriate initialization a local minimum of the associated energy function will give good results.

Taking the probabilistic approach, one usually wants to come up with a *probability measure* on the set Ω of all possible segmentations of $\mathcal{I}$ and then select the one with the highest probability. This probability measure is usually defined in a Bayesian framework [32, 77, 156, 199], in terms of a set of observed and hidden random variables. In our context, observations consists in low-level features used for partitioning the image, and the hidden entity represents the segmentation itself. The *data likelihood* (or *imaging model*) quantify how well any segmentation fits the observations.

In addition, a *prior* define a set of properties that any segmentation must possess regardless the image data. Purely data driven methods cannot deal very well with high noise, cluttered background or occlusions. Hence the idea of incorporating some prior knowledge about the shape of the objects has been considered by many researchers. Early approaches for shape prior were quite generic, enforcing some kind of homogeneity and contour smoothness [24, 66, 74, 79, 97, 127]. For example, [24, 97] uses a Markovian smoothness prior (basically a Potts model [58]) on ω; [66, 97] uses a line process to control the formation of region boundaries; and active contour models [127] have been using elasticity, rigidity, contour length, balloon or area minimizing forces [74, 79] in order to favor smooth closed curves. In spite of their simplicity, these methods proved to be very efficient in dealing with noisy images.

Herein, we will present our main contributions to construct efficient Markovian models to solve various image analysis problems related to remote sensing and biomedical applications. The ultimate goal of these methods is to extract coherent, meaningfull regions corresponding to visual objects of a particular application, *e.g.* tree crowns in aerial images, land coverage in satellite images, cells and lipid droplets in microscope images, moving regions in video frames, etc.

Alignment of visual objects

Registration is a fundamental problem in various fields of image processing where images taken from different views, at different times, or by different sensors need to be compared or combined. In a general setting, one is looking for a transformation which aligns two images

such that one image (called the *observation*) becomes similar to the second one (called the *template*).

When registering an image pair, first we have to characterize the possible deformations. From this point of view, registration techniques can be classified into two main categories: physical model-based and parametric or functional representation [118]. Herein, we deal with the latter representation, which typically originate from interpolation and approximation theory. Most of the existing approaches assume a linear transformation (rigid-body, similarity, affine) between the images, but in many applications nonlinear deformations [202] (*e.g.* projective, polynomial, elastic) need to be considered. Typical applications include visual inspection [192], object matching [60] and medical image analysis [114]. Good surveys can be found in [143, 207].

From a methodological point of view, we can differentiate *landmark-based* and *area-based* (or *featureless*) methods [71, 115, 144, 179, 207]. *Landmark-based methods* rely on extracted corresponding landmarks [105, 207], then the aligning transformation is recovered as a solution of a system of equations constructed from the established correspondences. Unfortunately, the correspondence problem itself is far from trivial, especially in the case of strong deformations. On the other hand, many *featureless approaches* estimate the transformation parameters directly from image intensity values over corresponding regions [146] or define a cost function based on a similarity metric and find the solution via a complex nonlinear optimization procedure [110].

A common assumption of both approaches is that the strength of the transformation is limited or close to identity: The neighborhood of a *landmark* is searched for correspondences, while *area-based* methods may get stuck in local minima for strong deformations. Furthermore, both approaches rely on the availability of rich radiometric information: *Landmark-based* methods usually match local brightness patterns around salient points [142] while *featureless methods* make use of intensity correlation between image patches. In many cases, however, such information may not be available (*e.g.* binary shapes) or it is very limited (*e.g.* prints, images of traffic signs). Another common problem is strong radiometric distortion (*e.g.* X-ray images, differently exposed images). Although there are some time consuming methods to cope with brightness change across image pairs [126], such image degradations are difficult to handle. While these issues make classical brightness-based features unreliable thus challenging current registration techniques, the segmentation of such images can be straightforward or readily available within a particular application. Therefore a valid alternative is to solve the registration problem using a binary representation (*i.e.* segmentation) of the images [181].

For example, spline-based deformations have been commonly used to register medical images or volumes. The interpolating Thin-plate Splines (TPS) was originally proposed by [67], which relies on a set of point correspondences between the image pairs. However, these correspondences are prone to error in real applications and therefore [175] extended the bending energy of TPS to approximation and regularization by introducing the correspondence localization error. On the other hand, we [12] proposed a generic framework for non-rigid registration which does not require explicit point correspondences. In our subse-

quent work [41], this framework has been adopted to solve multimodal registration of MRI and TRUS prostate images for reliable cancer diagnosis.

Another prominent medical application is complex bone fracture reduction which frequently requires surgical care, especially when angulation or displacement of bone fragments are large. In such situations, computer aided surgical planning is done before the actual surgery takes place, which allows to gather more information about the dislocation of the fragments and to arrange and analyze the surgical implants to be inserted. A crucial part of such a system is the relocation of bone fragments to their original anatomic position. In [8], we applied our framework to reduce pelvic fractures using 3D rigid-body transformations. In cases of single side fractures, the *template* is simply obtained by mirroring intact bones of the patient.

Herein, we will present our general registration framework for linear and non-linear alignment of extracted visual objects. A unique feature of our approach is that a wide range of deformations are handled in a unified, correspondence-less framework. It provides an efficient solution for various applications ranging from medical imaging to industrial inspections, where classical methods perform poorly.

IN THIS CHAPTER:

1.

Markovian segmentation models

In this chapter, we summarize the main results of our early work related to Markovian image modeling:

A novel hierarchical MRF model and its application to satellite image segmentation.

A new annealing schedule for Simulated Annealing: Multi-temperature annealing allows to assign different temperatures to different cliques during the minimization of the energy of a MRF model. The convergence of the new algorithm has also been proved toward a global optimum.

Etimation of the hierarchical model parameters and application to remote sensing image segmentation.

1.1 Introduction

The primary goal of any segmentation algorithm is to divide the domain R of the input image into the disjoint parts R_i such that they belong to distinct objects in the scene. The solution of this problem sometimes requires high level knowledge about the shape and appearance of the objects under investigation [80, 129, 172, 188]. In many applications, however, such information is not available or impractical to use. Hence low-level features of the surface patches are used for the segmentation process [57, 137, 206]. Herein, we are interested in the latter approach. In either case, we have to summarize all relevant information in a model which is then adjusted to fit the image data.

One broadly used class of models is the so called *cartoon model*, which has been extensively studied from both probabilistic [97] and variational [66, 158] viewpoints. The model assumes that the real world scene consists of a set of regions whose observed low-level features changes slowly, but across the boundary between them, these features change abruptly. What we want to infer is a *cartoon* ω consisting of a simplified, abstract version of the input image $\mathcal{I}$: regions R_i has a constant value (called a *label* in our context) and the discontinuities between them form a curve Γ - the contour. The pair (ω, Γ) specifies a *segmentation*. Region based methods are mainly focusing on ω while edge based methods are trying to determine Γ directly.

Taking the probabilistic approach, one usually wants to come up with a *probability measure* on the set Ω of all possible segmentations of $\mathcal{I}$ and then select the one with the highest probability. Note that Ω is finite, although huge. A widely accepted standard, also motivated by the human visual system [128, 157], is to construct this probability measure in a Bayesian framework [77, 156, 199]: We shall assume that we have a set of observed (Y) and hidden (X) random variables. In our context, any observed value $y \in Y$ represents the low-level features used for partitioning the image, and the hidden entity $x \in X$ represents the segmentation itself. First, we have to quantify how well any occurrence of x fits y. This is expressed by the probability distribution $P(y|x)$ - the *imaging model*. Second, we define a set of properties that any segmentation x must posses regardless the image data. These are described by $P(x)$, the *prior*, which tells us how well any occurrence x satisfies these properties. Factoring these distributions and applying the Bayes theorem gives us the *posterior* distribution $P(x|y) \propto P(y|x)P(x)$. Note that the constant factor $1/P(y)$ has been dropped as we are only interested in $\widehat{x}$ which *maximizes* the posterior, *i.e.* the Maximum A Posteriori (MAP) estimate of the hidden field X.

The models of the above distributions depend also on certain parameters that we denote by Θ. Supervised segmentation assumes that these parameters are either known or a set of joint realizations of the hidden field X and observations Y (called a *training set*) is available [97, 191]. This is known in statistics as the *complete data* problem which is relatively easy to solve using Maximum Likelihood (ML) [77]. Although the prior knowledge of the parameters is a strong assumption, supervised methods are still useful alternatives when working in a controlled environment. Many industrial applications, like quality inspection of agricultural products [161], fall into this category. In the unsupervised case, however, we

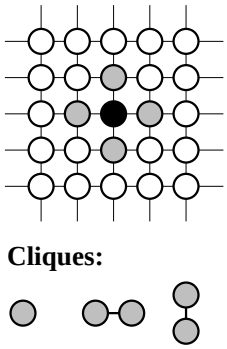


Figure 1.1: *First order neighborhood system with corresponding cliques.*

know neither Θ nor X. This is called the *incomplete data* problem where both Θ and X has to be inferred from the only observable entity Y. Hence our MAP estimation problem becomes $(\widehat{x}, \widehat{\Theta}) = \arg\max_{x,\Theta} P(x, \Theta|y)$. *Expectation Maximization* (EM) [81] and its variants (Stochastic EM [75, 149], Gibbsian EM [76]), as well as *Iterated Conditional Expectation* (ICE) [25, 68] are widely used to solve such problems. It is important to note, however, that these methods calculate a local maximum [77].

Due to the difficulty of estimating the number of pixel classes (or clusters), unsupervised algorithms often suppose that this parameter is *known a priori* [99,106,137,140,149]. When the number of pixel classes is also being estimated, the unsupervised segmentation problem may be treated as a *model selection* problem over a combined model space.

1.1.1 Markovian approach

In real images regions are usually homogeneous, neighboring pixels have similar properties. Markov Random Fields (MRF) are often used to capture such contextual constraints in a probabilistic framework. MRFs are well studied with a strong theoretical background hence providing a tool for rigorous and concise image modeling. Furthermore, they allow Markov Chain Monte Carlo (MCMC) sampling of the (hidden) underlying structure which greatly simplifies inference and parameter estimation.

Formally, a simple MRF image model is constructed as follows: we are given a set of sites (usually corresponding to pixels) $\mathcal{S} = \{s_1, s_2, \ldots, s_N\}$. For each site s, the region-type (or class) that the site belongs to is specified by a class label, ω_s, which is modeled as a discrete random variable taking values in $\Lambda = \{1, 2, \ldots, L\}$. The set of these labels $\omega = \{\omega_s, s \in \mathcal{S}\}$ is a random field, called the *label process*. Furthermore, the observed image features (*e.g.* graylevel, color, texture,...) are supposed to be a realization $\mathcal{F} = \{f_s|s \in \mathcal{S}\}$ from another random field, which is a function of the label process ω. Basically, the *image process* $\mathcal{F}$ represents the manifestation of the underlying label process. Thus, the overall segmentation model is composed of the hidden label process ω and the observable noisy image process

$\mathcal{F}$. If each pixel class is represented by a different model then the observed image may be viewed as a sample from a realization of the underlying label field.

$(\omega, \mathcal{F})$ is then regarded as a MRF with respect to an appropriate neighborhood-system $\mathcal{G} = \{\mathcal{G}_s\}_{s \in \mathcal{S}}$. The simplest example of such a neighborhood can be seen in Fig. 1.1. According to the Hammersley-Clifford theorem [54], $(\omega, \mathcal{F})$ must then follow a Gibbs distribution with an energy function $U(\omega, \mathcal{F}) = \sum_{C \in \mathcal{C}} V_C(\omega, \mathcal{F})$, where C denotes a clique of $\mathcal{G}$, and $\mathcal{C}$ is the set of all cliques. The restriction of ω to the sites of a given clique C is denoted by ω_C. The potential function $V_C(\omega_C)$ is defined for every $C \in \mathcal{C}$ and every $\omega \in \Omega$, where $\Omega = \Lambda^N$ is the set of all possible L^N discrete labelings. The advantage of such a decomposition is that these potentials are a function of the local configuration of the field making it possible to define the Gibbs distribution directly in terms of local interactions.

The MAP estimate $\hat{\omega}$ of the label field is then obtained by minimizing the non-convex energy function, which can be solved by stochastic or deterministic relaxation [3, 4, 33, 34].

1.2 Hierarchical MRF models and multi-temperature annealing

It is well known that multigrid methods can improve significantly the convergence rate and the quality of the final results of iterative relaxation techniques. Herein, we propose a new hierarchical model [14, 20–24], which consists of a label pyramid and a single observation field. The parameters of the coarse grid can be derived by simple computation from the finest grid. In addition, we have introduced a new local interaction between two neighboring grids which allows to propagate information more efficiently giving estimates closer to the global optimum for deterministic as well as for stochastic relaxation schemes. For the hierarchical model, we also propose a novel Multi-Temperature Annealing (MTA) algorithm [24,36]. The convergence towards the global optimum is proven by the generalization of the annealing theorem of Geman and Geman [97].

1.2.1 Multiscale and hierarchical model

In the following, we will focus on a MRF with a first order neighborhood (see Fig. 1.1) whose energy function is given by:

$$U(\omega, \mathcal{F}) = U_1(\omega, \mathcal{F}) + U_2(\omega) \tag{1.1}$$

where U_1 (resp. U_2) denotes the energy of the first order (resp. second order) cliques. To generate a multigrid MRF model, let us divide the initial grid into blocks of $n \times n$, typically 16 (4×4) neighboring pixels. We consider that the same label is assigned to each pixels of a given block. These configurations will describe the MRF at scale 1. Scale i is defined similarly by considering labels which are constant over blocks of size $n^i \times n^i$.

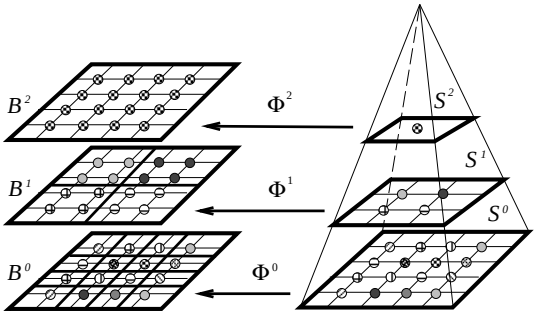


Figure 1.2: *The isomorphism Φ^i between $\mathcal{B}^i$ and $\mathcal{S}^i$.*

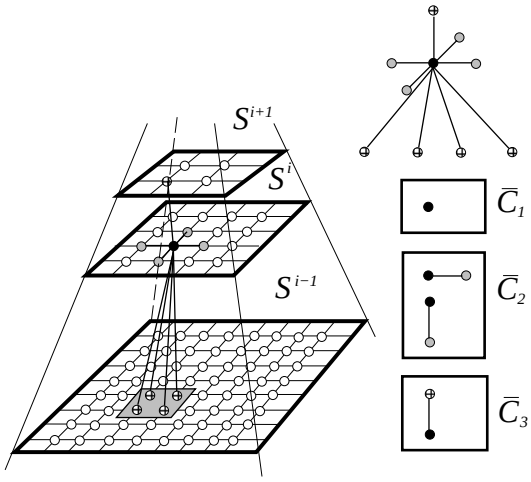


Figure 1.3: *The neighborhood system $\bar{\mathcal{G}}$ and the cliques $\bar{\mathcal{C}}_1$, $\bar{\mathcal{C}}_2$ and $\bar{\mathcal{C}}_3$.*

Let $\mathcal{B}^i = \{b_1^i, \ldots, b_{N_i}^i\}$ $(N_i = N/n^{2i})$ denote the set of blocks and Ω_i the configuration-space at scale i ($\Omega_i \subset \Omega_{i-1} \subset \cdots \subset \Omega_0 = \Omega$). The label associated with block b_k^i is denoted by ω_k^i. We can define the same neighborhood structure on $\mathcal{B}^i$ as on $\mathcal{S}$:

$$b_k^i \text{ and } b_l^i \text{ are neighbors} \Longleftrightarrow \begin{cases} b_k^i \equiv b_l^i \text{ or} \\ \exists C \in \mathcal{C}\text{: } C \cap b_k^i \neq \emptyset \text{ and } C \cap b_l^i \neq \emptyset \end{cases} \tag{1.2}$$

Let us partition the original set $\mathcal{C}$ into two disjoint subsets $\mathcal{C}_k^i$ (cliques which are included in b_k^i) and $\mathcal{C}_{k,l}^i$(cliques which sit astride two neighboring blocks $\{b_k^i, b_l^i\}$). It is obvious from this partition that our energy function can be decomposed in the following way:

$$U_1(\omega, \mathcal{F}) = \sum_{s \in \mathcal{S}} V_1(\omega_s, f_s) = \sum_{b_k^i \in \mathcal{B}^i} \underbrace{\sum_{s \in b_k^i} V_1(\omega_s, f_s)}_{V_1^{\mathcal{B}^i}(\omega_k^i, \mathcal{F})} = \sum_{b_k^i \in \mathcal{B}^i} V_1^{\mathcal{B}^i}(\omega_k^i, \mathcal{F}) \tag{1.3}$$

$$\begin{aligned} U_2(\omega) = \sum_{C \in \mathcal{C}} V_2(\omega_c) &= \sum_{b_k^i \in \mathcal{B}^i} \underbrace{\sum_{C \in \mathcal{C}_k^i} V_2(\omega_c)}_{V_k^{\mathcal{B}^i}(\omega_k^i)} + \sum_{\{b_k, b_l\} neighbors} \underbrace{\sum_{C \in \mathcal{C}_{k,l}^i} V_2(\omega_c)}_{V_{k,l}^{\mathcal{B}^i}(\omega_k^i, \omega_l^i)} \\ &= \sum_{b_k^i \in \mathcal{B}^i} V_k^{\mathcal{B}^i}(\omega_k^i) + \sum_{\{b_k, b_l\} neighbors} V_{k,l}^{\mathcal{B}^i}(\omega_k^i, \omega_l^i) \end{aligned} \tag{1.4}$$

Now, we define a pyramid (see Figure 1.2) where level i contains the coarse grid $\mathcal{S}^i$ which is isomorphic to the scale $\mathcal{B}^i$. The coarse grid has a reduced configuration space $\Xi^i = \Lambda^{N_i}$.

The isomorphism $\Phi^i : \mathcal{S}^i \to \mathcal{B}^i$ is just a projection of the coarse label field to the finest grid $\mathcal{S}^0 \equiv \mathcal{S}$. The energy function on the grid $\mathcal{S}^i$ $(i = 0, \ldots, M)$ is derived from Eq. (1.3)–(1.4):

$$\begin{aligned} U^i(\omega^i, \mathcal{F}) &= U_1^i(\omega^i, \mathcal{F}) + U_2^i(\omega^i) = U_1(\Phi^i(\omega_i), \mathcal{F}) + U_2(\Phi^i(\omega_i)) \\ \text{where } U_1^i(\omega^i, \mathcal{F}) &= \sum_{k \in \mathcal{S}^i} (V_1^{\mathcal{B}^i}(\omega_k^i, \mathcal{F}) + V_k^{\mathcal{B}^i}(\omega_k^i)) = \sum_{k \in \mathcal{S}^i} V_1^i(\omega_k^i, \mathcal{F}) \qquad (1.5) \\ \text{and } U_2^i(\omega^i) &= \sum_{\{k,l\} neighbors} V_{k,l}^{\mathcal{B}^i}(\omega_k^i, \omega_l^i) = \sum_{C^i \in \mathcal{C}^i} V_2^i(\omega_C^i) \qquad (1.6) \end{aligned}$$

where C^i is a second order clique corresponding to the definition in Eq. (1.2) and $\mathcal{C}^i$ is the set of cliques on grid i.

Let $\bar{\mathcal{S}} = \{\bar{s}_1, \ldots, \bar{s}_{\bar{N}}\} = \bigcup_{i=0}^{M} \mathcal{S}^i$ ($\bar{N} = \sum_{i=0}^{M} N^i$) denote the sites of the pyramid. We define the following function Ψ between two neighboring levels, which assigns to a site its descendants (that is the sites of the corresponding block):

$$\Psi : \mathcal{S}^i \longrightarrow \mathcal{S}^{i-1}, \quad \Psi(\bar{s}) = \{\bar{r} \mid \bar{s} \in \mathcal{S}^i \Rightarrow \bar{r} \in \mathcal{S}^{i-1} \text{ and } b_{\bar{r}}^{i-1} \subset b_{\bar{s}}^i\} \qquad (1.7)$$

It is clear that Ψ^{-1} will assign to a site its ancestor (that is the site at the upper level corresponding to the block of this site). Now we can define on these sites the following neighborhood-system (see Fig. 1.3):

$$\bar{\mathcal{G}} = (\bigcup_{i=0}^{M} \mathcal{G}^i) \cup \{\Psi^{-1}(\bar{s}) \cup \Psi(\bar{s}) \mid \bar{s} \in \bar{\mathcal{S}}\} \qquad (1.8)$$

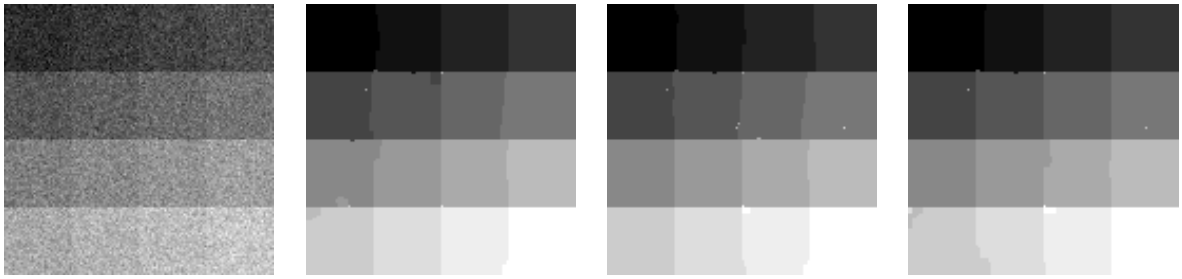
where $\mathcal{G}^i$ is the neighborhood structure of the i^{th} level. We will consider only the first and second order cliques, potentials for other cliques are supposed to be 0. Let $\bar{\mathcal{C}}$ denote the set of these cliques which can be partitioned into three disjoint subsets $\bar{\mathcal{C}}_1, \bar{\mathcal{C}}_2, \bar{\mathcal{C}}_3$ corresponding to first order cliques, second order cliques which are on the same level and second order cliques which sit astride two neighboring levels (see Figure 1.3). Let $\bar{\Omega}$ denote the configuration-space of the pyramid:

$$\bar{\Omega} = \Xi^0 \times \Xi^1 \times \cdots \times \Xi^M = \{\bar{\omega} \mid \bar{\omega} = (\omega^0, \omega^1, \ldots, \omega^M)\} \qquad (1.9)$$

The model on the pyramid defines a MRF, whose energy function is given by:

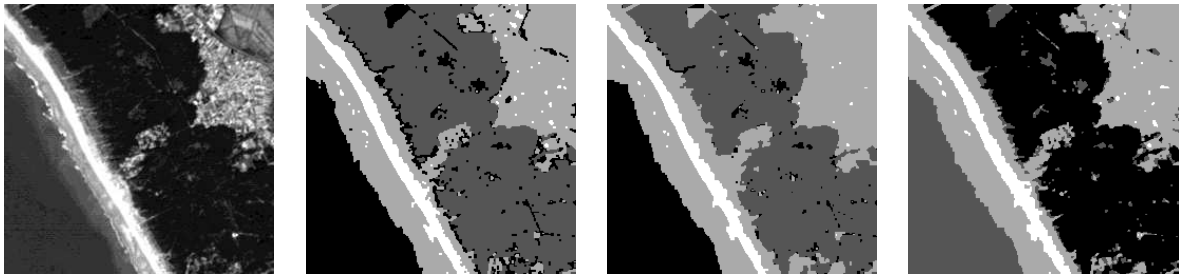
$$\begin{aligned} \bar{U}(\bar{\omega}, \mathcal{F}) &= \bar{U}_1(\bar{\omega}, \mathcal{F}) + \bar{U}_2(\bar{\omega}) \qquad (1.10) \\ \bar{U}_1(\bar{\omega}, \mathcal{F}) &= \sum_{\bar{s} \in \bar{\mathcal{S}}} \bar{V}_1(\bar{\omega}_{\bar{s}}, \mathcal{F}) = \sum_{i=0}^{M} \sum_{s^i \in \mathcal{S}^i} V_1^i(\omega_{s^i}^i, \mathcal{F}) = \sum_{i=0}^{M} U_1^i(\omega^i, \mathcal{F}) \\ \bar{U}_2(\bar{\omega}) &= \sum_{C \in \bar{\mathcal{C}}_2} \bar{V}_2(\bar{\omega}_C) + \sum_{C \in \bar{\mathcal{C}}_3} \bar{V}_2(\bar{\omega}_C) = \sum_{i=0}^{M} U_2^i(\omega^i) + \sum_{C \in \bar{\mathcal{C}}_3} \bar{V}_2(\bar{\omega}_C) \\ &= \sum_{i=0}^{M} \sum_{C \in \mathcal{C}^i} V_2^i(\omega_c^i) + \sum_{C \in \bar{\mathcal{C}}_3} \bar{V}_2(\bar{\omega}_C) \end{aligned}$$

model	num. of iter.	CPU time	time/iter.	error rate	β	γ
monogrid	89	10.39 sec.	0.117 sec.	2576	1.0	—
multiscale	146	14.7 sec.	0.1 sec.	2118	1.0	—
hierarchical	42	460.9 sec.	10.97 sec.	1231	1.0	0.2



Noisy image ($SNR = 10dB$) Monogrid Multiscale Hierarchical

Figure 1.4: *Results obtained by the Gibbs Sampler [97] on a noisy synthetic image (*128×128*,* $SNR = 10dB$*) with* 16 *classes [14, 20–24]. In the table, we show for each model the number of iterations, the CPU time, the error rate of the segmentation (= the number of misclassified pixels) and the inter- and intra-clique potentials* β *and* γ*.*



Original image Monogrid Multiscale Hierarchical

Figure 1.5: *Results obtained by ICM [64] on a (*256×256*) SPOT image with* 4 *classes [14, 20–24].*

The above energy of the hierarchical model can be minimized using classical combinatorial optimization algorithms [3, 4, 33, 34, 136]. The only difference is that we work on a pyramid here and not on a rectangular lattice as in the case of classical monogrid models. We have applied the model for supervised image segmentation and compared the segmentation results of the classical monogrid [3, 33–35], multiscale and hierarchical models on synthetic (Fig. 1.4) and real (Fig. 1.5) images. For both images, the label pyramid has been generated with 4 levels. The detailed equations can be found in [14, 24]. All tests have been conducted on a Connection Machine CM200 with $8K$ processors. In terms of segmentation quality, the hierarchical model clearly outperforms the other methods. Further results can be found in [14, 24].

1.2.2 Multi-temperature annealing

In the following we will focus on Simulated Annealing (SA) [97], where the temperature-change is controlled by the so-called *annealing schedule*. There are two well known schemes, *homogeneous* and *inhomogeneous* annealing [136], which works also on the hierarchical model. Herein, we propose a new annealing schedule, called *Multi-Temperature Annealing* (MTA), which is the most efficient with the new model. The basic idea is to associate higher temperatures to coarser levels in the pyramid which makes the algorithm less sensitive to local minima. However at a finer resolution, the relaxation is performed at a lower temperature (at the bottom level, it is close to 0). For the cliques siting between two levels, we use either the temperature of the finer level or the one of the coarser level (but once chosen, we always keep the same choice throughout the algorithm). More generally, we have the following problem:

Let $\mathcal{S} = \{s_1, \ldots, s_N\}$ be a set of sites, $\mathcal{G}$ some neighborhood system with cliques $\mathcal{C}$ and ω a MRF over these sites with energy function U. π_0 denotes the uniform distribution on the set of globally optimal configurations, and define $U^{sup} = \max_\omega U(\omega)$, $U^{inf} = \min_\omega U(\omega)$ and $\Delta = U^{sup} - U^{inf}$. Furthermore, let us suppose that the sites are visited for updating in the order $\{n_1, n_2, \ldots\} \subset \mathcal{S}$. We now define an annealing scheme where the temperature T depends on the iteration k and on the cliques C. For that purpose, let $\oslash$ denotes the following operation:

$$P(X = \omega) = \pi_{T(k,C)}(\omega) = \frac{\exp(-U(\omega) \oslash T(k, C))}{Z} \tag{1.11}$$

$$\text{where } U(\omega) \oslash T(k, C) = \sum_{C \in \mathcal{C}} \frac{V_C(\omega)}{T(k, C)} \,. \tag{1.12}$$

As usual with SA [97, 136], the transition from one configuration to another is governed by the energy change between the two states. Assuming $\omega' \in \Omega_{opt}$ is a *globally* optimal configuration, $U(\omega') - U^{inf}$ equals to 0 (*i.e.* there is no more energy change, the system is frozen). In the case of a classical annealing, dividing by a constant temperature does not change this relation (obviously, $\forall k\colon (U(\omega') - U^{inf})/T_k$ is still 0). But it is not necessarily true

that $(U(\omega') - U^{inf}) \oslash T(k, C)$ is also 0! Because choosing sufficiently small temperatures for the cliques where ω'_C is *locally* not optimal (*i.e.* strengthening the non-optimal cliques) and choosing sufficiently high temperatures for the cliques where ω'_C is *locally* optimal (*i.e.* weakening the optimal cliques), we obtain $(U(\omega') - U^{inf}) \oslash T(k, C) > 0$, meaning that ω' is no longer *globally* optimal (*i.e.* in such cases, SA may not be able to reach a global optimum).

Thus, we have to impose further conditions on the temperature to guarantee the convergence toward global optimum. First, let us examine the decomposition over the cliques of $U(\omega) - U(\eta)$ for arbitrary ω and η, $\omega \neq \eta$:

$$U(\omega) - U(\eta) = \sum_{C \in \mathcal{C}} (V_C(\omega) - V_C(\eta)). \tag{1.13}$$

Indeed, there may be negative and positive members in the decomposition. According to this fact, we have the following subsums:

$$\begin{aligned} \sum_{C \in \mathcal{C}} (V_C(\omega) - V_C(\eta)) \quad = \quad & \underbrace{\sum_{C \in \mathcal{C}: (V_C(\omega) - V_C(\eta)) < 0} (V_C(\omega) - V_C(\eta))}_{\Sigma^-(\omega, \eta)} \\ & + \underbrace{\sum_{C \in \mathcal{C}: (V_C(\omega) - V_C(\eta)) \geq 0} (V_C(\omega) - V_C(\eta))}_{\Sigma^+(\omega, \eta)}. \end{aligned} \tag{1.14}$$

Furthermore, let us define Σ^+_Δ as:

$$\Sigma^+_\Delta = \min_{\substack{\omega' \in \Omega_{sup} \\ \omega'' \in \Omega_{opt}}} \Sigma^+(\omega', \omega''). \tag{1.15}$$

Then the following theorem gives an annealing schedule, where *the temperature is a function of k and $C \in \mathcal{C}$* [24]:

Theorem 1.2.1 (Multi-Temperature Annealing) *Assume that there exists an integer $\kappa \geq N$ such that for every $k = 0, 1, 2, \ldots$, $\mathcal{S} \subseteq \{n_{k+1}, n_{k+2}, \ldots, n_{k+\kappa}\}$. For all $C \in \mathcal{C}$, let $T(k, C)$ be any decreasing sequence of temperatures in k for which*

(i) $\lim_{k\to\infty} T(k, C) = 0$.
Let us denote respectively by T_k^{inf} and T_k^{sup} the maximum and minimum of the temperature function at k ($\forall C \in \mathcal{C}$: $T_k^{inf} \leq T(k, C) \leq T_k^{sup}$).

(ii) For all $k \geq k_0$, for some integer $k_0 \geq 2$: $T_k^{inf} \geq N\Sigma_\Delta^+ / \ln(k)$.

(iii) If $\Sigma^-(\omega, \omega') \neq 0$ for some $\omega \in \Omega \setminus \Omega_{opt}$, $\omega' \in \Omega_{opt}$ then a further condition must be imposed:
For all k: $\frac{T_k^{sup} - T_k^{inf}}{T_k^{inf}} \leq R$ with

$$R = \min_{\substack{\omega \in \Omega \setminus \Omega_{opt} \\ \omega' \in \Omega_{opt} \\ \Sigma^-(\omega,\omega') \neq 0}} \frac{U(\omega) - U^{inf}}{\mid \Sigma^-(\omega, \omega') \mid}. \tag{1.16}$$

Then for any starting configuration $\eta \in \Omega$ and for every $\omega \in \Omega$:

$$\lim_{k\to\infty} P(X(k) = \omega \mid X(0) = \eta) = \pi_0(\omega). \tag{1.17}$$

The complete proof of this theorem can be found in Appendix A.1 and in [20, 24].

Remarks:

1. In practice, we cannot determine R and Σ_Δ^+, as we cannot compute Δ neither.
2. Considering Σ_Δ^+ in condition 1.2.1/ii, we have the same problem as in the case of a classical annealing. The only difference is that in a classical annealing, we have Δ instead of Σ_Δ^+. Consequently, the same solutions may be used: an exponential schedule with a sufficiently high initial temperature.
3. The factor R is more interesting. We propose herein two possibilities which can be used for practical implementations of the method: Either we choose a sufficiently small interval $[T_0^{inf}, T_0^{sup}]$ and suppose that it satisfies the condition 1.2.1/iii (we have used this technique in the simulations), or we use a more strict but easily verifiable condition instead of condition 1.2.1/iii, namely:
$$\lim_{k\to\infty} \frac{T_k^{sup} - T_k^{inf}}{T_k^{inf}} = 0. \tag{1.18}$$
4. What happens if $\Sigma^-(\omega, \omega')$ is zero for all ω and ω' in condition 1.2.1/iii and thus R is not defined? This is the best case because it means that all *globally* optimal

SNR$= 3dB$ Inhomogeneous MTA Energy plots

Figure 1.6: *Energy decrease and segmentation results of the Gibbs sampler on a synthetic image with the inhomogeneous and MTA schedules. In both cases, the parameters were strictly the same, the only difference is the applied schedule. We also show the global energy plot (computed at a fixed temperature on the finest level) versus the number of iterations. Note that both schedules reach practically the same minimum (53415.4 for the inhomogeneous and 53421.4 for the MTA), however the inhomogeneous schedule requires* 238 *iterations (*796.8 *sec. CPU time) while the MTA schedule requires only* 100 *iterations (*340.6 *sec. CPU time) for the convergence [20, 24].*

configurations are also *locally* optimal. That is we have no restriction on the interval $[T_k^{inf}, T_k^{sup}]$, thus any *local* temperature schedule satisfying conditions 1.2.1/i–1.2.1/ii is good.

In Fig. 1.6, we compare the inhomogeneous and MTA schedules on a noisy synthetic image using the Gibbs sampler. Since each site interacts with its ancestor and its descendants, the hierarchical model usually requires more computing time than a monogrid model. However, as we have shown in [20, 24], experiments prove that this model with the MTA schedule yields faster convergence (with respect to the number of iterations) for the stochastic relaxation algorithms and gives estimates which are closer to the global optimum. Other tests can be found in [20, 24].

1.3 Parameter estimation

In real life applications, the model parameters are usually unknown, one has to estimate [56] them from the observable image. Here we develop an algorithm for hierarchical Markovian models [25, 36–38]. Our approach is similar in spirit to Iterative Conditional Estimation [149, 167] as well as to the Estimation-Maximization algorithm: we recursively look at the Maximum a Posteriori (MAP) estimate of the label field given the estimated parameters then we look at the Maximum Likelihood (ML) estimate of the parameters given a tentative labeling obtained in the previous step. The only parameter supposed to be known is the number of labels, all the other parameters are estimated.

When both the model parameters Θ and ω are unknown, the estimation problem be-

comes [38, 96, 137]

$$(\widehat{\omega}, \widehat{\Theta}) = \arg\max_{\omega,\Theta} P(\omega, \mathcal{F} \mid \Theta). \tag{1.19}$$

The pair $(\widehat{\omega}, \widehat{\Theta})$ is the global maximum of the joint probability $P(\omega, \mathcal{F} \mid \Theta)$. If we regard Θ as a random variable, the above maximization is an ordinary MAP estimation in the following way [96]: Let us suppose, that Θ is restricted to a finite volume domain $\mathcal{D}_\Theta$ and suppose that Θ is uniform on $\mathcal{D}_\Theta$ (that is $P(\Theta)$ is constant). Then, we get [38, 96]:

$$\begin{aligned}
\arg\max_{\omega,\Theta} P(\omega, \Theta \mid \mathcal{F}) &= \arg\max_{\omega,\Theta} \frac{P(\omega, \mathcal{F} \mid \Theta)P(\Theta)}{P(\mathcal{F})} && (1.20)\\
&= \arg\max_{\omega,\Theta} \frac{P(\omega, \mathcal{F} \mid \Theta)}{\int_{\mathcal{D}_\Theta} \sum_{\omega\in\Omega} P(\omega, \mathcal{F} \mid \Theta) d\Theta} && (1.21)\\
&= \arg\max_{\omega,\Theta} P(\omega, \mathcal{F} \mid \Theta). && (1.22)
\end{aligned}$$

However, this maximization is very difficult, having no direct solution. Even Simulated Annealing (SA) is not implementable because the local characteristics with respect to the parameters Θ cannot be computed from $P(\omega, \mathcal{F} \mid \Theta)$. One possible solution is to adopt the following criterion instead [38, 96, 137]:

$$\begin{aligned}
\widehat{\omega} &= \arg\max_{\omega} P(\omega, \mathcal{F} \mid \widehat{\Theta}) && (1.23)\\
\widehat{\Theta} &= \arg\max_{\Theta} P(\widehat{\omega}, \mathcal{F} \mid \Theta) && (1.24)
\end{aligned}$$

Clearly, Eq. (1.23) is equivalent to Eq. (1.19) for $\Theta = \widehat{\Theta}$ and Eq. (1.24) is equivalent to Eq. (1.19) with $\omega = \widehat{\omega}$. Furthermore, Eq. (1.23) is equivalent to the MAP estimate of ω in the case of known parameters:

$$\arg\max_{\omega} P(\omega, \mathcal{F} \mid \widehat{\Theta}) = \arg\max_{\omega} P(\omega \mid \mathcal{F}, \widehat{\Theta}) P(\mathcal{F} \mid \widehat{\Theta}) = \arg\max_{\omega} P(\omega \mid \mathcal{F}, \widehat{\Theta}).$$

Hence in the following we will concentrate on Eq. (1.24) which gives the ML estimate of the parameters. Considering the hierarchical MRF segmentation model (see Fig. 1.3), we have the following logarithmic likelihood function [25, 36–38]:

$$\begin{aligned}
&\sum_{i=0}^{M} \sum_{s^i\in\mathcal{S}^i} \sum_{s\in b^i_{s^i}} \left(-\ln(\sqrt{2\pi}\sigma_{\widehat{\omega}_s}) - \frac{(f_s - \mu_{\widehat{\omega}_s})^2}{2\sigma^2_{\widehat{\omega}_s}} \right)\\
&-\beta \underbrace{\sum_{i=0}^{M} q^i \sum_{C^i\in\mathcal{C}^i} \delta(\widehat{\omega}_{C^i})}_{N^{ih}(\widehat{\omega})} - \gamma \underbrace{\sum_{C\in\bar{\mathcal{C}}_3} \delta(\widehat{\omega}_C)}_{\bar{N}^{ih}(\widehat{\omega})} - \ln(Z(\beta,\gamma))
\end{aligned} \tag{1.25}$$

where q^i is the number of cliques between two neighboring blocks at scale $\mathcal{B}^i$, $N^{ih}(\widehat{\omega})$ denotes the number of inhomogeneous cliques siting at the same scale and $\bar{N}^{ih}(\widehat{\omega})$ denotes

the number of inhomogeneous cliques siting astride two neighboring levels in the pyramid. Considering the first term, we get

$$\begin{aligned}&\sum_{i=0}^{M}\sum_{s^i\in\mathcal{S}^i}\sum_{s\in b^i_{s^i}}\left(-\ln(\sqrt{2\pi}\sigma_{\widehat{\omega}_s})-\frac{(f_s-\mu_{\widehat{\omega}_s})^2}{2\sigma^2_{\widehat{\omega}_s}}\right)\\ =&\sum_{\lambda\in\Lambda}\sum_{i=0}^{M}\sum_{s^i\in\mathcal{S}^i_\lambda}\sum_{s\in b^i_{s^i}}\left(-\ln(\sqrt{2\pi}\sigma_{\lambda})-\frac{(f_s-\mu_{\lambda})^2}{2\sigma^2_{\lambda}}\right)\end{aligned}\tag{1.26}$$

where $\mathcal{S}^i_\lambda$ is the set of sites at level i where $\widehat{\omega}_{s^i}=\lambda$. Derivating with respect to μ_λ and σ_λ, we get a closed form solution for the ML estimates of the Gaussian parameters:

$$\forall\lambda\in\Lambda:\quad \mu_\lambda=\frac{1}{\sum_{i=0}^{M}\mid\mathcal{S}^i_\lambda\mid}\sum_{i=0}^{M}\sum_{s^i\in\mathcal{S}^i_\lambda}\sum_{s\in b^i_{s^i}}f_s,\quad \sigma^2_\lambda=\frac{1}{\sum_{i=0}^{M}\mid\mathcal{S}^i_\lambda\mid}\sum_{i=0}^{M}\sum_{s^i\in\mathcal{S}^i_\lambda}\sum_{s\in b^i_{s^i}}(f_s-\mu_\lambda)^2\tag{1.27}$$

Notice that a grey-level value f_s may be considered several times. More precisely, f_s is considered m-times in the above sum for a given λ if there are m scales where $\widehat{\omega}$ assigns the label λ to the site s. m can also be seen as a weight. Obviously, the more s has been labeled by λ at different levels, the more is probable that s belongs to class λ and hence its grey-level value f_s characterizes better the class λ. The derivates of the logarithmic likelihood function with respect to β and γ are given by:

$$\frac{\partial}{\partial\beta}\left(-\beta N^{ih}(\widehat{\omega})-\ln(Z(\beta,\gamma))\right) = -N^{ih}(\widehat{\omega})-\frac{\partial}{\partial\beta}\ln(Z(\beta,\gamma))\tag{1.28}$$

$$\frac{\partial}{\partial\gamma}\left(-\gamma \bar{N}^{ih}(\widehat{\omega})-\ln(Z(\beta,\gamma))\right) = -\bar{N}^{ih}(\widehat{\omega})-\frac{\partial}{\partial\gamma}\ln(Z(\beta,\gamma))\tag{1.29}$$

From which, we get

$$N^{ih}(\widehat{\omega}) = \frac{\sum_{\omega\in\Omega}N^{ih}(\omega)\exp(-\beta N^{ih}(\omega)-\gamma\bar{N}^{ih}(\omega))}{\sum_{\omega\in\Omega}\exp(-\beta N^{ih}(\omega)-\gamma\bar{N}^{ih}(\omega))}\tag{1.30}$$

$$\bar{N}^{ih}(\widehat{\omega}) = \frac{\sum_{\omega\in\Omega}\bar{N}^{ih}(\omega)\exp(-\beta N^{ih}(\omega)-\gamma\bar{N}^{ih}(\omega))}{\sum_{\omega\in\Omega}\exp(-\beta N^{ih}(\omega)-\gamma\bar{N}^{ih}(\omega))}\tag{1.31}$$

The solution of the above equations can be approximated using the following algorithm.

Algorithm 1.3.1 (Hyperparameter Estimation)

① *Set $k = 0$ and initialize $\widehat{\beta}^0$ and $\widehat{\gamma}^0$. Furthermore, let $N^{ih}(\widehat{\omega})$ denote the number of inhomogeneous cliques at the same scale and $\bar{N}^{ih}(\widehat{\omega})$ denotes the number of inhomogeneous cliques between levels.*

② *Using SA at a fixed temperature T, generate a new labeling η sampling from*

$$P(\mathcal{X} = \omega) = \frac{\exp\left(-\frac{\widehat{\beta}^k}{T}\sum_{i=0}^{M}\sum_{\{s,r\}\in\mathcal{C}^i}\delta(\omega_s,\omega_r)\right)}{Z(\widehat{\beta}^k,\widehat{\gamma}^k)} + \frac{\exp\left(-\frac{\widehat{\gamma}^k}{T}\sum_{\{s,r\}\in\bar{\mathcal{C}}}\delta(\omega_s,\omega_r)\right)}{Z(\widehat{\beta}^k,\widehat{\gamma}^k)}. \tag{1.32}$$

Compute the number of inhomogeneous cliques $N^{ih}(\eta)$ and $\bar{N}^{ih}(\eta)$ in η.

③ *If $N^{ih}(\eta) \approx N^{ih}(\widehat{\omega})$* and *$\bar{N}^{ih}(\eta) \approx \bar{N}^{ih}(\widehat{\omega})$ then stop, else $k = k + 1$. If $N^{ih}(\eta) < N^{ih}(\widehat{\omega})$ then decrease $\widehat{\beta}^k$, if $N^{ih}(\eta) > N^{ih}(\widehat{\omega})$ then increase $\widehat{\beta}^k$. $\widehat{\gamma}^k$ is obtained in the same way. Continue Step ② with $(\widehat{\beta}^k, \widehat{\gamma}^k)$.*

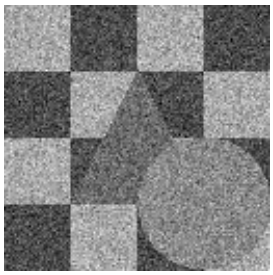
This algorithm completes the computation of the ML estimate of the parameters given $\widehat{\omega}$. The unsupervised segmentation is then carried out using *Adaptive Simulated Annealing* [38, 96], which is an iterative algorithm generating tentative labelings based on current parameter estimates (*i.e.* solving Eq. (1.23)) then updating the parameter values to their ML estimate based on the current labeling (*i.e.* solving Eq. (1.24) by making use of Eq. (1.27) and Algorithm 1.3.1). In fact, it is a classical Simulated Annealing with an additional step to reestimate model parameters during segmentation. The convergence of ASA has been proven in [137].

The algorithm has been tested on several synthetic and real images [25, 37, 38]. In Fig. 1.7, we show one of these results. In summary, the presented unsupervised algorithm provide results comparable to those obtained by supervised segmentations, but of course at the price of higher computing time.

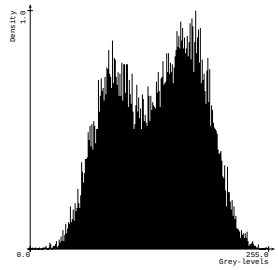
1.4 Application in remote sensing

Land cover classification is a common task in analysing satellite images [32]. Our MRF models can be readily applied to such problems as using appropriate sensors, different land properties can be distinguished based on the gray-level distribution of pixels. Herein we show two examples of SPOT satellite image segmentation using different models and stochastic optimizations techniques [14, 20–24, 38].

In Fig. 1.8 we present a SPOT image of size 512×512 with ground truth data (see Fig. 1.9). In Table 1.1, we give the mean (μ) and the variance (σ^2) for each of the 6 classes correspondig to different land coverages.



$SNR = 5$dB



Histogram

Supervised (0.6% error)

Unsupervised (0.68% error)

Parameter	Unsupervised		Supervised
	Initial	Final	
μ_0	83.5	84.3	85.48
σ_0^2	256.0	483.9	446.60
μ_1	100.0	115.5	115.60
σ_1^2	169.0	444.6	533.97
μ_2	152.5	146.7	146.11
σ_2^2	676.0	502.1	540.32
μ_3	181.5	177.9	178.01
σ_3^2	100.0	500.0	504.34
β	0.7	1.0	0.7
γ	0.1	0.1	0.1

Figure 1.7: *Supervised and unsupervised segmentation results and misclassification rate with the Gibbs Sampler. We also compare the parameters obtained by the unsupervised algorithm to the ones used for the supervised segmentation [25,36–38].*

class	1	2	3	4	5	6
μ	65.3	81.3	75.4	98.5	82.5	129.0
σ^2	6.4	12.7	14.9	16.8	9.46	183.2

Table 1.1: *Parameters of the "assalmer" image.*

As we can see, the classes 2 and 5 have nearly the same parameters, it is difficult to distinguish between them. Fig. 1.10 (resp. Fig. 1.11) shows the results obtained with the ICM (resp. the Gibbs Sampler). For these results, we give a map drawn by an expert (ground truth data, see Fig. 1.9). The classes $1-6$ correspond to the regions $B_{3c}, B_{3b}, B_{3d}, a_2, hc$ and 92_a on the map. For the hierarchical model a slight improvement can be noticed for the results of the Gibbs sampler, however, for the ICM, the improvement is more significant.

In Fig. 1.12, another SPOT image with 10 classes is presented with overlayed ground truth data (The regions are drawn by an expert (Unfortunately, they are shifted up by some pixels. Please take it into account when evaluating the results.) In Table 1.2, we give the mean (μ) and the variance (σ^2) for each class. Fig. 1.13 shows a supervised segmentation using the parameters listed in Table 1.2. Unsupervised result in Fig. 1.14 is comparable to the supervised one, but it requires more computing time and is more sensitive to noise. However, the main advantage is that unsupervised methods are completely data-driven. The only input parameter is the number of regions.

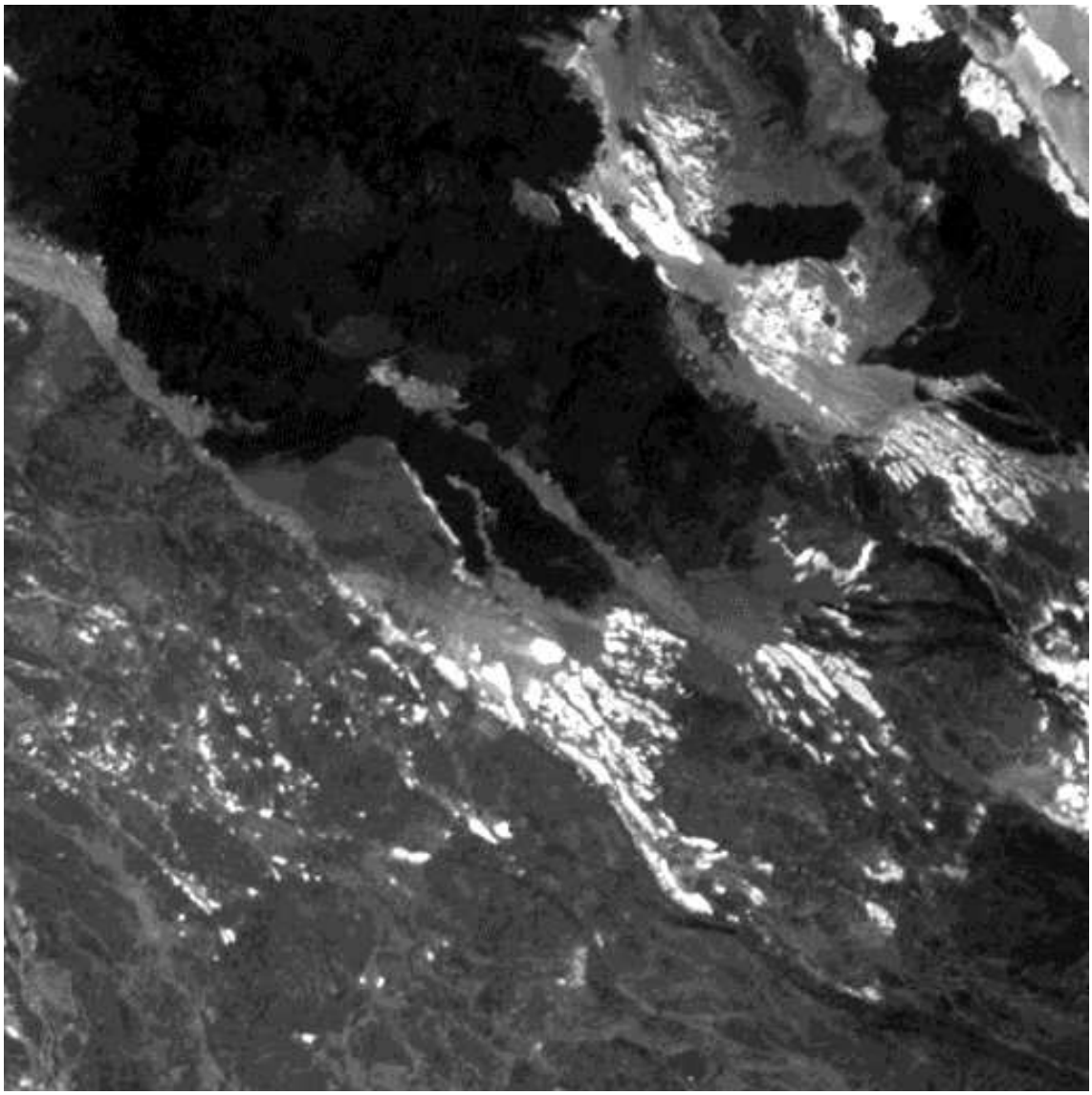


Figure 1.8: *Original SPOT image "assalmer" with 6 classes.*

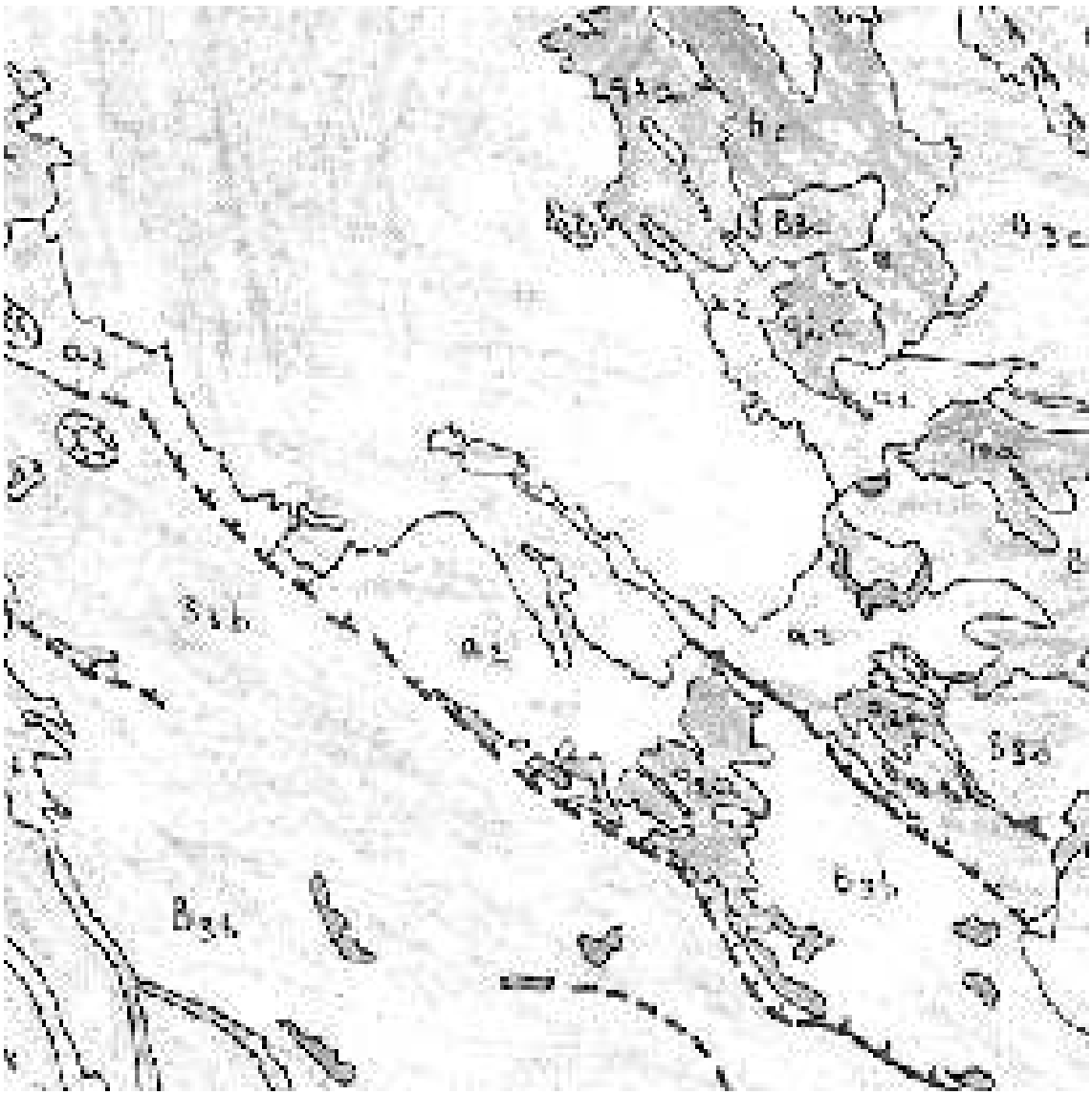


Figure 1.9: *Ground truth data.*

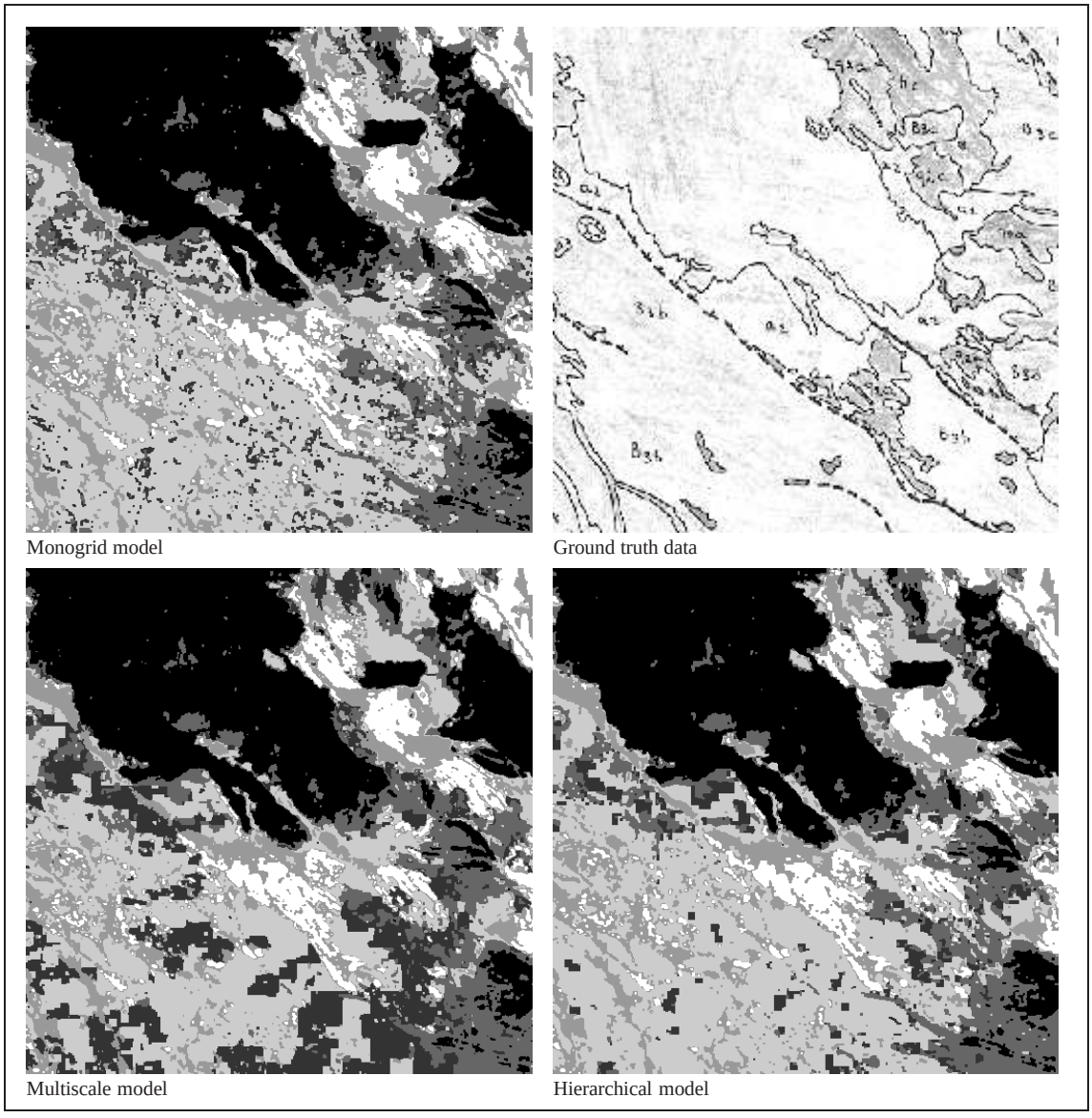


Figure 1.10: *Results of the ICM algorithm. Comparison with ground truth data.*

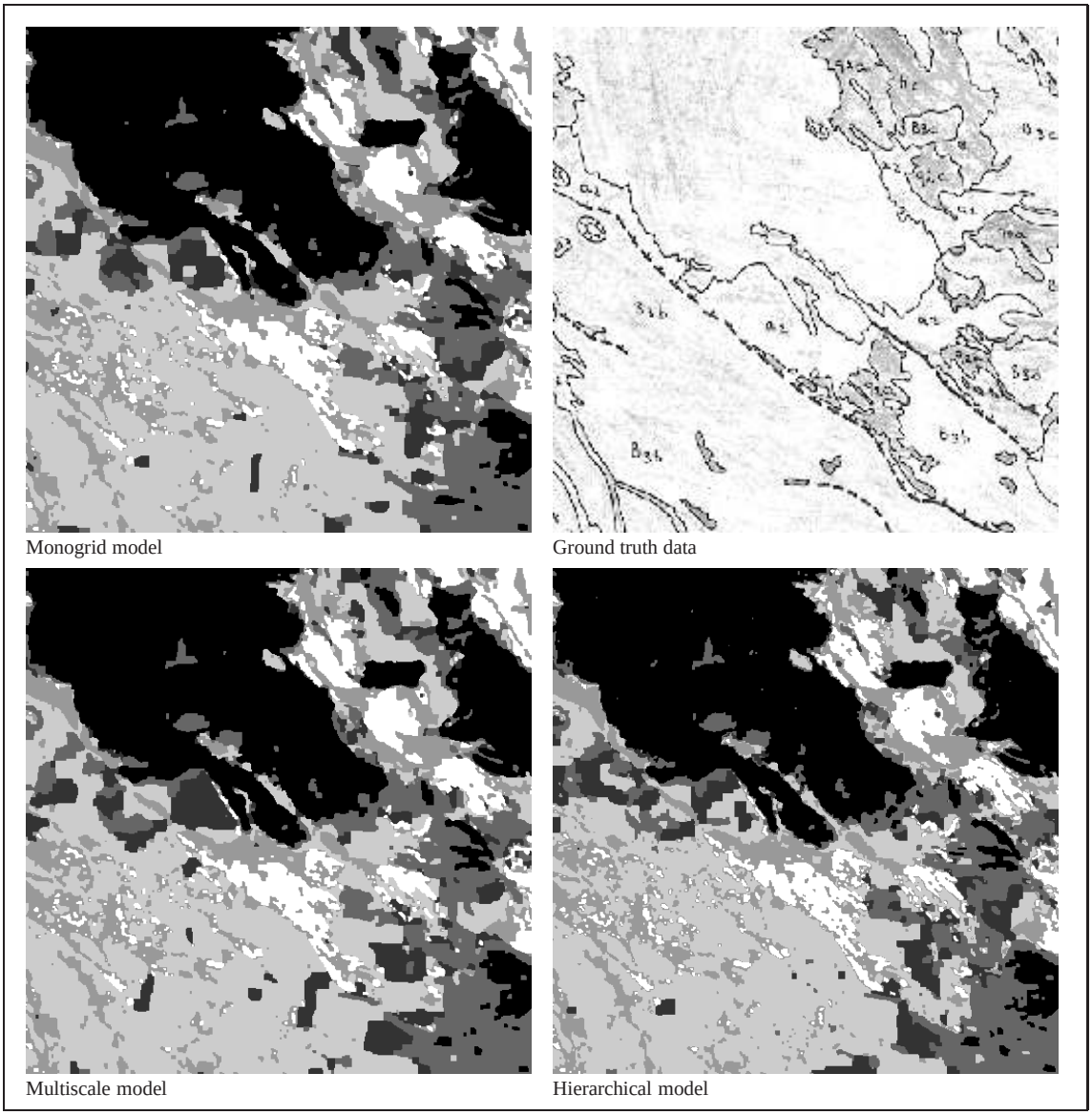


Figure 1.11: *Results of the Gibbs Sampler. Comparison with ground truth data.*

Figure 1.12: *Training areas on the "holland" image.*

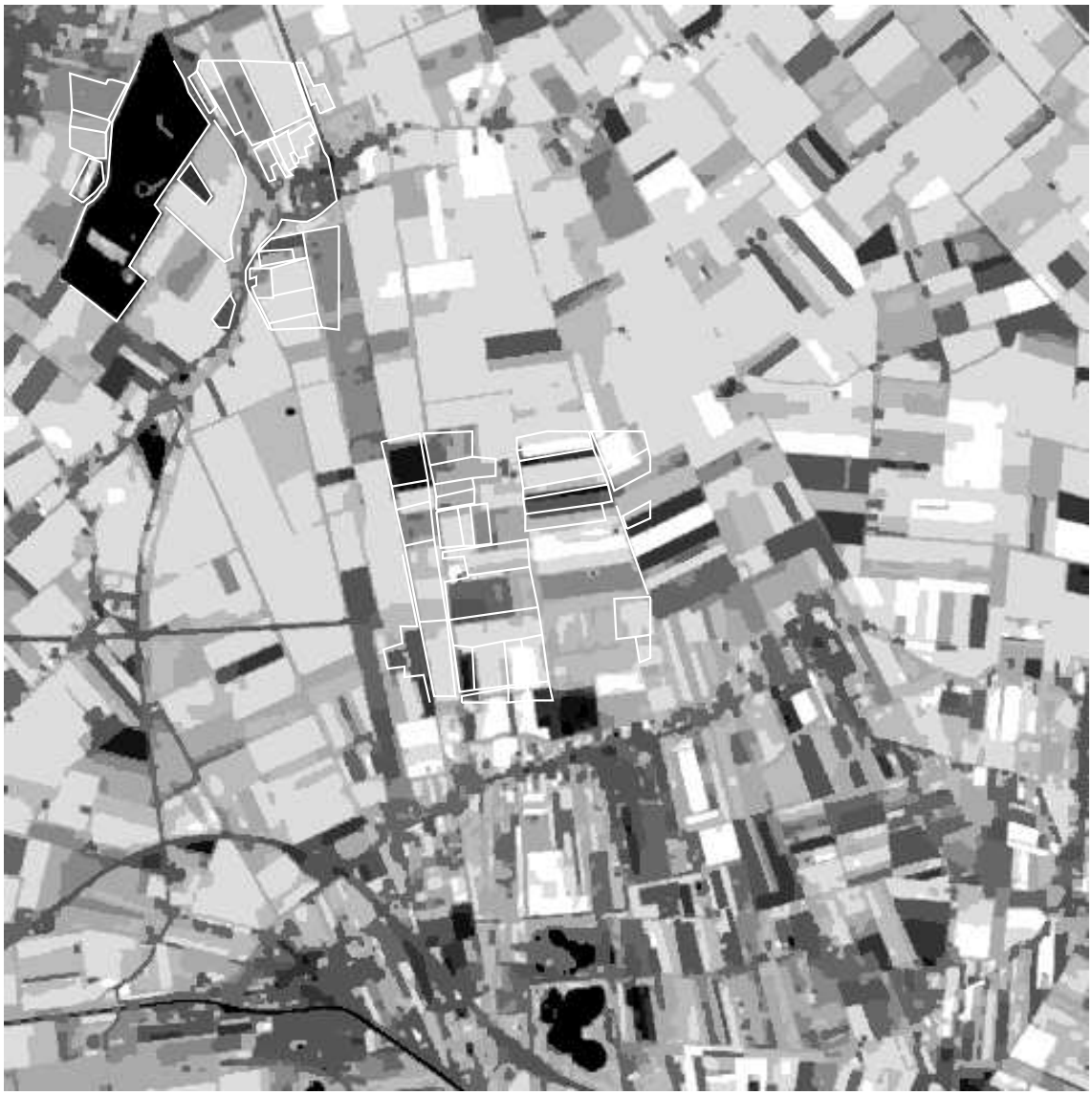


Figure 1.13: *Supervised segmentation result with* 10 *classes (Gibbs Sampler).*

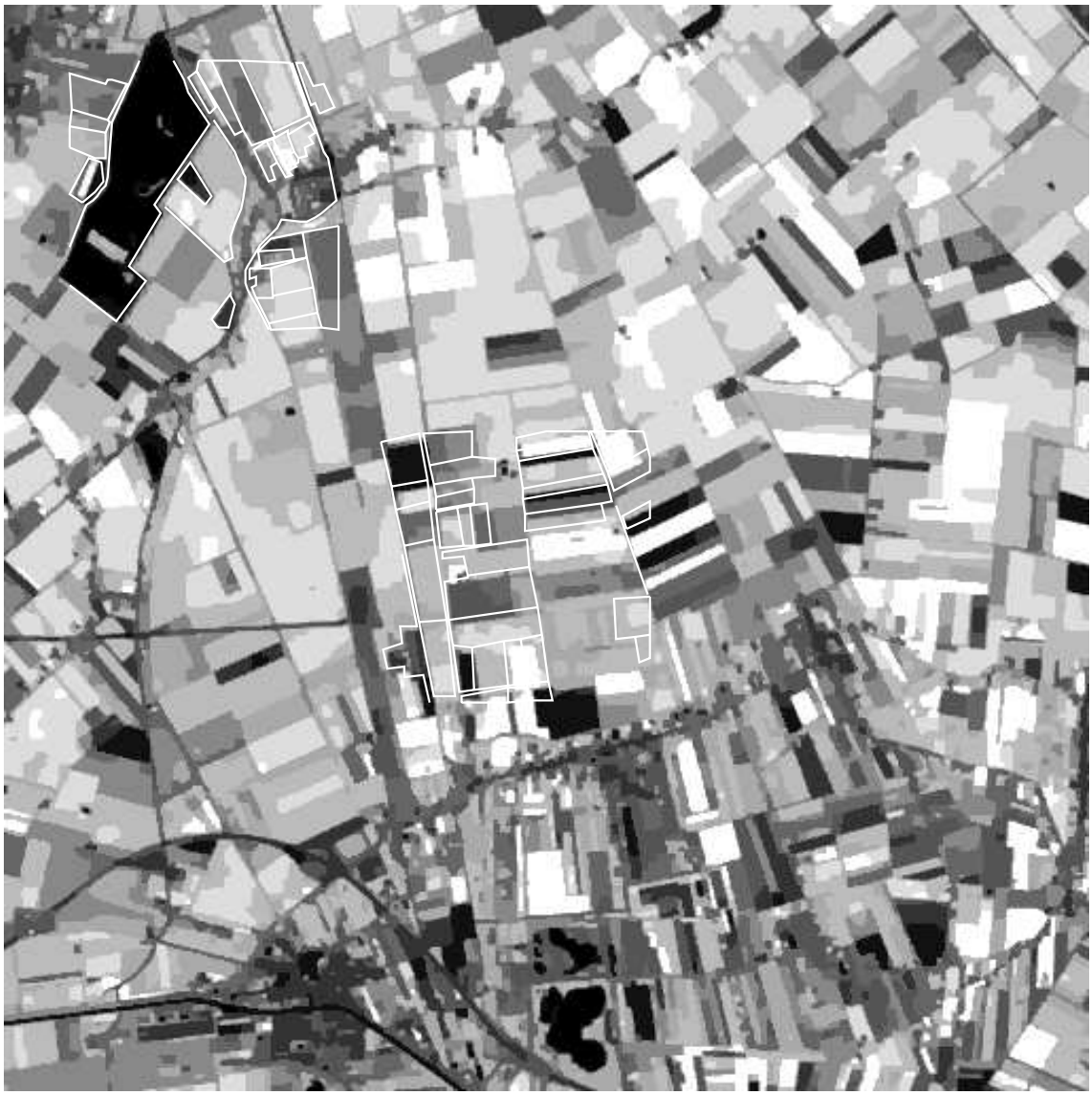


Figure 1.14: *Unsupervised segmentation result with* 10 *classes (Gibbs Sampler).*

class	1	2	3	4	5	6	7	8	9	10
μ	54.61	73.57	159.96	122.84	129.90	146.65	82.56	100.57	93.85	182.34
σ^2	93.10	4.10	31.31	8.90	37.42	15.83	35.58	308.86	93.71	73.18

Table 1.2: *Parameters of the "holland" image.*

IN THIS CHAPTER:

2.

Complex features and parameter estimation

In this chapter, we present our results on using more complex features (*e.g.* color, texture, motion) in MRF models and we also address the associated parameter estimation problems:

A monogrid MRF model which is able to combine color and texture features in order to improve the quality of unsupervised segmentations.

A novel RJMCMC sampling method which is able to identify multi-dimensional Gaussian mixtures. This technique has been applied to fully automatic color image segmentation.

A new multilayer MRF model has been proposed which is able to segment an image based on multiple cues (such as color, texture, or motion).

Application areas include motion segmentation (a crucial step in *e.g.* MPEG coding) as well as change detection in aerial images.

2.1 Introduction

There are many features that one can take as observation for the segmentation process: gray-level, color, motion, different texture features, etc. However, most of the segmentation algorithms presented in the literature are based on only one of these features.

One way to combine various features is to design a joint probability distribution which is able to represent the union of the complex observation. This approach works well when the combined features are of similar nature (*e.g.* define a multivariate Gaussian density). Such a model is proposed in our work [28] for color textured image segmentation.

However, the human visual system is not treating different features jointly. Instead, multiple cues are perceived simultaneously but in a parallel fashion and then they are integrated by our visual system in order to explain the observations. Following these ideas, we have developed multi-layer Markovian models and successfully applied it to color-texture [30, 31] and color-motion segmentation [27, 29]. For example, an important problem is extracting regions of object motions in the presence of camera drift. This is a key issue in several applications of aerial imagery. In surveillance and exploitation tasks [135] it can be used as a preliminary step of object detection, tracking and event analysis. On the other hand, in 2-D mosaicking [168] and in 3-D stereo reconstruction [55] independent object motions generate outlier regions for image alignment, thus, they should be detected and skipped from the resulting static scene models. An efficient solution to this problem consists in a three-layer Markov Random Field which integrates two different features to statistically characterize the background membership of the pixels [1, 2].

In the following, we give a brief overview of these approaches.

2.2 Unsupervised segmentation of color textured images

The proposed segmentation model [26, 28] consists of a monogrid MRF defined over a nearest neighborhood system (see Fig. 1.1) with pixel classes represented by multivariate Gaussian distributions. This kind of modelization corresponds well to our features: Texture feature images (extracted by Gabor filters) are constructed in such a way that similar textures map to similar intensities. Hence pixels with a given texture will be assigned a well determined value with some variance. Furthermore, pixels with similar color map to their average color. Putting these feature distributions into one multivariate Normal mixture, the modes will correspond to clusters of pixels which are homogeneous in both color and texture properties. Therefore regions will be formed where both features are homogeneous while boundaries will be present where there is a discontinuity in either color or texture. Applying these ideas, the *image process* $\mathcal{F}$ can be formalized as follows: $P(\mathbf{f}_s \mid \omega_s)$ follows a Normal distribution $N(\mu, \mathbf{\Sigma})$, each pixel class $\lambda \in \Lambda = \{1, 2, \ldots, L\}$ is represented by its mean vector μ_λ and covariance matrix $\mathbf{\Sigma}_\lambda$. The whole posterior can now be expressed as a first order MRF by including the contribution of the likelihood term via the singletons (*i.e.* pixel sites

$s \in \mathcal{S}$). Indeed, the singleton energies directly reflect the probabilistic modeling of labels without context, while doubleton clique potentials express relationship between neighboring pixel labels. Thus the energy function of the so defined MRF image segmentation model has the following form:

$$\sum_{s\in\mathcal{S}} \left(\ln(\sqrt{(2\pi)^n \mid \mathbf{\Sigma}_{\omega_s} \mid}) + \frac{1}{2}(\mathbf{f}_s - \mu_{\omega_s})\mathbf{\Sigma}_{\omega_s}^{-1}(\mathbf{f}_s - \mu_{\omega_s})^T \right) + \beta \sum_{\{s,r\}\in\mathcal{C}} \delta(\omega_s, \omega_r) \qquad (2.1)$$

where $\beta > 0$ is a weighting parameter controlling the importance of the prior. As β increases, the resulting regions become more homogeneous.

The proposed segmentation model has the following parameters:

1. The weight β of the prior term,
2. the number of pixel classes L,
3. the mean vector μ_λ and covariance matrix $\mathbf{\Sigma}_\lambda$ of each class $\lambda \in \Lambda$.

The automatic determination of L will be addressed in Section 2.3. While L strongly depends on the input image data, β is largely independent of it. Experimental evidence suggests that the model is not sensitive to a particular setting of β [26, 28]. We found that setting $\beta \geq 2.0$ gives satisfactory and stable segmentations. Unlike the first two parameters, the mean and covariance of the Gaussians must be computed directly from the input image. Our solution to this problem [28] adopts a general iterative algorithm, known as the *EM algorithm*, to compute the maximum likelihood estimates of the parameters of a mixture density. Basically, we will fit a Gaussian mixture of L components to the histogram of the image features. The observations consist of the histogram data $\mathbf{d}_i (i = 1, \ldots, D)$ of the feature images. D denotes the number of histogram points and the dimension of a data point equals to the dimension of the combined color-texture feature space. Assuming there are L classes, we want to estimate the mean values μ_λ and covariance matrices $\mathbf{\Sigma}_\lambda$ for each pixel class $\lambda \in \Lambda$.

The *EM algorithm* aims at finding parameter values which maximize the normalized log-likelihood function:

$$\mathcal{L} = \frac{1}{D} \sum_{i=1}^{D} \log \left(\sum_{\lambda\in\Lambda} P(\mathbf{d}_i \mid \lambda) P(\lambda) \right) \qquad (2.2)$$

The underlying model is that the *complete data* includes not only the observable $\mathbf{d}_i$ but also the *hidden data* labels ℓ_i specifying which Gaussian process generated the data $\mathbf{d}_i$. Actually, ℓ_i is also a vector of dimension L and $\ell_i^\lambda = 1$ if $\mathbf{d}_i$ belongs to class λ and 0 otherwise. The idea is that if labels were known, the estimation of model parameters would be equivalent to the supervised case. Hence the following algorithm is alternating two steps: The estimation of a tentative labeling of the data followed by updating the parameter values based on the tentatively labeled data.

Algorithm 2.2.1 (EM for Gaussian mixture identification)

① **[Estimation]** *Replace ℓ_i with its conditional expectation based on the current parameter estimates. Since the labels may only take values 0 or 1, the expectation is basically equivalent to the posterior probability:*

$$P(\lambda \mid \mathbf{d}_i) = \frac{P(\mathbf{d}_i \mid \lambda)P(\lambda)}{\sum_{\lambda \in \Lambda} P(\mathbf{d}_i \mid \lambda)P(\lambda)}, \tag{2.3}$$

where $P(\lambda)$ denotes the component weight.

② **[Maximization]** *Then, using the current expectation of the labels ℓ_i as the current labeling of the data, the estimation of the parameters is simple:*

$$P(\lambda) = \frac{K_\lambda}{D} \tag{2.4}$$

$$\mu_\lambda = \frac{1}{K_\lambda} \sum_{i=1}^{D} P(\lambda \mid \mathbf{d}_i)\mathbf{d}_i \tag{2.5}$$

$$\mathbf{\Sigma}_\lambda = \frac{1}{K_\lambda} \sum_{i=1}^{D} P(\lambda \mid \mathbf{d}_i)(\mathbf{d}_i - \mu_\lambda)^T(\mathbf{d}_i - \mu_\lambda) \tag{2.6}$$

where $K_\lambda = \sum_{i=1}^{D} P(\lambda \mid \mathbf{d}_i)$. Basically the posteriors $P(\lambda \mid \mathbf{d}_i)$ are used as a weight of the data vectors. They express the contribution of a particular data point $\mathbf{d}_i$ to the class λ.

③ *Go to Step ① until convergence. Each iteration is guaranteed to increase the likelihood of the estimates. The algorithm is stopped when the change of the log-likelihood $\mathcal{L}$ is less than a predetermined threshold (our test cases used 10^{-7}).*

The proposed algorithm has been tested on a variety of color images. We compared segmentation results using color-only, texture-only and combined (color+texture) features [26, 28] and found in all test-cases that segmentation based purely on texture gives fuzzy boundaries but usually homogeneous regions, whereas segmentation based on color is more sensitive to local variations but provides sharp boundaries. As for the combined features, the advantages of both color and texture based segmentation have been preserved: we obtained sharp boundaries and homogeneous regions. Results has also been compared to those obtained by the JSEG algorithm [82], a recent unsupervised method for color textured image segmentation. Our method clearly outperforms JSEG (see Fig. 2.1) but JSEG's advantage is that we do not have to specify the image dependent parameter L.

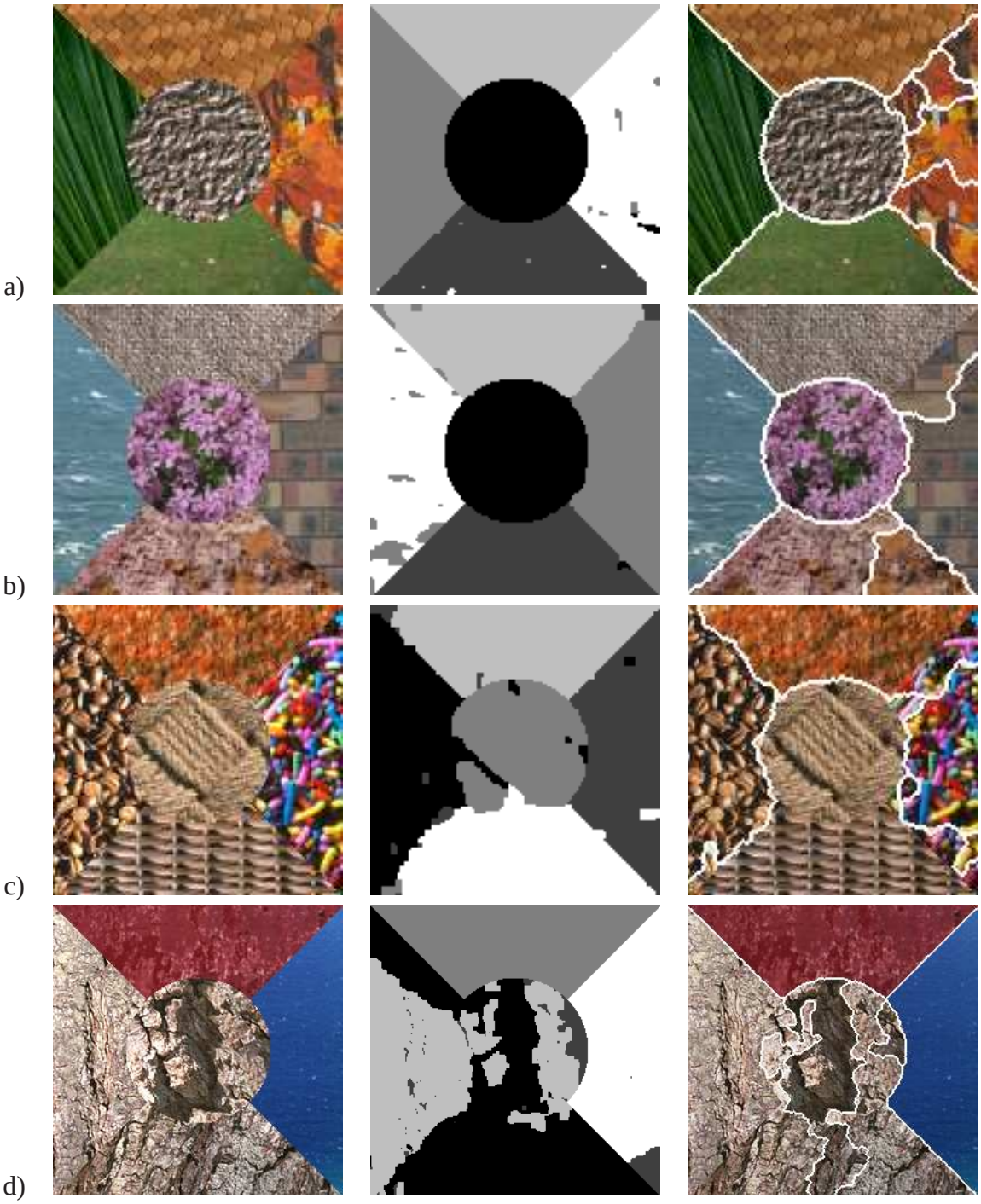


Figure 2.1: *Unsupervised segmentation results on color textured images, each with 5 classes [28].*

2.3 Segmentation of color images via reversible jump MCMC sampling

Our problem becomes much harder when the number of labels L is also unknown. We have addressed this problem in the context of color-based image segmentation [15, 16]. When this parameter is also being estimated, the unsupervised segmentation problem may be treated as a *model selection* problem over a combined model space. From this point of view, L becomes a *model indicator* and the observation $\mathcal{F}$ is regarded as a three-variate Normal *mixture* with L components corresponding to clusters of pixels which are homogeneous in color.

The goal of our analysis is inference about the number L of Gaussian mixture components (each one corresponds to a label), the component parameters $\Theta = \{\Theta_\lambda = (\mu_\lambda, \Sigma_\lambda) \mid \lambda \in \Lambda\}$, the component weights p_λ summing to 1, the inter-pixel interaction strength β, and the segmentation ω. A broadly used tool to sample from the posterior distribution is the Metropolis-Hastings method. Classical methods, however, can not be used due to the changing dimensionality of the parameter space. To overcome this limitation, a promising approach, called Reversible Jump MCMC (RJMCMC), has been adopted [15, 16]. When we have multiple parameter subspaces of different dimensionality, it is necessary to devise different *move types* between the subspaces. These will be combined in a so called *hybrid sampler*. For the color image segmentation model, the following move types are needed [15, 16]:

1. sampling the labels ω (*i.e.* re-segment the image);
2. sampling Gaussian parameters $\Theta = \{(\mu_\lambda, \Sigma_\lambda)\}$;
3. sampling the mixture weights $p_\lambda (\lambda \in \Lambda)$;
4. sampling the MRF hyperparameter β;
5. sampling the number of classes L (splitting one mixture component into two, or combining two into one).

The only randomness in scanning these move types is the random choice between splitting and merging in move (55). One iteration of the hybrid sampler, also called a *sweep*, consists of a complete pass over these moves. The first four move types are conventional in the sense that they do not alter the dimension of the parameter space. Hereafter, we will focus on move (55), which requires the use of the reversible jump mechanism. This move type involves changing L by 1 and making necessary corresponding changes to ω, Θ and p.

The *split proposal* begins by randomly choosing a class λ with a uniform probability $P^{split}_{select}(\lambda) = 1/L$. Then L is increased by 1 and λ is split into λ_1 and λ_2. In doing so, a new set of parameters need to be generated. Altering L changes the dimensionality of the variables Θ and p. Thus we shall define a deterministic function ψ as a function of these Gaussian mixture parameters:

$$(\Theta^+, p^+) = \psi(\Theta, p, u) \tag{2.7}$$

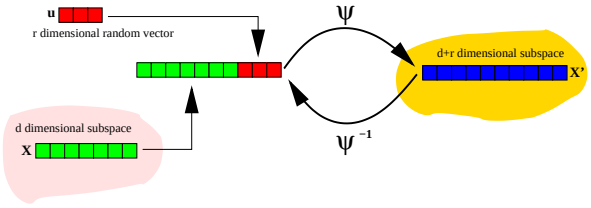


Figure 2.2: ψ *is a* diffeomorphism *which transforms back and forth between parameter subspaces of different dimensionality [15, 16].* Dimension matching *can be implemented by generating a random vector* u *such that the dimensions of* (X, u) *and* X' *are equal.*

where the superscript $+$ denotes parameter vectors after incrementing L. u is a set of random variables having as many elements as the degree of freedom of joint variation of the current parameters (Θ, p) and the proposal (Θ^+, p^+). Note that this definition satisfies the *dimension matching* constraint (see Fig. 2.2), which guarantees that one can jump back and forth between different parameter sub-spaces [15, 16]. This is needed to ensure the convergence of simulated annealing towards a global optimum. The new parameters of λ_1 and λ_2 are assigned by matching the 0^{th}, 1^{th}, 2^{th} moments of the component being split to those of a combination of the two new components [15, 16]:

$$p_\lambda = p^+_{\lambda_1} + p^+_{\lambda_2} \tag{2.8}$$

$$p_\lambda \mu_\lambda = p^+_{\lambda_1}\mu^+_{\lambda_1} + p^+_{\lambda_2}\mu^+_{\lambda_2} \tag{2.9}$$

$$p_\lambda(\mu_\lambda \mu_\lambda^T + \mathbf{\Sigma}_\lambda) = p^+_{\lambda_1}(\mu^+_{\lambda_1}\mu^{+T}_{\lambda_1} + \mathbf{\Sigma}^+_{\lambda_1}) + p^+_{\lambda_2}(\mu^+_{\lambda_2}\mu^{+T}_{\lambda_2} + \mathbf{\Sigma}^+_{\lambda_2}) \tag{2.10}$$

There are 10 degrees of freedom in splitting λ since covariance matrices are symmetric. Therefore, we need to generate a random variable $u1$, a random vector $\mathbf{u2}$ and a symmetric random matrix $\mathbf{u3}$. We can now define the diffeomorphism ψ which transforms the old parameters (Θ, p) into the new (Θ^+, p^+) using the above moment equations and the random numbers $u1$, $\mathbf{u2}$, and $\mathbf{u3}$ [15, 16]:

$$p^+_{\lambda_1} = p_\lambda u1 \tag{2.11}$$

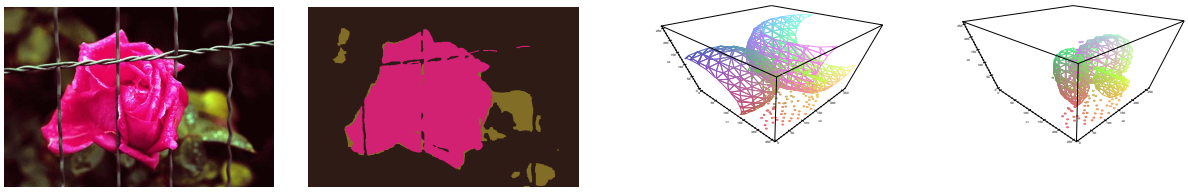
$$p^+_{\lambda_2} = p_\lambda(1-u1) \tag{2.12}$$

$$\mu^+_{\lambda_1,i} = \mu_{\lambda,i} + u2_i\sqrt{\Sigma_{\lambda,i,i}\frac{1-u1}{u1}} \tag{2.13}$$

$$\mu^+_{\lambda_2,i} = \mu_{\lambda,i} - u2_i\sqrt{\Sigma_{\lambda,i,i}\frac{u1}{1-u1}} \tag{2.14}$$

$$\Sigma^+_{\lambda_1,i,j} = \begin{cases} u3_{i,i}\left(1-{u2_i}^2\right)\Sigma_{\lambda,i,i}\dfrac{1}{u1} & \text{if } i=j \\ u3_{i,j}\Sigma_{\lambda,i,j}\sqrt{\left(1-{u2_i}^2\right)}\sqrt{\left(1-{u2_j}^2\right)u3_{i,i}u3_{j,j}} & \text{if } i\neq j \end{cases} \tag{2.15}$$

$$\Sigma^+_{\lambda_2,i,j} = \begin{cases} \left(1-u3_{i,i}\right)\left(1-{u2_i}^2\right)\Sigma_{\lambda,i,i}\dfrac{1}{u1} & \text{if } i=j \\ \left(1-u3_{i,j}\right)\Sigma_{\lambda,i,j}\sqrt{\left(1-{u2_i}^2\right)\left(1-{u2_j}^2\right)}\sqrt{\left(1-u3_{i,i}\right)\left(1-u3_{j,j}\right)} & \text{if } i\neq j \end{cases} \tag{2.16}$$



Original image | Segmentation (3 labels) | Initial Gaussians | Final (3 classes)

Figure 2.3: *Segmentation of image* rose41 *and the estimated Gaussian mixture [15, 16].*

The random variables u are chosen from the interval $(0, 1]$. In order to favor splitting a class into roughly equal portions, beta$(1.1, 1.1)$ distributions are used. The next step is the reallocation of those sites s where $\omega_s = \lambda$. This reallocation is based on the new parameters and has to be completed in such a way as to ensure the resulting labeling ω^+ is drawn from the posterior distribution with $\Theta = \Theta^+$, $p = p^+$ and $L = L + 1$.

Merging of a pair (λ_1, λ_2) is basically the inverse of the split operation [15, 16].

Finally, the split or merge proposal is accepted with a probability relative to the probability ratio of the current and the proposed states. The segmentation and parameter estimation is then obtained as a MAP estimation implemented via simulated annealing:

Algorithm 2.3.1 (RJMCMC Segmentation)

① *Set* $k = 0$. *Initialize* $\widehat{\beta}^0$, $\widehat{L}^0$, $\widehat{p}^0$, $\widehat{\Theta}^0$, *and the initial temperature* T_0.

② *A sample* $(\widehat{\omega}^k, \widehat{L}^k, \widehat{p}^k, \widehat{\beta}^k, \widehat{\Theta}^k)$ *is drawn from the posterior distribution using the* hybrid sampler *outlined earlier. Each sub-chain is sampled via the corresponding move-type while all the other parameter values are set to their current estimate.*

③ *Goto Step* ② *with* $k = k + 1$ *and* T_{k+1} *until* $k < \mathcal{K}$.

As usual, an exponential annealing schedule ($T_{k+1} = 0.98T_k$, $T_0 = 6.0$) was chosen so that the algorithm would converge after a reasonable number of iterations. In our experiments, the algorithm was stopped after 200 iterations ($T_{200} \approx 0.1$).

The proposed algorithm has been tested [15, 16] on a variety of real color images and results have also been compared to those produced by JSEG [82]. In Fig. 2.4, we show a couple of results obtained on the Berkeley Segmentation Dataset, and in Fig. 2.5, we plot the corresponding precision-recall curves. Note that RJMCMC has a slightly higher *F-measure* which ranks it over JSEG. However, it is fair to say that both method perform equally well but behave differently: while JSEG tends to smooth out fine details (hence it has a higher precision but lower recall value), RJMCMC prefers to keep fine details at the price of producing more edges (*i.e.* its recall values are higher at a lower precision value).

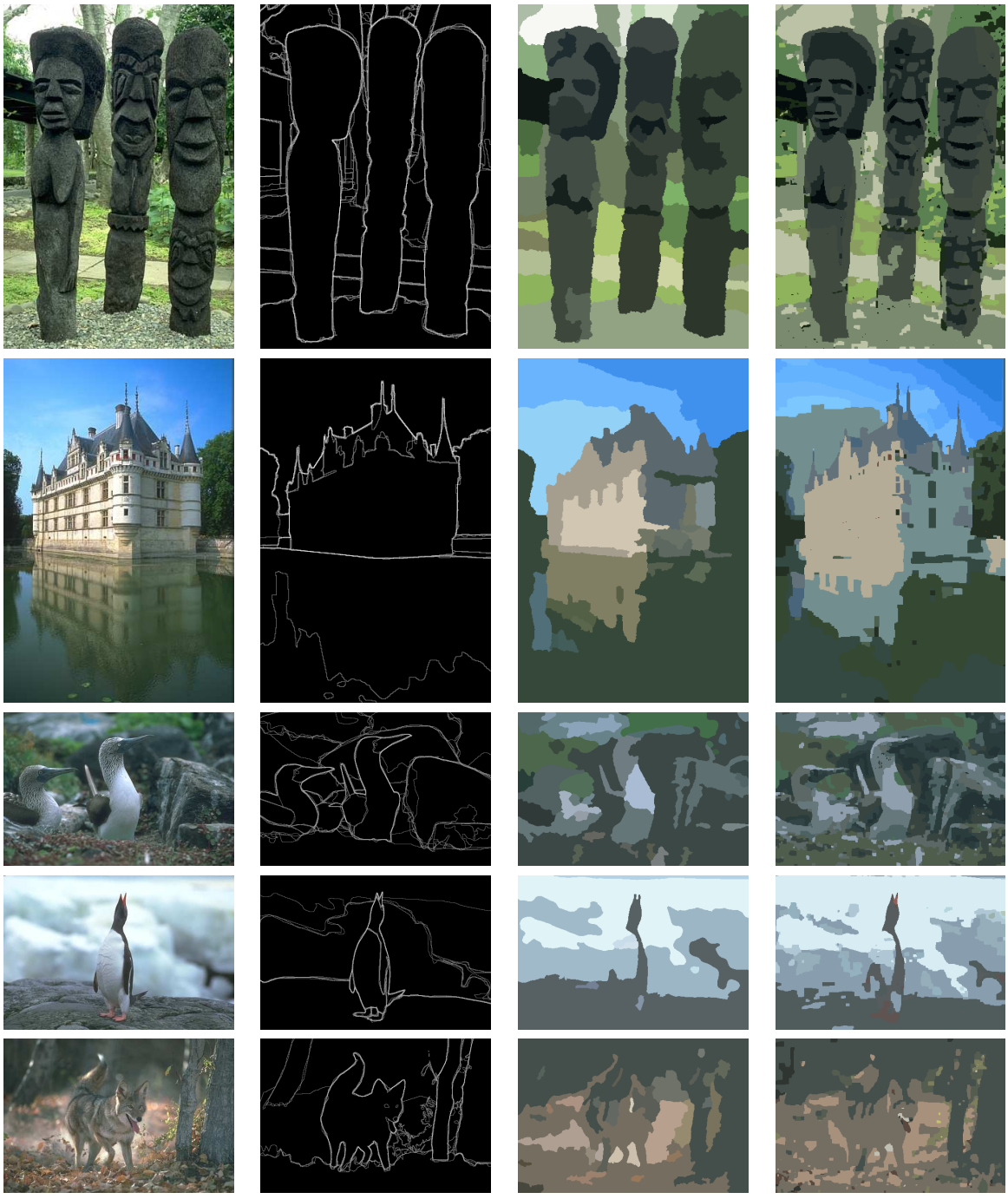
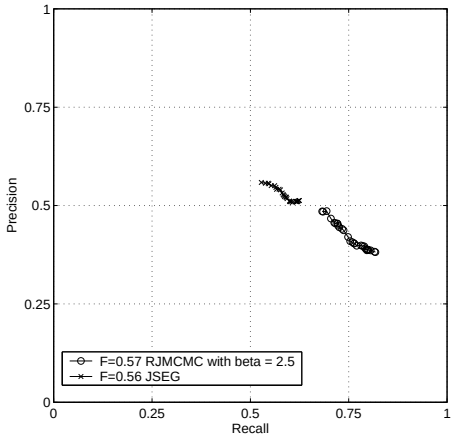


Figure 2.4: *Benchmark results on images from the Berkeley Segmentation Dataset [16]*



Method	F-measure	CPU time
Human segmentation	0.79	—
RJMCMC	0.57	15 min
JSEG	0.56	2 min

Figure 2.5: *Precision-recall curve, F-measure and CPU time comparison for JSEG and RJMCMC [16].*

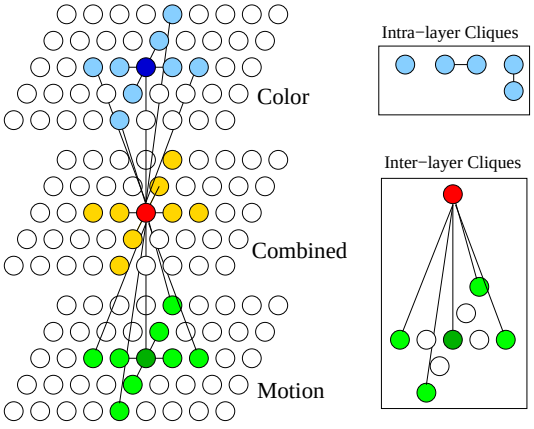


Figure 2.6: *Multi-layer MRF model [27,29].*

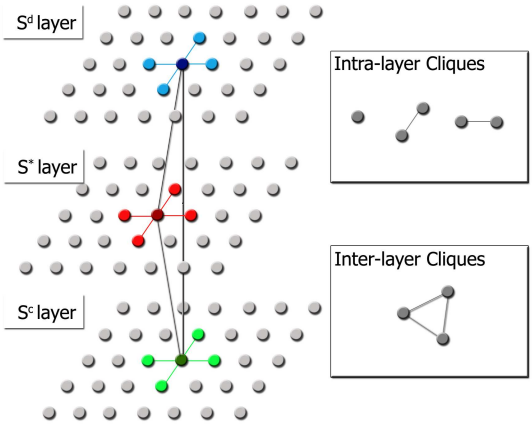


Figure 2.7: *Three-layer MRF model for change detection [2].*

2.4 Multilayer MRF modelization

The human visual system is not treating different features sequentially. Instead, multiple cues are perceived simultaneously and then they are integrated by our visual system in order to explain the observations. Therefore different image features has to be handled in a parallel fashion. We have developed such a model in a Markovian framework and successfully applied it to color-texture [30, 31] and color-motion segmentation [1, 2, 27, 29]. Herein, we present the MRF image segmentation model which aims at combining color and motion features for video object segmentation [27, 29]. The model has a multi-layer structure (see Fig. 2.6): Each feature has its own layer, called *feature layer*, where an MRF model is defined using only the corresponding feature. A special layer is assigned to the combined MRF model. This layer interacts with each feature layer and provides the segmentation based on the combination of different features. Unlike previous methods, our approach doesn't assume motion boundaries being part of spatial ones. The uniqueness of the proposed method is the ability to detect boundaries that are visible only in the motion feature as well as those visible only in the color one.

Perceptually uniform color values and precomputed optical flow data is used as features for the segmentation. The proposed model consists of 3 layers. At each layer, we use a first order neighborhood system and extra inter-layer cliques (Fig. 2.6). The image features are represented by multivariate Gaussian distributions. For example, on the color layer, the observed image $\mathcal{F}^c = \{\mathbf{f}_s^c | s \in \mathcal{S}^c\}$ consists of three spectral component values ($\mathrm{L^*u^*v^*}$) at each pixel s denoted by the vector $\mathbf{f}_s^c$. The class label assigned to a site s on the color layer is denoted by ψ_s. The energy function $U(\psi, \mathcal{F}^c)$ of the so defined MRF layer has the following form:

$$\sum_{s \in \mathcal{S}^c} \mathcal{G}^c(\mathbf{f}_s^c, \psi_s) + \beta \sum_{\{s,r\} \in \mathcal{C}} \delta(\psi_s, \psi_r) + \sum_{s \in \mathcal{S}^c} V^c(\psi_s, \eta_s^c)$$

where $\mathcal{G}^c(\mathbf{f}_s^c, \psi_s)$ denotes the Gaussian energy term. The last term ($V^c(\psi_s, \eta_s^c)$) is the inter-layer clique potential. The motion layer adopts a similar energy function with some obvious substitutions (*i.e.* for simplicity, we assume a translational motion model here – for a more elaborate model see [29]).

The combined layer only uses the motion and color features indirectly, through inter-layer cliques. A label consists of a pair of color and motion labels such that $\eta = \langle \eta^c, \eta^m \rangle$, where $\eta^c \in \Lambda^c$ and $\eta^m \in \Lambda^m$. The set of labels is denoted by $\Lambda^x = \Lambda^c \times \Lambda^m$ and the number of classes $L^x = L^c L^m$. Obviously, not all of these labels are valid for a given image. Therefore the combined layer model also estimates the number of classes and chose those pairs of motion and color labels which are actually present in a given image. The energy function $U(\eta)$ is of the following form:

$$\sum_{s \in \mathcal{S}^x} (V_s(\eta_s) + V^c(\psi_s, \eta_s^c) + V^m(\phi_s, \eta_s^m)) + \alpha \sum_{\{s,r\} \in \mathcal{C}} \delta(\eta_s, \eta_r)$$

where $V_s(\eta_s)$ denotes singleton energies, $V^c(\psi_s, \eta_s^c)$ (resp. $V^m(\phi_s, \eta_s^m)$ denotes inter-layer clique potentials. The last term corresponds to second order intra-layer cliques which ensures

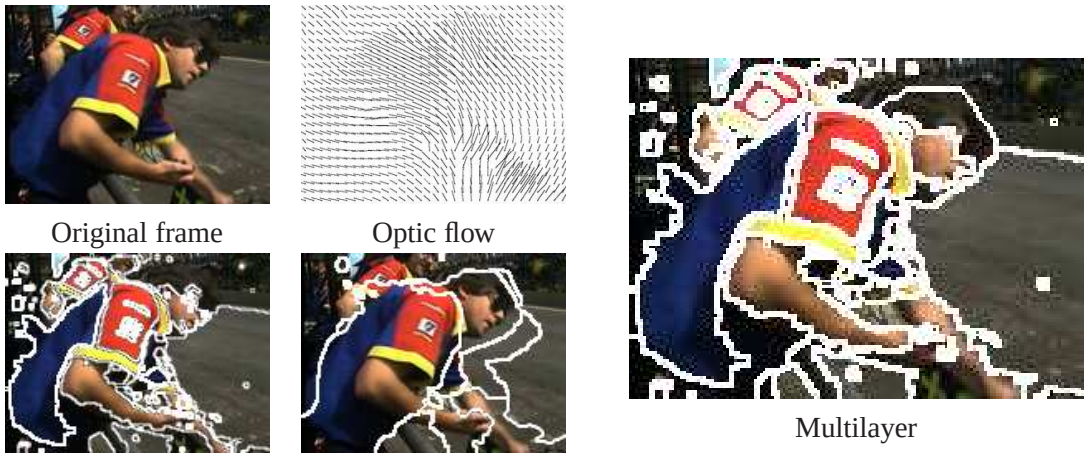


Figure 2.8: *Segmentation results [27, 29].*

Figure 2.9: *Comparison of the segmentation results obtained by the proposed method [27, 29] and those produced by the algorithm of Khan & Shah [130].*

homogeneity of the combined layer. α has the same role as β in the color layer model and $\delta(\eta_s, \eta_r) = -1$ if $\eta_s = \eta_r$, 0 if $\eta_s \neq \eta_r$ and 1 if $\eta_s^c = \eta_r^c$ *and* $\eta_s^m \neq \eta_r^m$ or $\eta_s^c \neq \eta_r^c$ *and* $\eta_s^m = \eta_r^m$. The idea is that region boundaries present at both color and motion layers are preferred over edges that are found only at one of the feature layers. At any site s, we have 5 inter-layer interactions between two layers: Site s interacts with the corresponding site on the other layer as well as with the 4 neighboring sites two steps away (see Fig. 2.6). This potential is based on the difference of the first order potentials at the corresponding feature layers. Clearly, the difference is 0 if and only if both the feature layer and the combined layer has the same label. If the labels are different then it is proportional to the energy difference between the two labels. Finally, the singleton energy controls the number of classes at the combined layer by penalizing small classes.

2.4.1 Application to motion segmentation and change detection

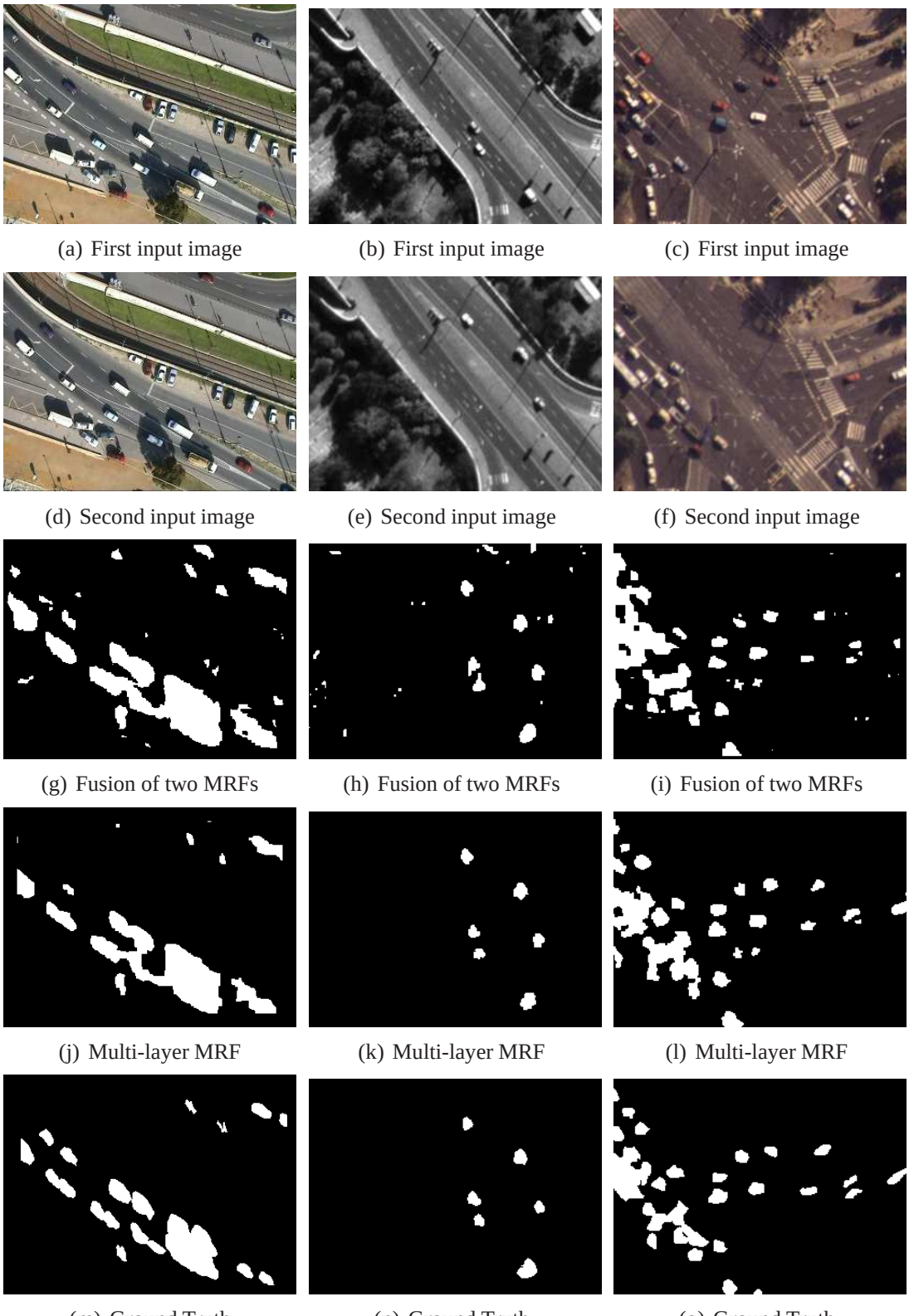
The proposed algorithm has been tested on real video sequences [27, 29]. We also compare the results to motion only and color only segmentation (basically a monogrid model similar to the one defined for the feature layers but without inter-layer cliques). The mean vectors and covariance matrices were computed over representative regions selected by the user. The number of motion and color classes is known a priori but classes on the combined layer are estimated during the segmentation process. Fig. 2.8 shows some segmentation results. Note that the head of the men on this image can only be separated from the background using motion features. Clearly, the multi-layer model provides significantly better results compared to color only and motion only segmentations. See Fig. 2.9 to compare the performance of the proposed method with the one from [130] on the *Mother and Daughter* standard sequence. Some of the contours are lost by [130] (the sofa, for example) while our method successfully identifies region boundaries. In particular, our method is able to separate the hand of the mother from the face of the daughter in spite of their similar color. This demonstrates again that the proposed method is quite powerful in combining motion and color features in order to detect boundaries visible only in one of the features. We can also handle occlusion and more complex motions using a modified multilayer model presented in [29]. The model has also been successfully applied to color-textured image segmentation [30, 31].

Finally, we present the application of multilayer modeling for automatic change detection on airborne images taken with moving cameras [2]. Essentially, we want to extract the accurate silhouettes of moving objects or object-groups in images taken by moving airborne vehicles in consecutive moments. This problem is solved in two steps: first a coarse (but robust) image registration is performed for camera motion compensation, then the aligned input image pair is segmented into moving (foreground) objects and background. Main challanges are camera motion, noise and the parallax artifacts caused by the static objects having considerable height (buildings, trees, walls etc.) from the difference image.

A three-layer MRF model is constructed on a graph $\mathcal{G}$ whose structure is shown in Fig. 2.7. The final goal is to perform a binary segmentation of the images into foreground (fg) and background (bg) classes. For the segmentation, two type of features are extracted from the aligned image pair: $d(s)$, the gray level difference of the corresponding pixels in the registered images; and $c(s)$, the maxima in the local correlation function around pixel s. The sites of $\mathcal{G}$ are arranged into three layers: S^d, S^c and S^*, each layer having the same size as the image lattice S. We assign to each pixel $s \in S$ a unique site in each layer: *e.g.* s^d is the site corresponding to pixel s on the layer S^d. We denote $s^c \in S^c$ and $s^* \in S^*$ similarly. The segmentation is obtained by assigning a label $\omega(.)$ to all sites of $\mathcal{G}$ from the label-set: $L \triangleq \{\text{fg, bg}\}$. The labeling of S^d/S^c corresponds to the segmentation based on the $d(.)/c(.)$ feature, respectively; while the labels at the S^* layer present the final change mask.

In Fig. 2.10, we show some results obtained on three pairs of aerial images. For each pair, we show the ground truth change masks obtained by manual segmentation, the multilayer MRF results and a simple fusion obtained as a logical AND operation on the change masks of two monolayer segmentations based on each features. The increased precision of

the multi-layer model is clearly visible. Another application of multilayer modeling can be found in [62].



(a) First input image (b) First input image (c) First input image

(d) Second input image (e) Second input image (f) Second input image

(g) Fusion of two MRFs (h) Fusion of two MRFs (i) Fusion of two MRFs

(j) Multi-layer MRF (k) Multi-layer MRF (l) Multi-layer MRF

(m) Ground Truth (n) Ground Truth (o) Ground Truth

Figure 2.10: *Experimental results [2].*

IN THIS CHAPTER:

3.

The 'gas of cicrcles' MRF model

Object extraction remains one of the key problems of computer vision, which can be stated as finding regions in the image domain occupied by a specified object or objects. The solution often requires high-level knowledge about the shape of the objects. HOAC models integrate shape knowledge via the inclusion of explicit long-range dependencies between region boundary points. Herein, we will show how to set the parameters of the HOAC model to favor regions consisting of any number of approximately circular connected components, each component having approximately the same, specified radius. This yields the 'gas of circles' HOACs.

A subsequent reformulation of HOAC models as phase fields can be interpreted as real-valued continuum Markov random fields. Discretizing the phase field GOC model, we will develop an equivalent 'gas of circles' Markov random field model that assigns high probability to regions in the image domain consisting of an unknown number of circles of a given radius. The MRF model is constructed in a principled way, thereby creating an *equivalent* MRF. The model can be used as a prior for object extraction whenever the objects conform to the 'gas of circles' geometry, *e.g.* tree crowns in aerial images or cells in biological images.

Here we present a theoretical and experimental analysis of these models. The performance is demonstrated on various synthetic images as well as on the problem of tree crown detection in aerial images.

3.1 Introduction

Object extraction remains one of the key problems of computer vision and image processing. The problem is easily stated: find the regions in the image domain occupied by a specified object or objects. The solution of this problem often requires high-level knowledge about the shape of the objects sought in order to deal with high noise, cluttered backgrounds, or occlusions [5, 13, 80, 129, 173, 188]. As a result, most approaches to extraction have, to differing degrees and in different ways, incorporated prior knowledge about the shape of the objects sought. Early approaches were quite generic, essentially encouraging smoothness of object boundaries [24, 74, 79, 97, 127]. For example, [24] uses a Markovian smoothness prior (basically a Potts model, *i.e.* boundary length is penalized); [97] uses a line process to control the formation of region boundaries and control curvature; while classical active contour models [127] use boundary length and curvature, and region area in order to favor smooth closed curves [74, 79].

Subsequently there has been a great deal of work on the inclusion of more specific prior shape knowledge. Many of these methods rely on a kind of template matching: shape variability is modeled as deformations of a reference shape or shapes. Often, borrowing ideas from Grenander's pattern theory [103], shape variability arising from changes of pose is modeled as the action of a transformation group on the shape. This may or may not be combined with further intrinsic shape variability. This knowledge is then summarized in a statistical model which is incorporated into a variational [78, 80, 92, 176] or probabilistic [86, 123, 153, 183] framework. Although these methods are useful for many applications, the major drawback of using a reference shape (or shapes) is that handling an unknown number of instances of an object in the same image is difficult.

In this latter case, marked point processes constitute a very natural approach [83, 154, 165]. However, while allowing a unknown number of object instances, computational constraints mean that individual objects are usually modeled as belonging to very low-dimensional families, with no other shape variability allowed. This restricts the shapes that can be modeled and the geometric accuracy that can be achieved.

An alternative approach, known as 'higher-order active contours' (HOACs), was presented and developed in [13, 173]. HOAC models integrate shape knowledge without using reference shapes via the inclusion of explicit long-range dependencies between region boundary points. The lack of reference shapes means that they can be used to extract multiple instances of the same object.

Herein, we will show how to set the parameters of the HOAC model introduced in [173] to favor regions consisting of any number of approximately circular connected components, each component having approximately the same, specified radius [13]. A subsequent reformulation of HOAC models (and active contour models in general) as equivalent phase field models [120, 174] brings a number of theoretical and algorithmic advantages. One of the most important of these is that phase field models can be interpreted as real-valued continuum Markov random fields (MRFs), thereby allowing the theoretical and algorithmic toolbox

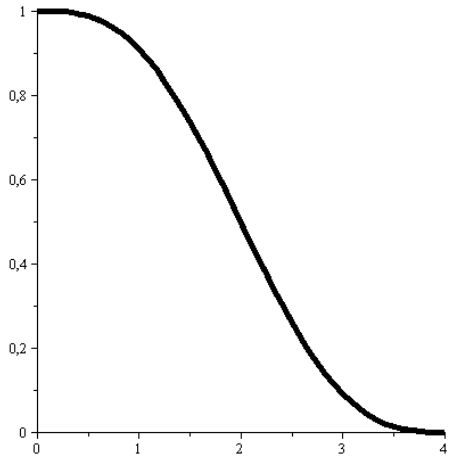


Figure 3.1: *The interaction function $\Psi_c(z)$ for $d = 2$.*

of random field theory to be brought to bear. Binarizing the phase field (in any case, regions are defined by thresholding the real-valued phase field), and of course discretize its domain, we will develop an equivalent 'gas of circles' Markov random field model [5]. Both HOAC and MRF formulation of the 'gas of circles' model has many real life applications, *e.g.* they have been succesfully applied to the extraction of tree crowns from aerial images [5, 13].

3.2 Higher order active contours

Higher-order active contours (HOACs), were presented and developed in [13, 173]. HOAC models integrate shape knowledge without using reference shapes via the inclusion of explicit long-range dependencies between region boundary points. The lack of reference shapes means that they can be used to extract multiple instances of the same object. Following [13], we show how to set the parameters of the model introduced in [173] to favor regions consisting of any number of approximately circular connected components, each component having approximately the same, specified radius. This 'gas of circles' (GOC) model was successfully used for the extraction of tree crowns from aerial images [13].

HOAC models, like all active contour models, represent a region R by its boundary, ∂R, a closed 1-chain γ in the image domain $\mathcal{D}$. The boundary ∂R is an equivalence class (under diffeomorphisms of their domain) of zero or more closed parameterized curves. The HOAC energy for the GOC model is [13]:

$$E_g(\partial R) \;=\; \lambda_c L(\gamma) + \alpha_c A(\gamma) - \frac{\beta_c}{2} \int_{T \times T'} n \cdot n' \Psi_c(|\gamma - \gamma'|) dt dt' \;, \tag{3.1}$$

where the parameterized curve γ with domain T is an arbitrary member of the equivalence class of parameterized curves corresponding to ∂R, and where L and A are the boundary length and interior area functionals. The last term of Eq. (3.1) controls the geometry of

the region, where n represents an (un-normalized) normal vector to the boundary, and where here and elsewhere, primed and unprimed maps are evaluated at primed and unprimed points of their domain. The *interaction function* Ψ_c takes the following form, plotted in Fig. 3.1:

$$\Psi_c(z) = \begin{cases} \frac{1}{2}\left(2 - \frac{|z|}{d} - \frac{1}{\pi}\sin\left(\frac{\pi(|z|-d)}{d}\right)\right) & \text{if } |z| < 2d, \\ 1 - H(|z| - d) & \text{otherwise,} \end{cases} \tag{3.2}$$

where d controls the range of interaction and H is the Heaviside step function.

3.3 The 'gas of circles' HOAC model

For certain ranges of the parameters involved, the energy in Eq. (3.1) favours regions in the form of networks, consisting of long narrow arms with approximately parallel sides, joined together at junctions, as described by [173]. It thus provides a good prior for network extraction from images. This behaviour does not persist for all parameter values, however. In [13], we showed that if the parameter triple $(\lambda_c, \alpha_c, \beta_c)$ satisfies certain constraints, circular regions of a given radius will be local minima of the energy, and thus stable, thereby yielding the HOAC 'gas of circles' (GOC) model.

For this to work, a circle of the given radius must be stable, that is, it must be a local minimum of the energy. In Section 3.3.1, we show that stable circles are indeed possible provided certain constraints are placed on the parameters. More specifically, we expand the energy E_g in a functional Taylor series to second order around a circle of radius r_0. The constraint that the circle be an energy extremum then requires that the first order term be zero, while the constraint that it be a minimum requires that the operator in the second order term be positive semi-definite. These requirements constrain the parameter values. In Section 3.3.2, we present numerical experiments using E_g that confirm the results of this analysis.

3.3.1 Stability analysis

We denote a member of the equivalence class of maps representing the 1-chain defining the circle by γ_0, and a small perturbation by $\delta\gamma$. To second order,

$$E_g(\gamma) = E_g(\gamma_0 + \delta\gamma) = E_g(\gamma_0) + \langle\delta\gamma|\frac{\delta E_g}{\delta\gamma}\rangle_{\gamma_0} + \frac{1}{2}\langle\delta\gamma|\frac{\delta^2 E_g}{\delta\gamma^2}|\delta\gamma\rangle_{\gamma_0} \,. \tag{3.3}$$

where $\langle\cdot|\cdot\rangle$ is a metric on the space of 1-chains.

Since γ_0 represents a circle, it is easiest to express it in terms of polar coordinates r, θ on $\mathcal{D}$. For a suitable choice of coordinate on S^1, a circle of radius r_0 centred on the origin is then given by $\gamma_0(t) = (r_0(t), \theta_0(t))$, where $r_0(t) = r_0$, $\theta(t) = t$, and $t \in [-\pi, \pi)$. We are interested in the behaviour of small perturbations $\delta\gamma = (\delta r, \delta\theta)$. Because the energy E_g

is defined on 1-chains, tangential changes in γ do not affect its value. We can therefore set $\delta\theta = 0$, and concentrate on δr.

On the circle, using the arc length parameterization t, the integrands of the different terms in E_g are functions of $t - t'$ only; they are invariant to translations around the circle. In consequence, the second derivative $\delta^2 E_g/\delta\gamma(t)\delta\gamma(t')$ is also translation invariant, and this implies that it can be diagonalized in the Fourier basis of the tangent space at γ_0. It is thus easiest to perform the calculation by expressing δr in terms of this basis: $\delta r(t) = \sum_k a_k e^{ir_0 kt}$, where $k \in \{m/r_0 : m \in \mathbb{Z}\}$. Below, we simply state the resulting expansions to second order in the a_k for the three terms appearing in Eq. (3.1). Details can be found in [13].

The boundary length and interior area of the region are given to second order by

$$L(\gamma) = \int_{-\pi}^{\pi} dt\, |\tau(t)| = 2\pi r_0 \left\{ 1 + \frac{a_0}{r_0} + \frac{1}{2}\sum_k k^2 |a_k|^2 \right\} \tag{3.4}$$

$$A(\gamma) = \int_{-\pi}^{\pi} d\theta \int_0^{r(\theta)} dr'\, r' = \pi r_0^2 + 2\pi r_0 a_0 + \pi \sum_k |a_k|^2 \,. \tag{3.5}$$

Note that there are no stable solutions using these terms alone. For the circle to be an extremum, we require $\lambda_c 2\pi + \alpha_c 2\pi r_0 = 0$, which tells us that $\alpha_c = -\lambda_c/r_0$. The criterion for a minimum is, for each k, $\lambda_c r_0 k^2 + \alpha_c \geq 0$. We must have $\lambda_c > 0$ for stability at high frequencies. Substituting for α_c, the condition becomes $\lambda_c(r_0 k^2 - r_0^{-1}) \geq 0$. Substituting $k = m/r_0$, gives the condition $m^2 - 1 \geq 0$: the zero frequency perturbation is never stable.

The quadratic term can be expressed to second order as

$$\int\int_{-\pi}^{\pi} dt\, dt'\, G(t,t') = 2\pi \int_{-\pi}^{\pi} dp\, F_{00}(p) + 4\pi a_0 \int_{-\pi}^{\pi} dp\, F_{10}(p) \tag{3.6}$$

$$+ \sum_k 2\pi |a_k|^2 \Big\{ \Big[2\int_{-\pi}^{\pi} dp\, F_{20}(p) + \int_{-\pi}^{\pi} dp\, e^{-ir_0 kp} F_{21}(p) \Big] \tag{3.7}$$

$$- \Big[2ir_0 k \int_{-\pi}^{\pi} dp\, e^{-ir_0 kp} F_{23}(p) \Big] + \Big[r_0^2 k^2 \int_{-\pi}^{\pi} dp\, e^{-ir_0 kp} F_{24}(p) \Big] \Big\} \tag{3.8}$$

where $G(t',t') = \tau(t') \cdot \tau(t)\ \Psi(R(t,t'))$. The F_{ij} are functionals of Ψ (and hence of d), and functions of r_0, as well as of p.

Combining Eq. (3.4), Eq. (3.5), and Eq. (3.8), we find, up to second order:

$$E_g(\gamma_0 + \delta\gamma) = e_0(r_0) + a_0 e_1(r_0) + \frac{1}{2}\sum_k |a_k|^2 e_2(k, r_0) \,, \tag{3.9}$$

where

$$\begin{aligned} e_0(r_0) &= 2\pi\lambda_c r_0 + \pi\alpha_c r_0^2 - \pi\beta_c G_{00}(r_0) \\ e_1(r_0) &= 2\pi\lambda_c + 2\pi\alpha_c r_0 - 2\pi\beta_c G_{10}(r_0) \\ e_2(k, r_0) &= 2\pi\lambda_c r_0 k^2 + 2\pi\alpha_c \\ &\quad -2\pi\beta_c \left[2G_{20}(r_0) + G_{21}(k, r_0) - 2ir_0 k G_{23}(k, r_0) + r_0^2 k^2 G_{24}(k, r_0) \right] \,, \end{aligned}$$

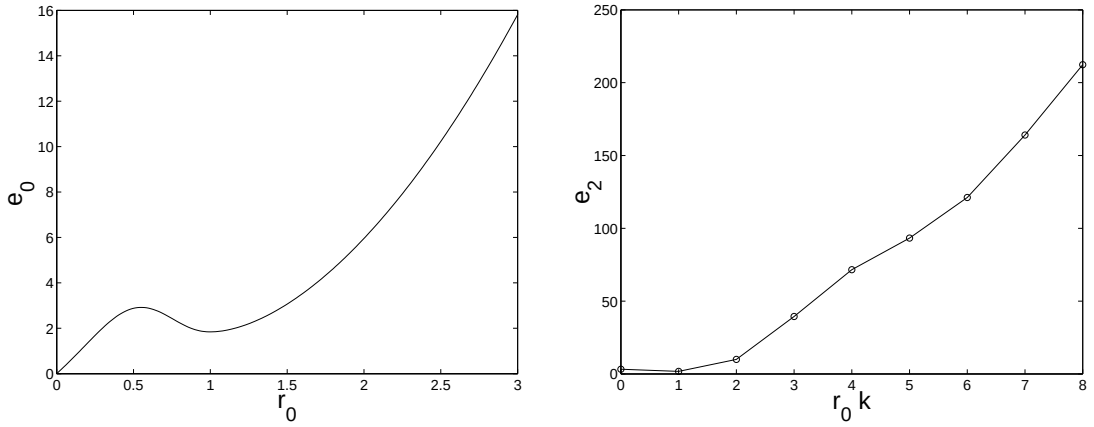


Figure 3.2: *Plots of e_0 against r_0 and e_2 against $\hat{r}_0 k$.* **Left***: the energy of a circle e_0 plotted against radius r_0 for $\lambda_c = 1.0$, $\alpha = 0.8$, and $\beta_c = 1.39$ calculated from Eq. (3.10) with $\hat{r}_0 = 1.0$. (The parameters of Ψ are $d = 1.0$ and $\epsilon = 1.0$, but note that it is not necessary in general that $d = \hat{r}_0$.) The function has a minimum at $r_0 = \hat{r}_0$ as desired.* **Right***: the second derivative of E_g, e_2, plotted against $\hat{r}_0 k$ for the same parameter values. The function is non-negative for all frequencies [13].*

where $G_{ij} = \int_{-\pi}^{\pi} dp\, e^{-ir_0(1-\delta(j))kp} F_{ij}(p)$. Note that there are no off-diagonal terms linking a_k and $a_{k'}$ for $k \neq k'$: the Fourier basis diagonalizes the second order term.

3.3.1.1 Parameter constraints

Note that a circle of any radius is always an extremum for non-zero frequency perturbations (a_k for $k \neq 0$), as these Fourier coefficients do not appear in the first order term (this is also a consequence of invariance to translations around the circle). The condition that a circle be an extremum for a_0 as well ($e_1 = 0$) gives rise to a relation between the parameters:

$$\beta_c(\lambda_c, \alpha_c, \hat{r}_0) = \frac{\lambda_c + \alpha_c \hat{r}_0}{G_{10}(\hat{r}_0)} \,, \tag{3.10}$$

where we have introduced $\hat{r}_0$ to indicate the radius at which there is an extremum, to distinguish it from r_0, the radius of the circle about which we are calculating the expansion Eq. (3.3). The left hand side of Fig. 3.2 shows a typical plot of the energy e_0 of a circle versus its radius r_0, with the β_c parameter fixed using the Eq. (3.10) with $\lambda_c = 1.0$, $\alpha = 0.8$, and $\hat{r}_0 = 1.0$. The energy has a minimum at $r_0 = \hat{r}_0$ as desired. The relationship between $\hat{r}_0$ and β_c is not quite as straightforward as it might seem though. As can be seen, the energy also has a maximum at some radius. It is not *a priori* clear whether it will be the maximum or the minimum that appears at $\hat{r}_0$. If we graph the positions of the extrema of the energy of a circle against β_c for fixed α_c, we find a curve qualitatively similar to that shown in Fig. 3.3 (this is an example of a fold catastrophe). The solid curve represents the minimum, the dashed the maximum. Note that there is indeed a unique β_c for a given choice of $\hat{r}_0$. Denote the point

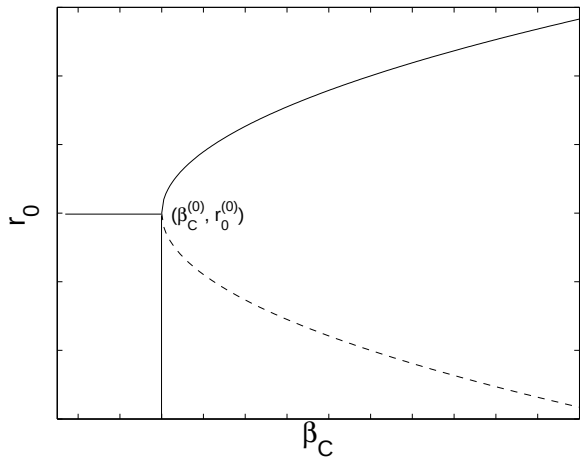


Figure 3.3: *Schematic plot of the positions of the extrema of the energy of a circle versus β_c [13].*

at the bottom of the curve by $(\beta_c^{(0)}, \hat{r}_0^{(0)})$. Note that at $\beta_c = \beta_c^{(0)}$, the extrema merge and for $\beta_c < \beta_c^{(0)}$, there are no extrema: the energy curve is monotonic because the quadratic term is not strong enough to overcome the shrinking effect of the length and area terms. In order to use Eq. (3.10) then, we have to ensure that we are on the upper branch of Fig. 3.3.

Eq. (3.10) gives the value of β_c that provides an extremum of e_0 with respect to changes of radius a_0 at a given $\hat{r}_0$ ($e_1(\hat{r}_0) = 0$), but we still need to check that the circle of radius $\hat{r}_0$ is indeed stable to perturbations with non-zero frequency, *i.e.* that $e_2(k, \hat{r}_0)$ is non-negative for all k. Scaling arguments mean that in fact the sign of e_2 depends only on the combinations $\tilde{r}_0 = r_0/d$ and $\tilde{\alpha}_C = (d/\lambda_c)\alpha_c$. The equation for e_2 can then be used to obtain bounds on $\tilde{\alpha}_C$ in terms of $\tilde{r}_0$. (Details of these calculations and bounds can be found in [13].) The right hand side of Fig. 3.2 shows a plot of $e_2(k, \hat{r}_0)$ against $\hat{r}_0 k$ for the same parameter values used for the left hand side, showing that it is non-negative for all $\hat{r}_0 k$.

We call the resulting model, the energy E_g with parameters chosen according to the above criteria, the ‘gas of circles’ model.

3.3.2 Geometric experiments

To illustrate the behaviour of ‘gas of circles’ model, in this section we show the results of some experiments using E_g (there are no image terms). Fig. 3.4 shows the result of gradient descent using E_g starting from various different initial regions. (For details of the implementation of gradient descent for higher-order active contour energies using level set methods, see [173].) In the first column, four different initial regions are shown. The other three columns show the final regions, at convergence, for three different sets of parameters. In particular, the three columns have $\hat{r}_0 = 15.0$, 10.0, and 5.0 respectively.

In the first row, the initial shape is a circle of radius 32 pixels. The stable states, which

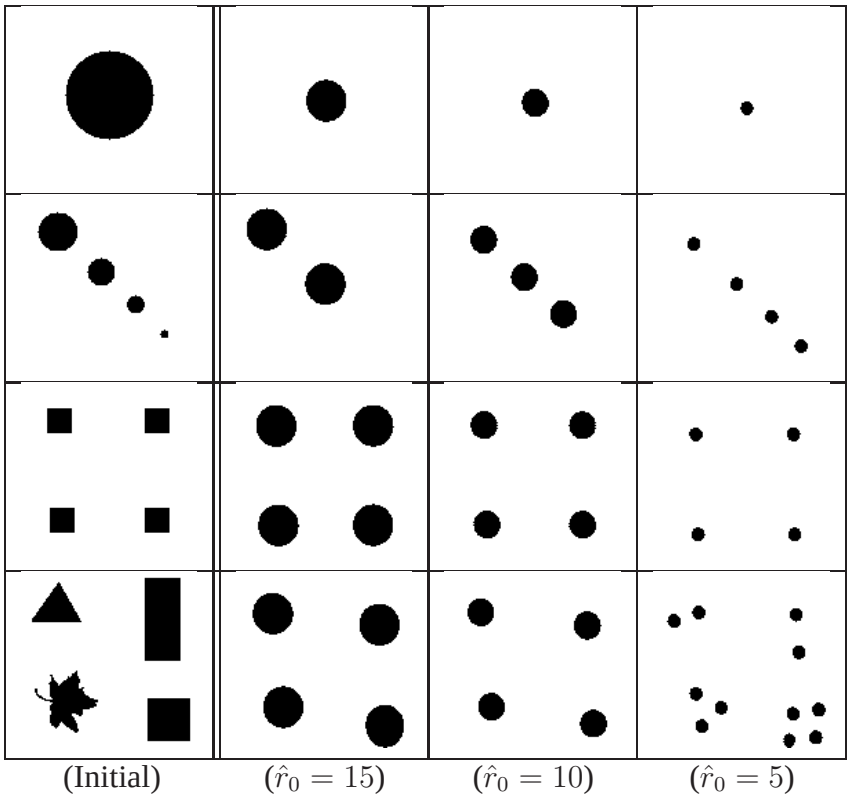


Figure 3.4: *Experimental results using the geometric term: the first column shows the initial conditions; the other columns show the stable states for various choices of the radius [13].*

can be seen in the other three columns, are circles with the desired radii in every case. In the second row, the initial region is composed of four circles of different radii. Depending on the value of $\hat{r}_0$, some of these circles shrink and disappear. This behaviour can be explained by looking at Fig. 3.2. As already noted, the energy of a circle e_0 has a maximum at some radius r_{max}. If an initial circle has a radius less than r_{max}, it will 'slide down the energy slope' towards $r_0 = 0$, and disappear. If its radius is larger than r_{max}, it will finish in the minimum, with radius $\hat{r}_0$. This is precisely what is observed in this second experiment. In the third row, the initial condition is composed of four squares. The squares evolve to circles of the appropriate radii. The fourth row has an initial condition composed of four differing shapes. The nature of the stable states depends on the relation between the stable radius, $\hat{r}_0$, and the size of the initial shapes. If $\hat{r}_0$ is much smaller than an initial shape, this shape will 'decay' into several circles of radius $\hat{r}_0$.

3.4 Phase field model

A subsequent reformulation of HOAC models (and active contour models in general) as equivalent phase field models [120, 174] brings a number of theoretical and algorithmic advantages. One of the most important of these is that phase field models can be interpreted as real-valued continuum Markov random fields (MRFs), thereby allowing the theoretical and algorithmic toolbox of random field theory to be brought to bear. Regions are defined by thresholding the real-valued phase field. Herein, we will develop an MRF GOC model [5] equivalent to the phase field GOC model, by binarizing the phase field and of course discretizing its domain. We start from the phase field formulation of the contour energy of Eq. (3.1) given in [174]:

$$E(\phi) = \int_{\mathcal{D}} \left\{ \frac{D_f}{2}|\nabla\phi|^2 + \alpha_f \left(\phi - \frac{\phi^3}{3}\right) + \lambda_f \left(\frac{\phi^4}{4} - \frac{\phi^2}{2}\right) \right\} d^2x - \\ -\frac{\beta_f}{2}\int_{\mathcal{D}^2} \nabla\phi \mathbf{G}(x,x')\nabla\phi' d^2x d^2x' \,, \tag{3.11}$$

where $\mathbf{G}(x,x')$ is defined as in [174]:

$$\mathbf{G}(x,x') = \Psi(\|x-x'\|)\mathbf{I}, \tag{3.12}$$

where $\mathbf{I}$ is a 2×2 unit matrix and Ψ is as follows:

$$\Psi(z) = \begin{cases} \frac{1}{2}\left(2 - \frac{z}{d} + \frac{1}{\pi}\sin\left(\frac{\pi z}{d}\right)\right) & \text{if } z < 2d, \\ 1 - H(z-d) & \text{else.} \end{cases} \tag{3.13}$$

where d is the range of interactions and H is the Heaviside function. Note that in our case $z = \|x-x'\|$, hence $z \geq 0$ is assumed in Eq. (3.13). The derivatives of Ψ will also be useful (note that $z < 2d$ is assumed):

$$\Psi'(z) = \frac{1}{2d}\left(\cos\left(\frac{\pi z}{d}\right) - 1\right) \tag{3.14}$$

$$\Psi''(z) = -\frac{\pi}{2d^2}\sin\left(\frac{\pi z}{d}\right) \tag{3.15}$$

The higher order term in Eq. (3.11) can be integrated by parts:

$$-\frac{\beta_f}{2}\int_{\mathcal{D}^2} \nabla\phi \mathbf{G}(x,x')\nabla\phi' d^2x d^2x' = -\frac{\beta_f}{2}\int_{\mathcal{D}^2} \phi(\nabla\nabla'\Psi(\|x-x'\|)\mathbf{I})\phi' d^2x d^2x' \tag{3.16}$$

$$= -\frac{\beta_f}{2}\int_{\mathcal{D}^2} \phi(-\nabla^2\Psi(\|x-x'\|)\mathbf{I})\phi' d^2x d^2x' \tag{3.17}$$

$$= \frac{\beta_f}{2}\int_{\mathcal{D}^2} \underbrace{\nabla^2\Psi}_{\mathcal{G}}(\|x-x'\|)\phi\phi' d^2x d^2x' \tag{3.18}$$

Using the followings:

$$\nabla\|x-x'\| = \frac{x-x'}{\|x-x'\|} \tag{3.19}$$

$$\nabla^2\|x-x'\| = \frac{1}{\|x-x'\|}\,, \tag{3.20}$$

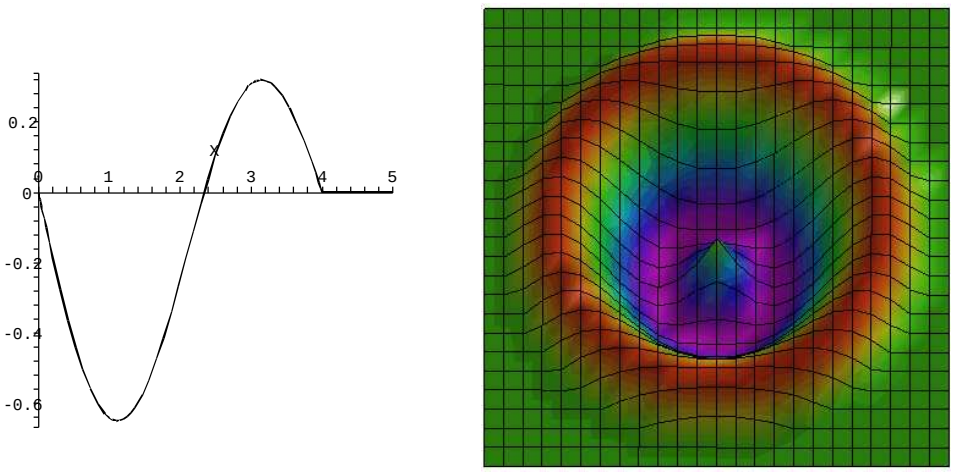


Figure 3.5: *Plots of the higher order interaction function $\mathcal{G}(\|x-x'\|)$ for $d=2$ (i.e. $\|x-x'\|<4$).* **Left***: Plot of $\mathcal{G}(z)$.* **Right***: Surface plot of $\mathcal{G}(\|x-x'\|)$.*

let us expand $\mathcal{G}(\|x-x'\|)=\nabla^2\Psi(\|x-x'\|)$:

$$\begin{aligned}
\mathcal{G}(\|x-x'\|) &= \nabla^2\Psi(\|x-x'\|)=\nabla\cdot(\Psi'(\|x-x'\|)\nabla\|x-x'\|) && (3.21)\\
&= \Psi''(\|x-x'\|)\underbrace{(\nabla\|x-x'\|)(\nabla\|x-x'\|)}_{=1}+\Psi'(\|x-x'\|)\nabla^2\|x-x'\| && (3.22)\\
&= \Psi''(\|x-x'\|)+\frac{\Psi'(\|x-x'\|)}{\|x-x'\|} && (3.23)\\
&= \frac{1}{2d\|x-x'\|}\left(\cos\left(\frac{\pi\|x-x'\|}{d}\right)-1\right)-\frac{\pi}{2d^2}\sin\left(\frac{\pi\|x-x'\|}{d}\right) && (3.24)
\end{aligned}$$

assuming $\|x-x'\|<2d$. Now we have the following energy functional to be discretized:

$$\begin{aligned}
E(\phi) &= \frac{D_f}{2}\int_{\mathcal{D}}|\nabla\phi|^2d^2x && (3.25a)\\
&\quad+\alpha_f\int_{\mathcal{D}}\left(\phi-\frac{\phi^3}{3}\right)d^2x && (3.25b)\\
&\quad+\lambda_f\int_{\mathcal{D}}\left(\frac{\phi^4}{4}-\frac{\phi^2}{2}\right)d^2x && (3.25c)\\
&\quad+\frac{\beta_f}{2}\int_{\mathcal{D}^2}\mathcal{G}(\|x-x'\|)\phi\phi'd^2xd^2x' && (3.25d)
\end{aligned}$$

$\phi:\mathcal{D}\to[-1,1]$ is a real valued function on $\mathcal{D}\subset\mathbb{R}^2$. Basically, ϕ takes the value $+1$ inside and -1 outside a region R while it changes linearly across the boundary ∂R. It is therefore a smoothed version of the characteristic function of R up to a scaling and shift. It can also be regarded as a fuzzy membership function.

3.5 Equivalence of the HOAC, phase field, and MRF models

HOAC and phase field models represent binary partitionings of the domain $\mathcal{D}$. If we denote the space of binary partitions of the domain $\mathcal{D}$ by $\mathcal{B}$, then the following mappings will relate these representations to a partitioning:

$$g: \quad \Gamma \to \mathcal{B}, \gamma \mapsto B : g(\gamma) = \mathbb{1}_B \tag{3.26}$$

$$f: \quad \Phi \to \mathcal{B}, \phi \mapsto B : f(\phi) = \mathbb{1}_B \tag{3.27}$$

where Γ and Φ denote the space of HOACs and phase fields respectively. The mappings f and g are *many to one*, because we can reach the same characteristic function via several parameterization of the phase field and contour models. In general, f may be regarded as a quantization of a given phase field function ϕ in order to obtain the characteristic function of the corresponding partition B, which is said to be represented by ϕ. Such a quantization can be achieved in several ways. $\mathbb{1}_B \equiv H(\phi)$ is the simplest one used for example in [174]. A more generic approach is to convert ϕ into a fuzzy membership function $\chi_B = (\phi+1)/2, \chi : B \to [0,1]$ and then *defuzzify* χ_B.

Both the HOAC and phase field models produce an optimal binary partitioning of the domain $\mathcal{D}$ in the sense that the resulting contour and field have the lowest energy state of the respective energy functionals:

$$\widehat{\gamma} = \arg\min_{\gamma \in \Gamma} E_g(\gamma) \tag{3.28}$$

$$\widehat{\phi} = \arg\min_{\phi \in \Phi} E(\phi) \tag{3.29}$$

Of course, the partitionings $\widehat{B}_f = f(\widehat{\phi})$ and $\widehat{B}_g = g(\widehat{\gamma})$ may not be the same.

While the above models are continuous, both the domain and state space of an MRF model are discrete. For now, let us consider a realization ω of the MRF as a *binary labeling* of the lattice $\mathcal{S}$ which is the discretization of the domain $\mathcal{D}$:

$$\omega : \mathcal{S} \to \{-1, +1\} \tag{3.30}$$

Depending on the discretization of $\mathcal{D}$, there may be several partitionings $B \in \mathcal{B}$ mapping to the same binary labeling ω:

$$m : \mathcal{B} \to \Omega, B \mapsto \omega, m(B) = 2 \cdot \mathbb{1}_B - 1 \equiv \omega \tag{3.31}$$

where Ω denote the space of binary labelings of $\mathcal{S}$. m is a *many to one* mapping as there are many continuous partitions which discretizes into the same binary labeling. Here the transfer of the characteristic function is easily defined but the discretization of the domain $\mathcal{D}$ can be achieved in several different ways.

In summary, starting from a phase field ϕ, we can reach a binary labeling ω, and this is a unique correspondence:

$$\phi \mapsto B \mapsto \omega : m(f(\phi)) = \omega \tag{3.32}$$

Note that we cannot directly relate a phase field ϕ to a labeling ω because none of the above mappings are invertible. However, we can define equivalence classes of binary partitions with respect to a given discretization of the domain $\mathcal{D}$ and equivalence classes of the phase fields with respect to a given defuzzification of the membership function χ_B.

Before constructing an equivalent MRF model, we have to clarify in what sense we want the phase field, and MRF models to be equivalent. It is clear, that each of these models is describing the same real world phenomenon. Since each model aims at minimizing some energy functional, it is expected that the optimums (*i.e.* arguments of the global minimum) be equivalent. Formally:

$$m(f(\widehat{\phi})) = \widehat{\omega}, \text{ where} \tag{3.33}$$

$$\widehat{\omega} = \arg\min_{\omega\in\Omega} U(\omega) \tag{3.34}$$

Unfortunately, it is not easy to quantify such an equivalence, partly because these optimums are defined indirectly, via the different energy functionals. Hence we will state a slightly more restricted, but easily formalized equivalence, which is also a sufficient (but not necessary) condition of the above optimum equivalence. Let us consider a map $m \circ f = W : \Phi \to \Omega$. Since we will compute MAP estimates, and since we wish to preserve the property that circles of a given radius have higher probability than neighbouring configurations, we define the MRF energy to be [5]

$$U(\omega) = \min_{\Phi:W(\phi)=\omega} E(\phi). \tag{3.35}$$

We thus set the energy $U(\omega)$ of a particular binary field ω to the energy minimum of the phase fields $\{\phi : W(\phi) = \omega\}$ belonging to the equivalence class of ω.

3.6 Discretization

Following our development in Section 3.5, we will now attempt to discretize the phase field model whose energy $E(\phi)$ is defined in Eq. (3.25). For that purpose, we have to discretize the domain $\mathcal{D}$, the function ϕ, and the energy itself.

3.6.1 Quantization of the function ϕ

The simplest way to quantize ϕ is taking $2H(\phi) - 1$. The resulting discrete function will be denoted by $\phi^{\pm} : \mathcal{D} \to \{-1, +1\}$. The energy of the discrete function $\phi^{\pm}$ can be derived

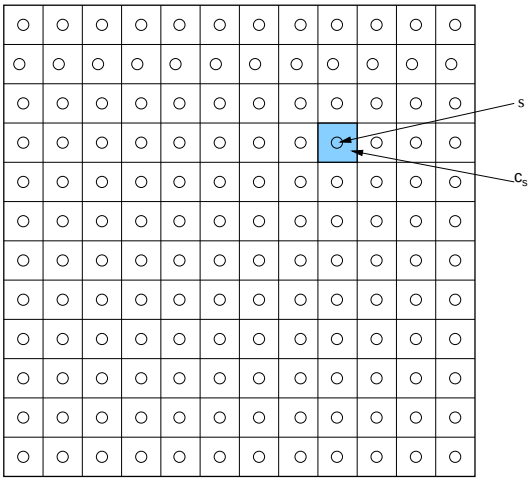


Figure 3.6: *Discretization of the domain $\mathcal{D}$. Each lattice site $s \in \mathcal{S}$ represents a unit square c_s in $\mathcal{D}$, that we call a* cell.

from Eq. (3.25) as follows:

$$
\begin{aligned}
E(\phi^{\pm}) &= E(2H(\phi)-1) && (3.36)\\
&= \int_{\mathcal{D}} \left\{ \frac{D_f}{2} |\nabla(2H(\phi)-1)|^2 + \alpha_f \left((2H(\phi)-1) - \frac{(2H(\phi)-1)^3}{3} \right) \right\} d^2x \\
&\quad + \int_{\mathcal{D}} \lambda_f \left(\frac{(2H(\phi)-1)^4}{4} - \frac{(2H(\phi)-1)^2}{2} \right) d^2x \\
&\quad + \frac{\beta_f}{2} \int_{\mathcal{D}^2} \mathcal{G}(\|x-x'\|)(2H(\phi)-1)(2H(\phi')-1) d^2x d^2x' && (3.37)\\
&= \int_{\mathcal{D}} \left\{ \frac{D_f}{2} |2\delta(\phi)\nabla\phi|^2 + \frac{2\alpha_f}{3}(2H(\phi)-1) - \frac{\lambda_f}{4} \right\} d^2x \\
&\quad + \frac{\beta_f}{2} \int_{\mathcal{D}^2} \mathcal{G}(\|x-x'\|)(2H(\phi)-1)(2H(\phi')-1) d^2x d^2x' && (3.38)
\end{aligned}
$$

Note that the discrete representation $\phi^{\pm}$ of the field ϕ corresponds to a contour representation γ and this is a one to one mapping. Furthermore, the gradient term $\delta(\phi)\nabla\phi$ contributes a non-zero value only on the region boundaries, where $\phi = 0$. Therefore this term corresponds to the contour length term $L(\gamma)$ in Eq. (3.1).

3.6.2 Discretization of the domain $\mathcal{D}$

$\mathcal{D}$ is simply discretized as a finite rectangular lattice $\mathcal{S} \subset \mathbb{Z}^2$. Each lattice site $s \in \mathcal{S}$ corresponds to (or represents) a rectangular area c_s in $\mathcal{D}$, that we call a *cell*. In our case, these cells are squares of unit size (see Fig. 3.6). The energy $U(\phi)$ over the lattice $\mathcal{S}$ can be

derived from Eq. (3.25) as follows:

$$\begin{aligned} E(\phi) \equiv U(\phi) &= \sum_{s\in\mathcal{S}} \left\{ \frac{D_f}{2} \int_{c_s} |\nabla\phi|^2 d^2x + \alpha_f \int_{c_s} \left(\phi - \frac{\phi^3}{3}\right) d^2x + \lambda_f \int_{c_s} \left(\frac{\phi^4}{4} - \frac{\phi^2}{2}\right) d^2x \right\} \\ &+ \frac{\beta_f}{2} \sum_{s\in\mathcal{S}} \sum_{r\in\mathcal{S}} \int_{c_s\times c_r} \mathcal{G}(\|x - x'\|)\phi\phi' d^2x d^2x' \end{aligned} \tag{3.39}$$

3.6.3 Discretization of the energy functional

In the following, we will derive the discrete energy functional when both the domain $\mathcal{D}$ and function ϕ are discretized. This procedure will convert the continuous phase field model into a discrete one. The resulting discrete field will be denoted by $\omega : \mathcal{S} \to \{-1, +1\}$. In our case, it is natural to assume the following discretization and quantization:

$$\omega_s := 2H\left(\int_{c_s} \phi d^2x\right) - 1 \tag{3.40}$$

Similarly, higher powers of ϕ are discretized as

$$\omega_s^n := 2H\left(\int_{c_s} \phi^n d^2x\right) - 1; \quad \forall n = 2, 3, \ldots \tag{3.41}$$

$$\equiv \begin{cases} +1 & \forall n = 2, 4, \ldots \\ \omega_s & \forall n = 3, 5, \ldots \end{cases} \tag{3.42}$$

Let us now consider the phase field energy Eq. (3.25). As usual, the gradient operator ∇ is replaced by the finite difference operator $\triangle$:

$$\nabla\phi \approx \triangle\omega \tag{3.43}$$

$$\int_{c_s} |\nabla\phi|^2 \approx \sum_{\|s-r\|=1} (\omega_s - \omega_r)^2 \tag{3.44}$$

Hence the first term Eq. (3.25a) becomes

$$\frac{D_f}{2} \int_{\mathcal{D}} |\nabla\phi|^2 d^2x = \frac{D_f}{2} \sum_s \int_{c_s} |\nabla\phi|^2 d^2x \tag{3.45}$$

$$\approx D \sum_{\|s-r\|=1} (\omega_s - \omega_r)^2, \tag{3.46}$$

where

$$D = \frac{1}{2} D_f \tag{3.47}$$

The next term Eq. (3.25b) is approximated by

$$\begin{aligned}\alpha_f \int_{\mathcal{D}} \left(\phi - \frac{\phi^3}{3}\right) d^2x &= \alpha_f \sum_s \left(\int_{c_s} \phi d^2x - \frac{1}{3}\int_{c_s} \phi^3 d^2x\right) \\ &\approx \alpha_f \sum_s \left(\omega_s - \frac{1}{3}\omega_s\right) = \alpha_f \sum_s \frac{2}{3}\omega_s \\ &\approx \alpha \sum_s \omega_s, \end{aligned} \tag{3.48}$$

yielding

$$\alpha = \frac{2}{3}\alpha_f \tag{3.49}$$

Eq. (3.25c) becomes

$$\begin{aligned}\lambda_f \int_{\mathcal{D}} \left(\frac{\phi^4}{4} - \frac{\phi^2}{2}\right) d^2x &= \lambda_f \sum_s \left(\frac{1}{4}\int_{c_s} \phi^4 d^2x - \frac{1}{2}\int_{c_s} \phi^2 d^2x\right) \\ &\approx \lambda_f \sum_s \left(\frac{1}{4} - \frac{1}{2}\right) = \lambda_f \sum_s -\frac{1}{4} \\ &\approx \lambda|\mathcal{S}|, \end{aligned} \tag{3.50}$$

from which we get

$$\lambda = -\frac{1}{4}\lambda_f \tag{3.51}$$

Finally the non-linear term Eq. (3.25d) is as follows

$$\begin{aligned}\frac{\beta_f}{2}\int_{\mathcal{D}^2} \mathcal{G}(\|x-x'\|)\phi\phi' d^2x d^2x' &= \beta_f \sum_{\|s-r\|<2d} \int_{c_s\times c_r} \mathcal{G}(\|x-x'\|)\phi\phi' d^2x d^2x' \\ &\approx \beta_f \sum_{\|s-r\|<2d} \omega_s\omega_r \int_{c_s\times c_r} \mathcal{G}(\|x-x'\|) d^2x d^2x' \\ &\approx \beta \sum_{\|s-r\|<2d} \omega_s\omega_r F_{sr}, \end{aligned} \tag{3.52}$$

where F_{sr} denotes the discrete version of the operator $\mathcal{G}(\|x-x'\|)$. Setting

$$\beta = \beta_f \tag{3.53}$$

yields

$$F_{sr} = \int_{c_s\times c_r} \mathcal{G}(\|x-x'\|) d^2x d^2x' \tag{3.54}$$

Putting these terms together we get the energy of the discrete model:

$$U(\omega) = \lambda|\mathcal{S}| + \alpha\sum_s \omega_s + D\sum_{\|s-r\|=1} (\omega_s - \omega_r)^2 + \beta \sum_{\|s-r\|<2d} \omega_s\omega_r F_{sr} \tag{3.55}$$

Note that the first term is constant hence it can be omitted when $U(\omega)$ is minimized. Now let us have a closer look at the parameters of the above energy functional.

3.6.3.1 Relationship between the parameters of the contour and field energies

The following relationship between the contour and field parameters has been derived in [174]:

$$\lambda_c^2 = \frac{16 D_f \lambda_f K_f}{15} \text{ with } K_f = 1 + 5(\alpha_f/\lambda_f)^2 \quad (3.56)$$

$$\alpha_c = \frac{4}{3}\alpha_f \quad (3.57)$$

$$\beta_c = 4\beta_f \quad (3.58)$$

The equivalence of the HOAC and phase field energy is thus established using the above parameter settings [174]. The value of K_f is derived from an approximation of the phase field energy, see [174] for details.

3.6.3.2 Parameters of the discrete energy functional

In summary, we have the following relationship between the parameters of the HOAC energy Eq. (3.1), the continuous Eq. (3.25) and discrete Eq. (3.55) field energies:

$$\alpha = \frac{2}{3}\alpha_f = \frac{1}{2}\alpha_c \quad (3.59)$$

$$D = \frac{1}{2} D_f \quad (3.60)$$

$$\lambda = -\frac{1}{4}\lambda_f \quad (3.61)$$

$$\beta = \beta_f = \frac{1}{4}\beta_c \quad (3.62)$$

Finally, from Eq. (3.56) and the above equations, we get

$$\lambda_c^2 = \frac{16 D_f \lambda_f K_f}{15} \text{ with } K_f = 1 + 5(\alpha_f/\lambda_f)^2 \quad (3.63)$$

$$= \frac{-128 D \lambda K}{15} \text{ with } K = 1 + 5(3\alpha/-8\lambda)^2 \quad (3.64)$$

3.7 Markovian interpretation

Now we will show that the discrete energy functional in Eq. (3.55) defines a Markov Random Field (MRF) with respect to an appropriate neighborhood system ν. Since the first term of the energy is a constant ($\lambda|\mathcal{S}|$), let us first remove it:

$$U(\omega) = \alpha \sum_s \omega_s + D \sum_{\|s-r\|=1} (\omega_s - \omega_r)^2 + \beta \sum_{\|s-r\|<2d} \omega_s \omega_r F_{sr} \quad (3.65)$$

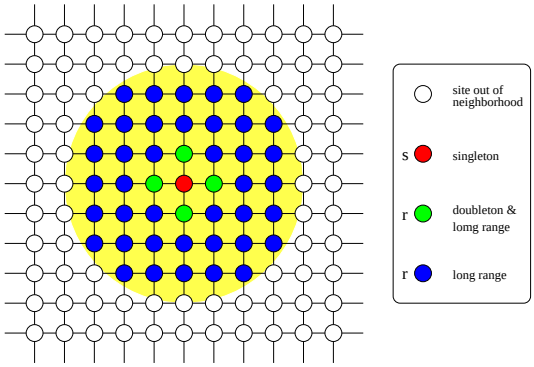


Figure 3.7: *MRF neighborhood system corresponding to the higher order interaction function* $\mathcal{G}(\|x-x'\|)$ *for* $d = 2$ *(i.e.* $\|x - x'\| < 4$*).*

In this context, $\omega = \{\omega_s : s \in \mathcal{S}\}$ is called the *label process* which is modeled as a MRF. Ω denotes the set of all possible labelings of the lattice $\mathcal{S}$. Since we have two labels (± 1), $|\Omega| = 2^{|\mathcal{S}|}$.

Definition 3.7.1 (Gibbs distribution) *A Gibbs distribution is a probability measure* P *on* Ω *with the following representation:*

$$P(\omega) = \frac{\exp\left(-U(\omega)\right)}{Z}, \tag{3.66}$$

where Z *is the* normalizing constant *or partition function:*

$$Z = \sum_{\omega \in \Omega} \exp\left(-U(\omega)\right) \tag{3.67}$$

Clearly, $U(\omega)$ from Eq. (3.65) defines a Gibbs distribution, although the computation of Z is usually not tractable. Note however, that for sampling the field, we do not need to compute the actual value of Z as long as the parameters D, α, β are fixed a priori. ω and $P(\omega)$ must also satisfy the following conditions:

Definition 3.7.2 (Markov random field) *$\mathcal{X}$ is a Markov random field (MRF) with respect to ν if*

(i) for all $\omega \in \Omega$*:* $P(\mathcal{X} = \omega) > 0$*,*

(ii) for every $s \in \mathcal{S}$ *and* $\omega \in \Omega$*:*
$P(X_s = \omega_s \mid X_r = \omega_r, r \neq s) = P(X_s = \omega_s \mid X_r = \omega_r, r \in \nu_s)$.

3.7.2/i) is satisfied by definition ($P(\omega)$ belongs to the exponential family). For 3.7.2/ii), we have to find the neighborhood system satisfying the Markovian constraint. In Eq. (3.65), there are two type of interactions (see Fig. 3.7):

1. The approximation of the gradient term by the finite differences in the second term, which corresponds to a classical nearest neighborhood.
2. The higher order pairwise interactions governed by the discrete operator F which corresponds to a neighborhood of diameter $2d$.

Since the size of the neighborhood is dominated by the latter interaction ($d \geq 2$ in practice), we conclude that the neighborhood of a site $s \in \mathcal{S}$ consists of the set of sites $\nu_s = \{r \in \mathcal{S} : \|s - r\| < 2d\}$ (see Fig. 3.7). The Gibbs - MRF equivalence is then established by the *Hammersley-Clifford* theorem [63].

Theorem 3.7.1 (Hammersley-Clifford) *$\mathcal{X}$ is a MRF with respect to the neighborhood system ν if and only if $P(\omega) = P(\mathcal{X} = \omega)$ is a Gibbs distribution with an energy equal to the sum of* clique potentials*, that is*

$$P(\omega) = \frac{\exp\left(-U(\omega)\right)}{Z} = \frac{\exp\left(-\sum_{C \in \mathcal{C}} V_C(\omega)\right)}{Z} \tag{3.68}$$

Definition 3.7.3 (Clique) *A subset $C \subseteq \mathcal{S}$ is a clique if every pair of distinct sites in C are neighbors. $\mathcal{C}$ denotes the set of cliques and $\deg(\mathcal{C}) = \max_{C \in \mathcal{C}} |C|$.*

The advantage of such a decomposition is that these potentials are a function of the local configuration of the field making it possible to define the Gibbs distribution directly in terms of local interactions. Since our neighborhood consists of three types of cliques (*singleton*, *doubleton*, and *long range* pairs), the definition of the energy function $U(\omega)$ can be completed by defining the corresponding clique potentials.

3.7.1 Singleton potential

$$\forall s : V_s = \alpha \omega_s \tag{3.69}$$

Depending on the sign of α, the singleton potential will either prefer $\omega_s = -1$ or $\omega_s = +1$ everywhere. Hence setting $\alpha > 0$ will prefer a homogeneous background ($\omega_s = -1$). It can also be interpreted as an *area term*. Thus as α is increased, typical configurations have less foreground pixels, yielding less circles. This is illustrated in Fig. 3.8), where samples from the MRF are shown for different α.

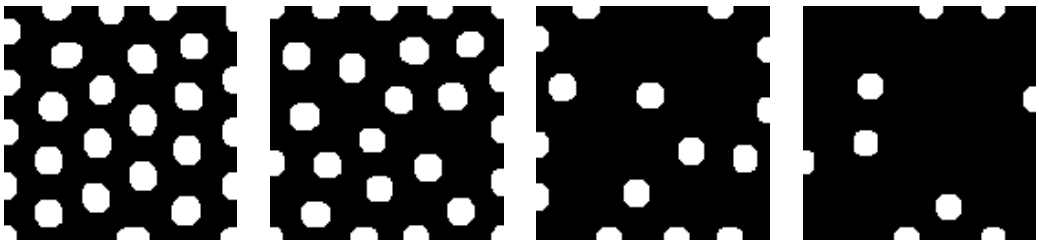
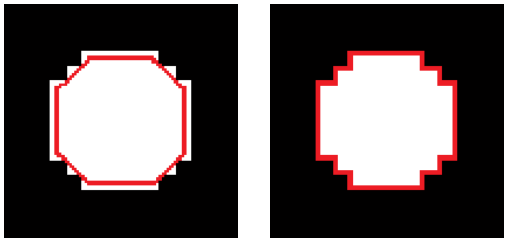


Figure 3.8: *Typical samples from the MRF defined by U: the effect of altering α ($d = 8$, $\beta = 0.096$, $D = 0.1545$) [5].*



r	Phase field	MRF	PF/MRF
2	8	12	0.66
3	16	20	0.8
4	21.5	28	0.77
5	27.3	36	0.75
10	62.6	76	0.82
15	95.5	116	0.82
20	128.5	156	0.82
50	326.4	396	0.82

Figure 3.9: *The contour length in the continuous (left) and in the discrete model (right). The table shows the continuous and discrete lengths vs. different radius [5].*

3.7.2 Doubleton potential

In the MRF model, the contour length is expressed by the doubleton term: when neighbors have different labels then there is an (implicit) contour element between them. In such cases the doubleton potential is non-zero, otherwise it vanishes. Hence the contour length in the MRF model is proportional to the overall energy of inhomogeneous doubletons:

$$\forall\{s,r\}, \|s-r\| = 1 : V_{\{s,r\}} = D(\omega_s - \omega_r)^2 = \begin{cases} 4D & \text{if } \omega_s \neq \omega_r \\ 0 & \text{otherwise} \end{cases} \tag{3.70}$$

However, due to the discretization, this contour is longer than the one in the phase field, even when the phase field is discretized (see Fig. 3.9). In order to correct for this mismatch, we have to multiply D_f by the ratio of contour lengths ((≈ 0.82) according to the table in Fig. 3.9). Furthermore, each doubleton potential is counted twice in the summation in Eq. (3.65), hence the final energy must be divided by 2. We thus get the following modified formula (see Eq. (3.47) for the original formula) to compute D from the phase field parameter D_f:

$$D = \frac{0.82}{4} D_f \tag{3.71}$$

3.7.3 Long range potential

Introduces the prior shape knowledge:

$$\forall\{s,r\}, \|s-r\| < 2d: \quad V'_{\{s,r\}} = \beta F_{sr}\omega_s\omega_r \tag{3.72}$$

$$= \begin{cases} -\beta F_{sr} & \text{if } \omega_s \neq \omega_r \\ +\beta F_{sr} & \text{otherwise} \end{cases} \tag{3.73}$$

Looking at Fig. 3.5 and Fig. 3.7, it is clear that the above potential at sites $\{s,r\}$ will favor the same label when $\|s-r\| < d'$ (*attractive case*) and a different label if $d' < \|s-r\| < 2d$ (*repulsive case*), where $d' = d + \epsilon$, because the zero point of the high order interaction function is not in d, see Fig. 3.5. Furthermore, it has no effect when $\|s-r\|$ equals to 0, or d'. Therefore this potential is only meaningful when $d \geq 2$.

3.8 The 'gas of circles' MRF model

Using the energy function of Eq. (3.65), we can now easily define the probability $P(\omega)$ of the 'gas of circles' Markov random field ω as

$$P(\omega) = \frac{1}{Z}\exp\left(-U(\omega)\right) \tag{3.74}$$

$$= \frac{1}{Z}\exp\left(-\alpha\sum_s \omega_s - D\sum_{\|s-r\|=1}(\omega_s-\omega_r)^2 - \beta\sum_{\|s-r\|<2d} F_{sr}\omega_s\omega_r\right) \tag{3.75}$$

where Z is the *partition function*.

Of course, for the extraction of circular objects from real images, we also need a *data likelihood*, $P(\mathcal{I} \mid \omega)$, which completes the definition of the *posterior* $P(\omega \mid \mathcal{I}) = P(\mathcal{I} \mid \omega)P(\omega)$. Obviously, the definition of $P(\mathcal{I} \mid \omega)$ is problem dependent. Herein, we will use a data likelihood that represents the background and foreground pixel classes by Gaussian distributions. This adds inhomogeneous terms to V_s. The result is that in the posterior probability for ω, V_s is given by

$$V_s = \alpha\omega_s + \gamma\left(\ln(\sqrt{(2\pi)}\sigma_{\omega_s}) + \frac{(\mathcal{I}_s - \mu_{\omega_s})^2}{2\sigma^2_{\omega_s}}\right). \tag{3.76}$$

The parameters of the Gaussian distributions $\mu_{\pm 1}$ and $\sigma_{\pm 1}$ are learned from representative samples provided by the user.

Using standard algorithms like simulated annealing (*e.g. Gibbs Sampler* [97] or the *Metropolis-Hastings* method [113, 152]), we can find MAP estimates $\widehat{\omega}$. In our experiments [5], we used a standard Gibbs sampler [97]. The initial temperature was set to 3 and we used an exponential annealing schedule $T_{k+1} = 0.97T_k$. The iterations were stopped

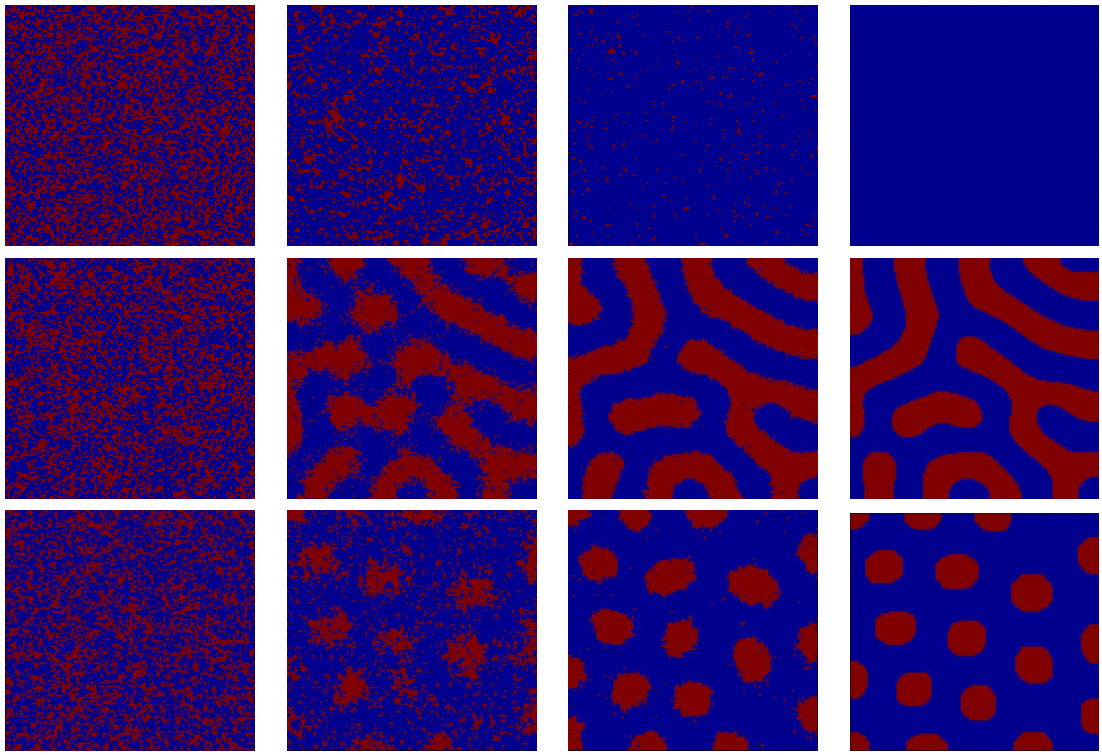


Figure 3.10: *The evolution of the MRF model* $(\alpha = 0.1863; D = 0.1545; d = 10)$*. From left to right we can see results at different temperatures. In the first row* $(\beta = 0.05)$ *the contour vanishes, in the second row* $(\beta = 0.6)$ *contour grows arms, and in the third row* $(\beta = 0.0911)$*, where* β *is computed from the GOC phase field model, the final regions are stable circles.*

when the temperature decreased below 0.01. Fig. 3.10 shows typical configurations sampled at differenet temperatures from the prior $P(\omega)$ in Eq. (3.75) with various parameter settings. These experiments confirm that the model behaves like the continuous models: when parameters are set according to the satibility analysis of the GOC model, then low energy configurations consist of stable circles.

3.8.1 Experiments

Table 3.1 shows the quantitative results obtained on a set of 160 synthetic noisy images. We compare the segmentation results to a classical MRF model [3], which doesn't include a shape prior. For a fair comparison, the false-positive (FP) and false-negative (FN) rates were computed while excluding the small circular regions. This is to avoid biasing the measure: the classical MRF should detect all regions having a particular intensity while our model will only detect the desired circles. Based on these numbers, it is clear that the proposed model is less sensitive to noise. Fig. 3.11 and Fig. 3.12 show sample results on synthetic images and demonstrates the results for various noise levels.

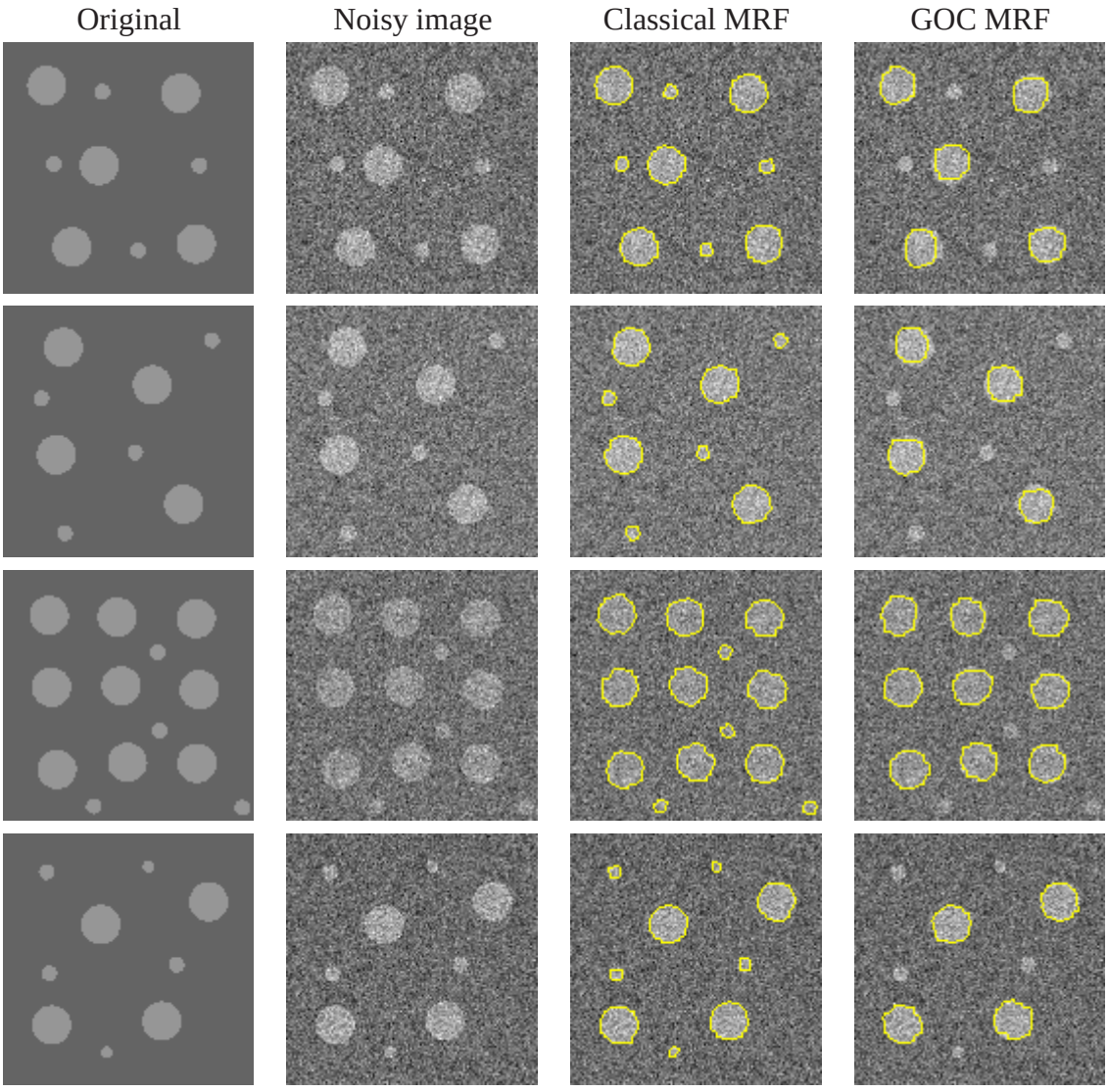


Figure 3.11: *For moderate noise levels (SNR $= -5$dB), the classical MRF model finds all circles, but -as expected- the GOC MRF model detects only circles with the appropriate radius.*

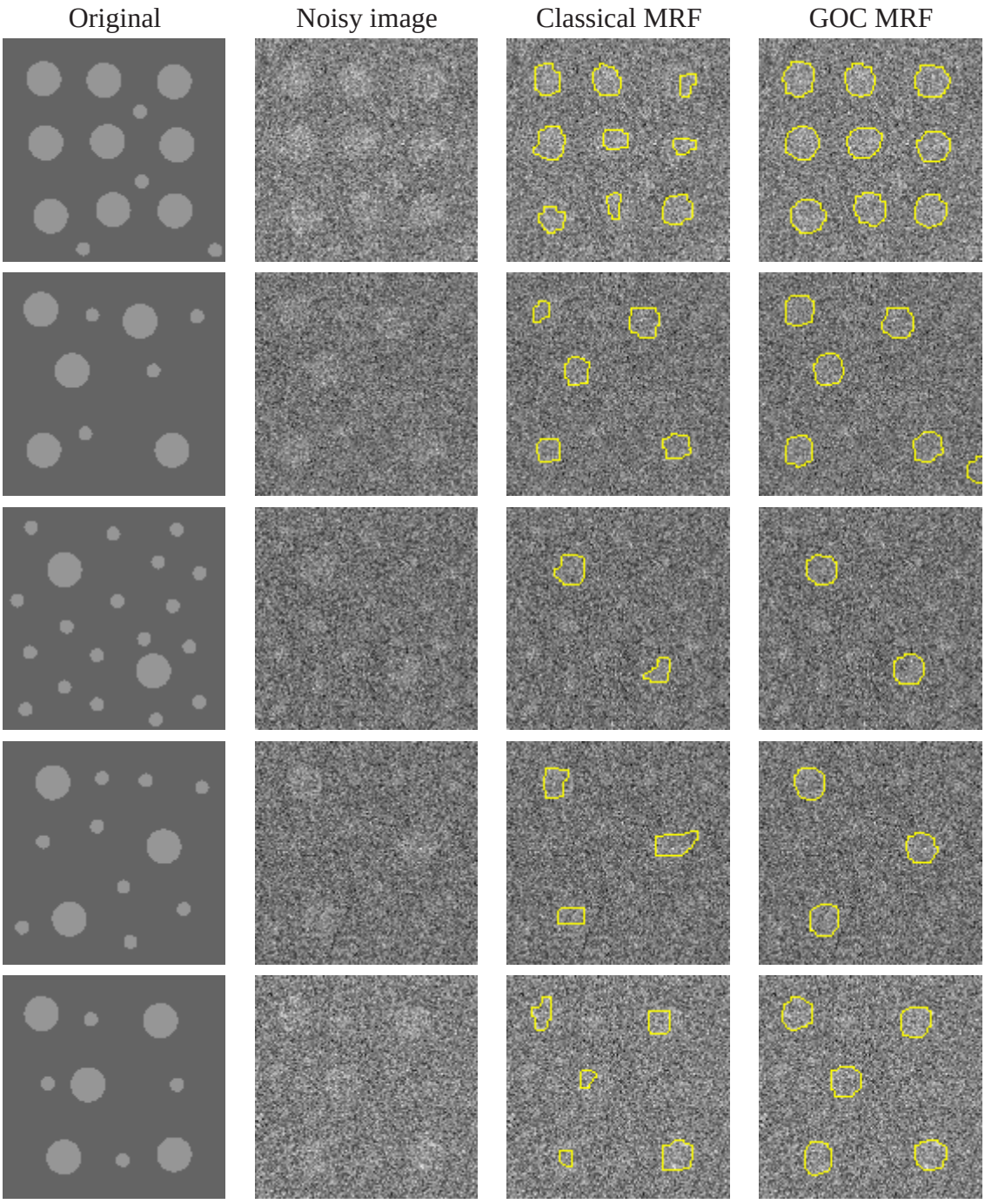


Figure 3.12: *Results on synthetic noisy images. In the first row SNR* $= -12dB$*, otherwise SNR* $= -16dB$*. The GOC MRF model segments the circles accurately while the classical MRF model is challenged by the high noise level.*

MRF		
Noise (dB)	FP	FN
0	13	46
-5	22	76
-10	34	144
-14	42	223
-16	49	431
-20	102	651

GOC MRF		
Noise (dB)	FP	FN
0	11	10
-5	21	32
-10	51	82
-14	62	120
-16	87	154
-20	147	394

Table 3.1: *Results on a set of* 160 *noisy synthetic images.* **Left:** *classical MRF;* **Right:** *GOC MRF. The slightly higher false-positive rate in the case of the GOC MRF model is probably due to the fact that a small error in the position of the detected circles results in more background pixels classified as foreground [5].*

3.9 Application in remote sensing

Forestry is a domain in which image processing and computer vision techniques can have a significant impact. Resource management and conservation require information about the current state of a forest or plantation. Much of this information can be summarized in statistics related to the size and placement of individual tree crowns (*e.g.* mean crown area and diameter, density of the trees). Currently, this information is gathered using expensive field surveys and time-consuming semi-automatic procedures, with the result that partial information from a number of chosen sites frequently has to be extrapolated. An image processing method capable of automatically extracting tree crowns from high resolution aerial or satellite images and computing statistics based on the results would greatly aid this domain.

The tree crown extraction problem can be viewed as a special case of a general image understanding problem: the identification of the region R in the image domain $\mathcal{D}$ corresponding to some entity or entities in the scene. In order to solve this problem in any particular case, we have to construct, even if only implicitly, a probability distribution on the space of regions $P(R|I,K)$. This distribution depends on the current image data I and on any prior knowledge K we may have about the region or about its relation to the image data, as encoded in the likelihood $P(I|R,K)$ and the prior $P(R|K)$ appearing in the Bayes' decomposition of $P(R|I,K)$ (or equivalently in their energies $-\ln P(I|R,K)$ and $-\ln P(R|K)$). This probability distribution can then be used to make estimates of the region we are looking for.

In the automatic solution of realistic problems, the prior knowledge K, and in particular prior knowledge about the 'shape' of the region, as described by $P(R|K)$, is critical. The tree crown extraction problem provides a good example: particularly in plantations, R takes the form of a collection of approximately circular connected components of similar size. There is thus a great deal of prior knowledge about the region sought which can be modeled by the 'gas of circles' model.

The main challenge to successful detection of crowns is the cluttered background, which

causes traditional segmentation methods to fail. Fig. 3.13, Fig. 3.15, and Fig. 3.16 show some results. In Fig. 3.13 and Fig. 3.16, the trees are difficult to separate due to shadows, blur, and vegetation between neighbouring crowns. In Fig. 3.13, results with the HOAC, phase field, and MRF models are shown. In Fig. 3.15, the classical MRF model fails to separate trees from background vegetation because they have similar intensity distributions. Obviously, the d parameter of our model, controlling the approximate radius of the detected trees, must be set correctly in order to achieve the best performance. Fig. 3.14 demonstrates the effect of various d settings.

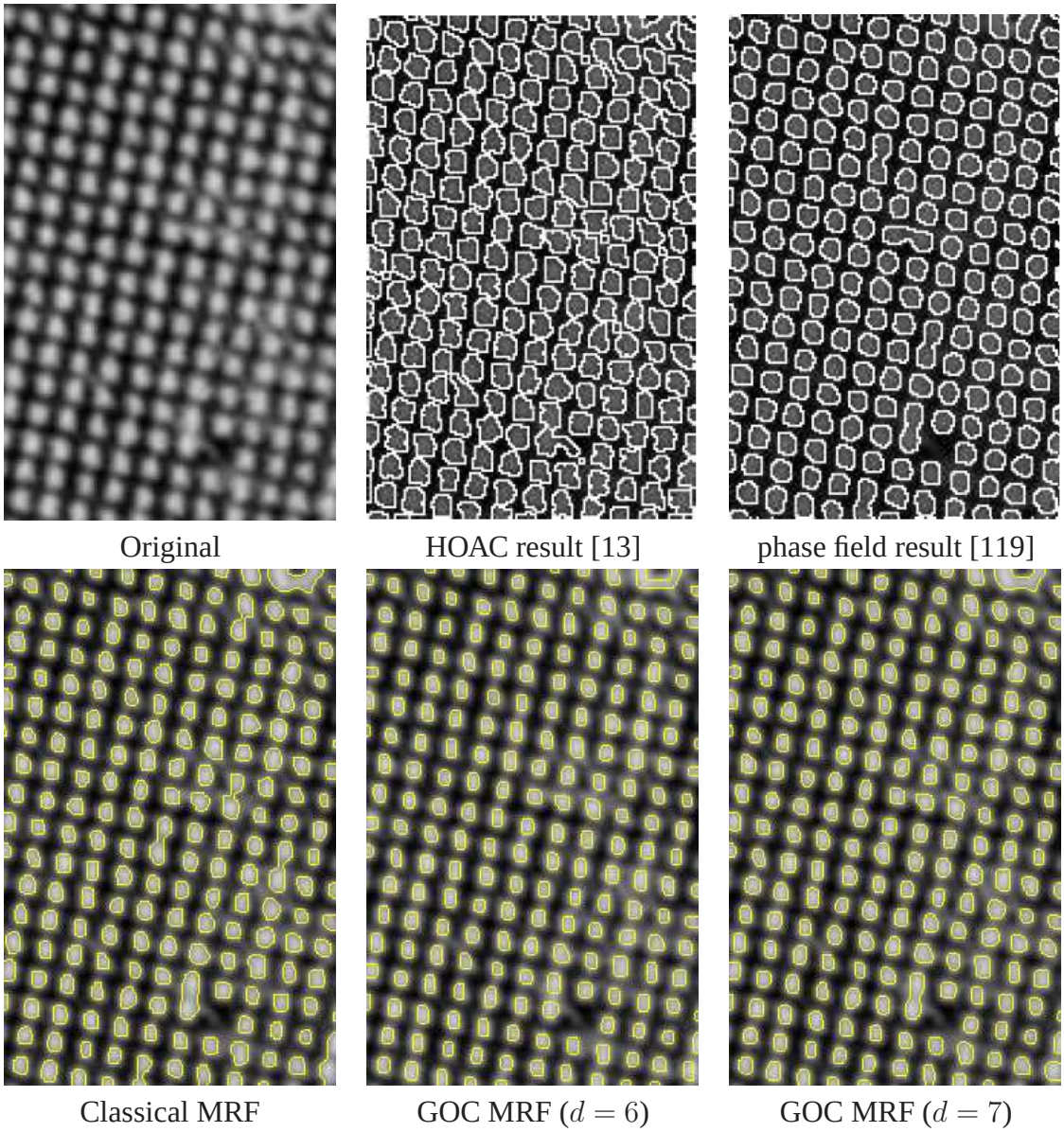
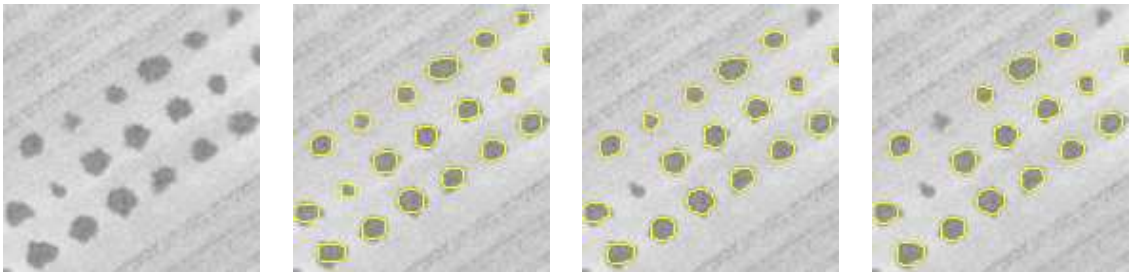
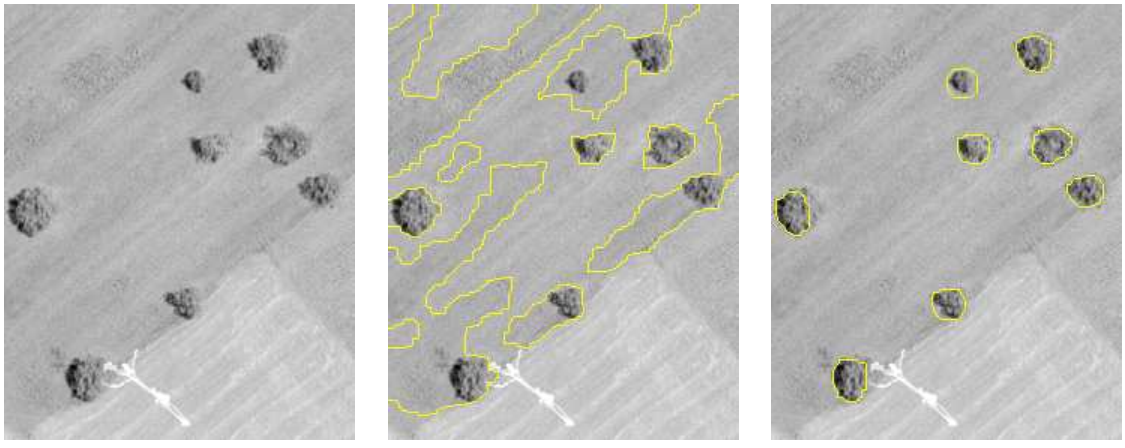


Figure 3.13: Top: *Results of the continuous models [13, 119].* **Bottom**: *Results with various MRF models [5].*



$d = 5$ $d = 6$ $d = 7$

Figure 3.14: *The effect of the d parameter. As d is increasing, smaller trees are not detected.*



Original image Classical MRF GOC MRF

Figure 3.15: *The classical MRF model fails to separate trees from background vegetation because they have similar intensity distributions [5].*

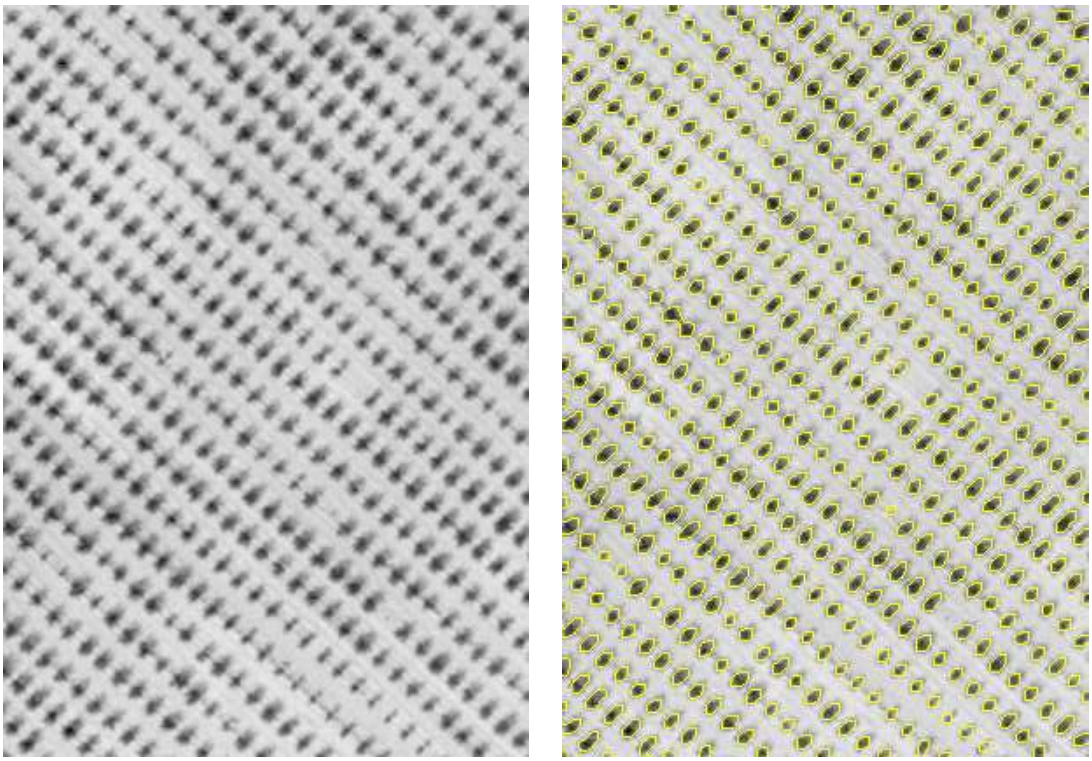


Figure 3.16: *Tree crown extraction result with the 'gas of circles' MRF model on a regularly planted pine forest [5].*

IN THIS CHAPTER:

4.

The multi-layer 'gas of circles' model

A major limitation of the 'gas of circles' model is that touching or overlapping objects cannot be represented. A generalization of the original GOC model that overcomes these limitations while maintaining computational efficiency is the multi-layer phase field GOC model. It consists of multiple instances of the phase field GOC model, each instance being known as a 'layer'. Each layer has an associated energy function, regions being defined by thresholding. Intra-layer interactions assign low energy to configurations consisting of non-overlapping near-circular regions, while overlapping regions are represented in separate layers. Inter-layer interactions penalize overlaps. This makes it possible to represent overlapping objects as subsets on different layers, thereby removing the above limitation.

The Markovian formulation yields a multi-layer binary Markov random field model that assigns high probability to object configurations in the image domain consisting of an unknown number of possibly touching or overlapping near-circular objects of approximately a given size. Each layer has an associated binary random field that specifies a region corresponding to objects. Overlapping objects are represented by regions in different layers. Within each layer, long-range clique potentials favor connected components of approximately circular shape, while regions in different layers that overlap are penalized through inter-layer cliques.

4.1 Introduction

An important subset of object extraction problems involve multiple objects of near-circular shape, *e.g.* tree crowns in remote sensing images, and cells and other structures in biological images, and are thus difficult to solve using standard shape modelling methods. To address these problems, the HOAC model has been developed favouring subsets of the image domain consisting of any number of near-circular components with approximately a given radius [13, 120]. This 'gas of circles' (GOC) model was successfully used for the extraction of tree crowns from aerial images. The model suffers, however, from two limitations that render it unsuitable for many important applications. The first arises from the representation: because the configuration space consists of subsets of the image domain, as opposed to sets of subsets, touching or overlapping objects cannot be represented. The second arises from the model: the long-range interactions that favour near-circular shapes also create repulsive interactions between nearby objects, meaning that objects in low-energy configurations are typically separated by a distance comparable to their size.

Herein, we present a generalization of the GOC model that overcomes all these limitations while maintaining computational efficiency: the multi-layer phase field GOC model [42]. This model consists of multiple instances of the phase field GOC model, each instance being known as a 'layer'. This makes it possible to represent overlapping objects as subsets on different layers, thereby removing the first limitation. The only inter-layer interaction is an overlap penalty: the long-range interaction does not act between different layers. As a result, objects in separate layers do not repel, thereby removing the second limitation. MAP estimates can be computed by minimizing the energy of the model via gradient descent, which is relatively computationally efficient if a good initialization is available.

In [45], we have developed an equivalent binary Markov random field model, the multi-layer GOC MRF model. The main difference compared to the continuous phase field model is that the MRF energy can be minimized via standard stochastic optimization, which - although computationally more expensive than gradient descent- do not require any initialization.

With a suitable data likelihood, these models can be used for object extraction in the many cases in which the 'gas of circles' geometry is relevant. Herein, we demonstrate their use for the extraction of cells and lipid droplets from biological images.

4.2 Layered representation of overlapping near-circular shapes

We now extend the single-layer model of Eq. (3.11) to a *multi-layer* GOC model. The use of multiple layers enables the representation, not just of subsets, but of sets of subsets of $\mathcal{D}$, because subsets with non-empty intersection can now be represented on separate layers. As a result, the new model can represent objects that touch and overlap in the image.

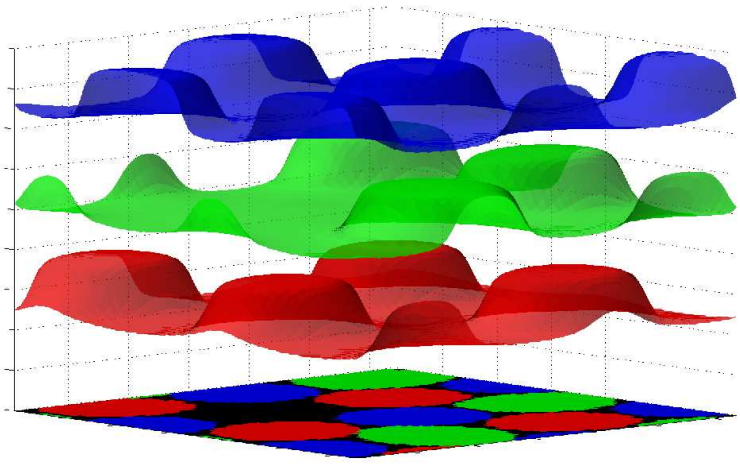


Figure 4.1: *Layered phase fields.*

This is not enough on its own because the long-range interaction in Eq. (3.11) creates a repulsion between connected components, favouring configurations in which the objects are separated by a distance comparable to their size. While appropriate for some problems, *e.g.* tree crowns in regular plantations [120], it fails for problems in which objects are touching or overlapping, see *e.g.* Fig. 4.8. To overcome this limitation, in the new model the long-range interactions act intra-layer but *not* inter-layer. This has two effects. First, the low-energy configurations in each layer are still 'gas of circles' configurations, as required. Second, the repulsive interaction is eliminated, because repulsively interacting regions can 'escape' to separate layers. The result is that overlapping 'gas of circles' configurations on separate layers can now be combined without penalty. To avoid degenerate configurations, in which a given object is duplicated across all layers, an inter-layer area overlap penalty is introduced.

To proceed, we redefine the phase field as a multi-component object: $\phi = \{\phi_i\}_{i\in[1..\ell]} : [1..\ell]\times\mathcal{D} \to \mathbb{R}$, ϕ, where ℓ is the number of layers. The total energy $\tilde{E}$ of the new multi-layer model then takes the form

$$\tilde{E}(\phi) = \sum_{i=1}^{\ell} E(\phi_i) + \frac{\kappa}{4}\sum_{i\neq j}\int_{\mathcal{D}}(1+\phi_i)(1+\phi_j)\ , \tag{4.1}$$

where E is defined in Eq. (3.11), and κ is a new parameter controlling the strength of the overlap penalty. An example of a low-energy configuration is shown in Fig. 4.1.

Note that 'background' points, with $\phi_i \simeq -1$, do not generate overlap penalty. Note also that if they do not overlap, objects in range of the repulsive interaction will tend to lie in different layers. If they do overlap, there is competition between the repulsive interaction and the overlap penalty. If κ is not too large, they will exist on separate layers; if κ is large enough, they will exist on the same layer, perhaps reducing to one object.

4.2.1 Functional derivative of the layered energy

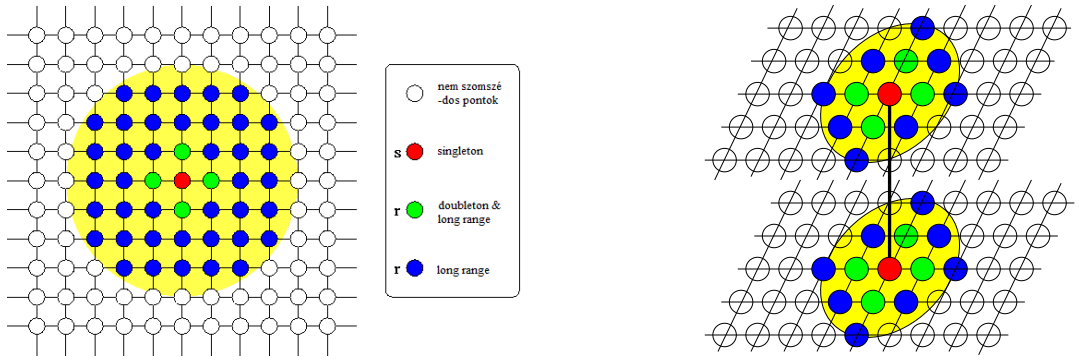
The layered phase field energy will be minimized via gradient descent, for which we need to compute its functional derivative:

$$\frac{\delta \tilde{E}(\phi)}{\delta \phi_k(x')} = \frac{\delta E(\phi_k)}{\delta \phi_k(x')} + \frac{\kappa}{4} \frac{\delta E_O(\phi)}{\delta \phi_k(x')} \tag{4.2}$$

where $E_O(\phi)$ denotes the overlap energy term from Eq. (4.1). The first term is simply the functional derivative of E evaluated at the ϕ_k, and so is known [120, 174]. The derivative of the overlap energy is

$$\begin{aligned}
\frac{\delta E_O(\phi)}{\delta \phi_k(x')} &= \sum_{i \neq j} \int_{\mathcal{D}} dx \, \{\delta_{ik}\delta(x,x')(1+\phi_j(x)) + (1+\phi_i(x))\delta_{jk}\delta(x,x')\} && (4.3)\\
&= \sum_{i \neq j} \delta_{ik}(1+\phi_j(x')) + \delta_{jk}(1+\phi_i(x')) && (4.4)\\
&= \sum_{j} (1-\delta_{kj})(1+\phi_j(x')) + \sum_{i} (1-\delta_{ki})(1+\phi_i(x')) && (4.5)\\
&= \sum_{j} ((1+\phi_j(x')) - (1+\phi_k(x'))) + \sum_{i} ((1+\phi_i(x')) - (1+\phi_k(x'))) && (4.6)\\
&= 2\sum_{i \neq k} (1+\phi_i(x')) && (4.7)\\
&= 2(\ell - 1) + 2\sum_{i \neq k} \phi_i(x') && (4.8)
\end{aligned}$$

4.3 The multi-layer MRF 'gas of circles' model



Intra-layer interactions Inter-layer interaction

Figure 4.2: *MRF neighborhoods.*

The multi-layer MRF GOC model, similar to the multi-layer phasefield model, uses multiple copies of the MRF GOC model, as follows [45]. The domain of the binary random field becomes $\tilde{\mathcal{S}} = \ell \times \mathcal{S}$, or alternatively, the field is a map from $\mathcal{S}$ to $\mathbb{B}^\ell$, where $\mathbb{B} = \{-1, +1\}$ and ℓ denotes either $\ell \in \mathbb{Z}^+$ or the set $\{1, \ldots, \ell\}$. Hence $\omega = \{\omega^{(i)}\}$ for $i \in \ell$, where $\omega^{(i)} : \mathcal{S} \to \mathbb{B}$. In principle, we would like $\ell = \mathbb{Z}^+$,*i.e.* an infinite number of layers, as this would place no restrictions on the possible configurations. In practice, there is always a maximum number of mutual overlaps, and ℓ need be no larger than this.

Sites that only differ in the value of i correspond to the same spatial point. Thus the domain $\tilde{\mathcal{S}}$ can be thought of as a series of layers, each of which is isomorphic to $\mathcal{S}$, hence the name 'multi-layer'. It is clear that the multi-layer field can represent overlapping objects, simply by placing the regions corresponding to them on different layers.

The Gibbs energy $\tilde{U}$ of the multi-layer model is the sum of the MRF GOC energies of each layer, plus an inter-layer interaction term that penalizes overlaps (see Fig. 4.2):

$$\tilde{U}(\omega) = \sum_{i=1}^{\ell} U(\omega^{(i)}) + \frac{\kappa}{4} \sum_{i \neq j} \sum_{s} (1 + \omega_s^{(i)})(1 + \omega_s^{(j)}) \;, \tag{4.9}$$

where κ is a new parameter controlling the strength of the overlap penalty.[1] Note that the inter-layer energy is ultralocal: only corresponding sites on different layers interact. Thus two regions in different layers experience no interaction at all unless they overlap. This eliminates the repulsive energy that exists in the single-layer model, because nearby but non-overlapping regions in different layers always have lower energy than the same regions in the same layer, assuming the intra-layer interactions are repulsive.

4.3.1 Energy of two interacting circles

In order to understand the behavior of the model, in this section we analyze the energy of two circles, on the same layer and on different layers. We consider the configurations shown in Fig. 4.3, where w stands for the size of the intersection: $w < 0$ means the circles do not intersect, while $w > 0$ represents a non-empty intersection of width w. We want to express the energy of these configurations as a function of w. We take advantage of the equivalence of the 'gas of circles' MRF and HOAC models to use the higher-order active contour energy Eq. (3.1) to compute the energy of the two circles. The parameters of this energy come from the equivalences between the three formulations: $\beta_c = 4\beta$; the unit weight of a boundary point is $\frac{4D}{0.82}$; while the difference in energy between an interior and exterior point is 2α. Thus the MRF energy of a single circle with radius r can be written as

$$E(r) = \frac{4D}{0.82} 2r\pi + 2\alpha r^2 \pi - 2\beta \int \int_0^{2\pi} d\theta \, d\theta' \, r^2 \cos(\theta - \theta') \, G(\gamma(\theta) - \gamma(\theta')) \;, \tag{4.10}$$

[1] Notice that $\tilde{U}$ is invariant to permutations of the layers. This will remain true even after we add a likelihood energy. Thus all configurations, and in particular minimum energy configurations, are $\ell!$ times degenerate. In practice, this degeneracy will be spontaneously broken by the optimization algorithm.

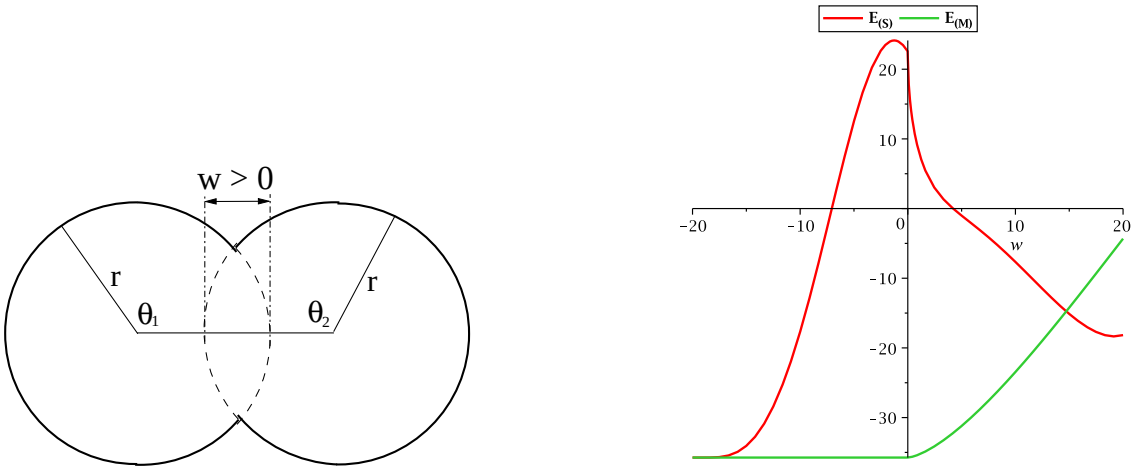


Figure 4.3: *Configurations of two overlapping circles and corresponding plots of* $E_{(M)}(r, w)$ *and* $E_{(S)}(r, w)$ *vs.* w *for two circles of radius* $r = 10$.

where γ is an embedding corresponding to the circle, parameterized, as shown in Fig. 4.3, by polar angle θ.

4.3.1.1 Different layers

When the two circles are in different layers, the only interaction energy is the inter-layer overlap penalty. Thus the energy is constant until the circles start to overlap. It then starts to increase:

$$E_{(M)}(r, w) = 2E(r) + \kappa A(r, w) \ , \tag{4.11}$$

where $A(r, w)$ is the area of the overlap given by

$$A(r, w) = \begin{cases} 2\left(r^2 \arccos\left(1 - \frac{w}{2r}\right) - \left(r - \frac{w}{2}\right)\sqrt{2rw - \frac{w^2}{4}}\right) & \text{if } w > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{4.12}$$

4.3.1.2 Same layer

When the two circles are in the same layer, they interact if $w > -2d$ for the particular form of interaction function in Eq. (3.2). (Note that we need only consider $w \leq 2r$, where r is the radius of the circles, due to symmetry.) Thus if $w \leq -2d$, the energy is simply $2E(r)$. For $w > -2d$, the energy increases with w until $w \cong 0$. As the circles start to overlap (and thus no longer form two circles, but a combined 'dumbbell' shape), there is effectively an attractive energy that causes an energy decrease with increasing w until the combined shape,

and thus the energy, becomes that of a single circle ($w = 2r$). More precisely, the energy of two circles is

$$\begin{aligned} E_{(S)}(r,w) &= \frac{4D}{0.82}2(2r\pi - L(r,w)) + 2\alpha(2r^2\pi - A(r,w)) \\ &- 4\beta \int\int_{\theta_s}^{\theta_f} d\theta_1\, d\theta_1'\, r^2 \cos(\theta_1 - \theta_1')\, G(\gamma_1(\theta_1) - \gamma_1(\theta_1')) \\ &- 2\beta \int\int_{\theta_s}^{\theta_f} d\theta_1\, d\theta_2\, r^2 \cos(\theta_1 - (\pi - \theta_2))\, G(\Delta(\theta_1, \theta_2, w))\ , \end{aligned} \quad (4.13)$$

where $\gamma_{1,2}$ are two embeddings corresponding to the two circles, parameterized by angles $\theta_{1,2}$ respectively, as shown in Fig. 4.3. We have taken advantage of symmetry to write the second line in terms of γ_1 only. $L(r,w)$ is the arc length of the intersection segment, while

$$\Delta(\theta_1, \theta_2, w) = \sqrt{(r(\sin(\theta_1) - \sin(\theta_2)))^2 + (2r - w - r(\cos(\theta_1) - \cos(\theta_2)))^2} \quad (4.14)$$

is the distance between the points $\gamma_1(\theta_1)$ and $\gamma_2(\theta_2)$. The limits $\theta_s = \cos^{-1}(\min(1, \frac{1-w}{2d}))$ and $\theta_f = 2\pi - \theta_s$ are the radial angles of the two intersection points.

The righthand side of Fig. 4.3 shows plots of $E_{(M)}(r,w)$ and $E_{(S)}(r,w)$ against w for circles with $r = 10$. When the overlap is greater than a certain threshold, controlled by κ, the energy of two circles in different layers becomes greater than two partially merged circles in one layer. Below this threshold, however, the two layer configuration has a lower energy. The stable configuration energy of two circles is given by the lower envelope of the curves in Fig. 4.3, and one can thus see that the repulsive energy that exists in the single-layer MRF GOC model is eliminated in the multi-layer MRF GOC model.

4.3.2 Experimental results

In this section, we report on the quantitative evaluation of the behavior and performance of the multi-layer MRF GOC model in object extraction problems involving simulated data and microscope images. Results were obtained as MAP estimates, using the multi-layer MRF GOC model as a prior, combined with a likelihood energy U_L to be described shortly: $\hat{\omega} = \arg\max_\omega P(I|\omega)P(\omega) = \arg\min_\omega U_L(I,\omega) + \tilde{U}(\omega)$, where $I : \mathcal{S} \to \mathbb{R}$ is the image data. Optimization was performed using Gibbs sampling coupled with simulated annealing [97]. The annealing schedule was exponential, with half-life at least 70 iterations, and a starting temperature of 3.0 for the parameter values used in the experiments.

4.3.2.1 Data likelihood

The data likelihood models the image in the interior and exterior regions using Gaussian distributions with constant means, and covariances equal to different multiples of the identity.

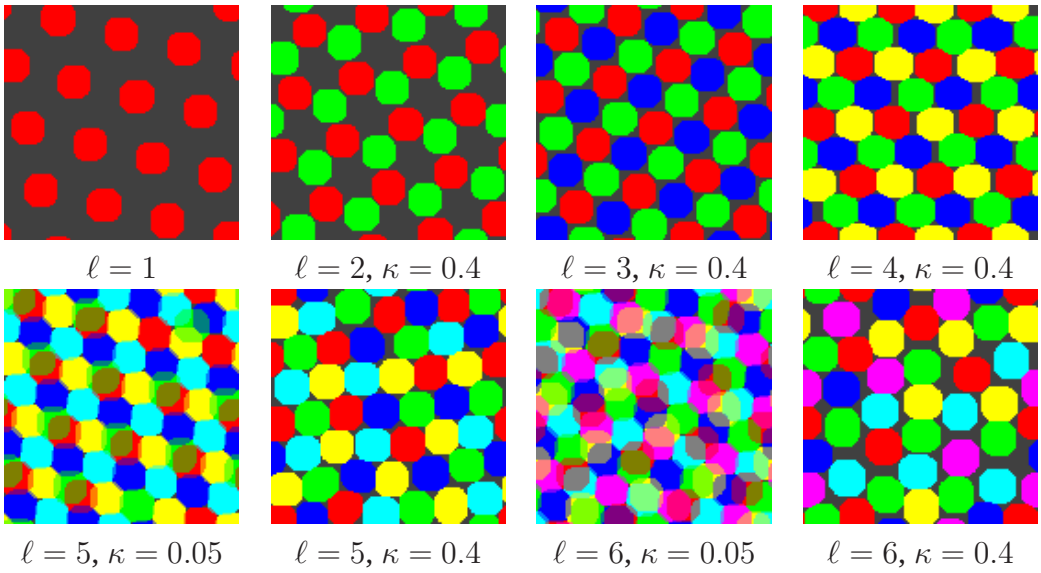


Figure 4.4: *Stable configurations of the multi-layer MRF GOC model for different numbers of layers and values of κ.*

In addition, we add an image gradient term connecting neighboring pixels, as follows. For each pair of neighboring sites, s and s', let (s, s') be the unit vector pointing from s to s'. Let $\hat{s} = \arg\max_{t \in \{s,s'\}}(|\nabla I(t)|)$. Let $h(s, s') = |(s, s') \cdot \nabla I(\hat{s})|$. Then define

$$g_i(s, s') = \begin{cases} h(s, s') & \omega_s^{(i)} = \omega_{s'}^{(i)}, \\ |\nabla I(\hat{s})| - h(s, s') & \text{otherwise.} \end{cases} \tag{4.15}$$

The likelihood energy then becomes

$$U_L(I, \omega) = \sum_i \gamma \left\{ \sum_s \left[\ln\left((2\pi)^{1/2} \sigma_{\omega_s^{(i)}} \right) + \frac{\left(I_s - \mu_{\omega_s^{(i)}} \right)^2}{2\sigma^2_{\omega_s^{(i)}}} \right] + \frac{\gamma_2}{2} \sum_s \sum_{s' \sim s} g_i(s, s') \right\}, \tag{4.16}$$

where γ and γ_2 are positive weights. In practice, the parameters $\mu_{\pm 1}$ and $\sigma_{\pm 1}$ of the Gaussian distributions were learned from representative samples.

4.3.2.2 Simulation results with the multi-layer MRF GOC model

In the first experiment, we study the global minima of $\tilde{U}$. Choosing, without loss of generality, $d = 10$, and choosing the intra-layer parameters $\alpha = 0.18634$, $D = 0.15451$, and $\beta = 0.091137$ according to the stability constraints [5, 13] and to ensure that stable circles have negative energy, $\tilde{U}$ was then minimized for different numbers of layers ℓ and values of κ. Fig. 4.4 shows representative examples of these optimal configurations. The top-left

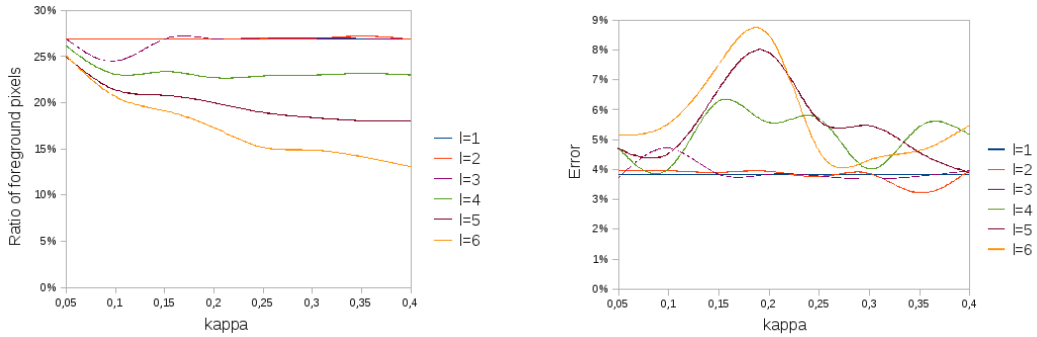


Figure 4.5: *Plots of the relative interior area (left) and shape error (right) of the stable configurations against* κ.

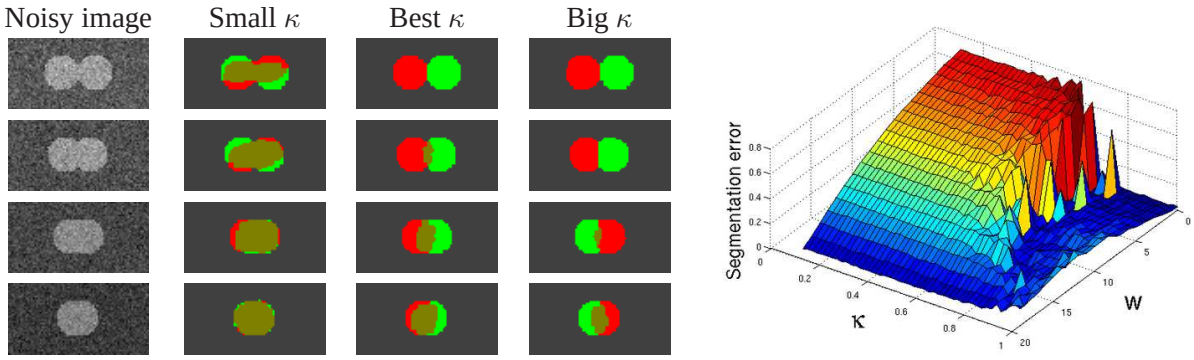


Figure 4.6: *Results on noisy synthetic images (SNR=* 0*dB) containing two circles of radius* 10 *with different degrees of overlap. Left: typical extraction results. Right: plot of segmentation error as a function of degree of overlap* (w) *and* κ.

result has $\ell = 1$: note the spacing of the circles due to the intra-layer repulsive energy. When there are more layers, the intra-layer energies favor a similarly dense ‘gas of circles’ in each layer. For $\ell \leq 3$, every layer may contain such a configuration without the circles in different layers overlapping. For $\ell > 3$, however, it is not possible to achieve both an optimal configuration in each layer and zero overlap energy. For small κ, the model tries to generate a dense configuration in each layer at the price of having overlaps. For large κ, the situation is the opposite: the model tries to avoid overlaps at the price of having less circles in each layer. Fig. 4.5 shows a plot of the relative interior area $\frac{1}{\ell N}\sum H(\omega)$ against κ, where $N = |\mathcal{S}|$. The value is almost constant for $\ell \leq 3$, while for $\ell > 3$, the value decreases with κ. The circularity of the regions was also evaluated. The righthand plot in Fig. 4.5 shows the percentage of pixels outside the ideal desired circles. Although for $\ell > 3$, these errors increase slightly, overall they remain low, meaning that the connected components remain circles to good accuracy for all ℓ and κ.

4.3.2.3 Quantitative evaluation on synthetic images

In this experiment, we demonstrate the efficiency of our model in separating overlapping circles. A series of noisy synthetic images were generated containing two circles of radius 10 with different degrees of overlap. The weights in the likelihood energy were set to $\gamma = 0.1$ and $\gamma_2 = 0$, *i.e.* no gradient term was used. We used two layers and differing κ values in the range $[0.01, 1]$. Segmentation error was evaluated as the proportion of incorrectly segmented pixels. A plot of these errors versus the amount of overlap w and κ is shown in Fig. 4.6. Note that there is a rather clear drop in the segmentation error for $\kappa \cong 0.7$. When $w > 10$ (corresponding to an overlap of greater than 50%), a larger κ is required to get an accurate segmentation ($\kappa = 0.88$ was needed in the last case in Fig. 4.6), and for $w > 15$, it is hard to get good quality results. In summary, the model performs well for reasonable overlaps and it is not sensitive to the value of κ. On the other hand, there is a performance drop for very large overlaps.

4.4 Application in biomedical imaging

Biomedical image segmentation aims to find the boundaries of various biological structures, *e.g.* cells, chromosomes, genes, proteins and other sub-cellular components in various image types [177]. Light microscope techniques are often used, but the resulting images are frequently noisy, blurred, and of low contrast, making accurate segmentation difficult. In many cases, the geometric structures involved are near-circular with many overlaps, so that our multi-layer 'gas of circles' model is well suited to extracting the desired structures. While the phase field model is computationally more efficient, it requires an appropriate initialization. The MRF model works without any particular initialization at the price of higher CPU times. Let us briefly review the results obtained on microscope data using these representations.

4.4.1 Performance of the phase field model

For the evaluation of the multi-layer phase field GOC model on real microscope images, we used the following phase field data term:

$$E_i(I, R) = \int_\Omega \Big\{ \gamma_1 \nabla I \cdot \nabla \phi \tag{4.17}$$

$$+ \gamma_2 \left[\frac{(I - \mu_{\text{in}})^2}{2\sigma_{\text{in}}^2} \phi_+ + \frac{(I - \mu_{\text{out}})^2}{2\sigma_{\text{out}}^2} \phi_- \right] \Big\} , \tag{4.18}$$

where: $\nabla\phi$, and $\phi_\pm = (1 \pm \phi)/2$ are approximately the normal vector to the boundary, and the characteristic functions of the region $(+)$ and its complement $(-)$ respectively [120]; $I : \mathcal{D} \to \mathbb{R}$ is the image data; $\mu_{\text{in,out}}$ and $\sigma_{\text{in,out}}$ are the parameters of pixel-wise Gaussian

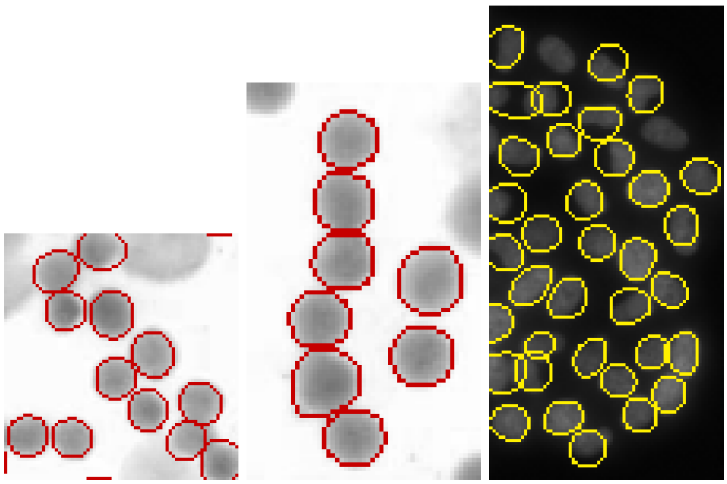


Figure 4.7: *Extraction from light microscope images of cells having a particular radius.*

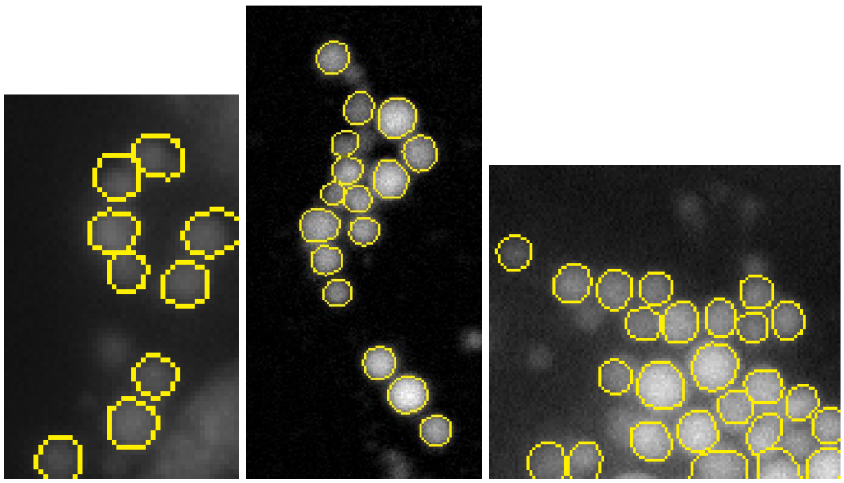


Figure 4.8: *Extraction from light microscope images of lipid drops having a particular radius.*

distributions modelling the image in the interior (in) and exterior (out) regions, learned from samples; and $\gamma_{1,2}$ are positive weights.

To initialize the multi-layer phase field, we used a simple thresholding and connected component detection, plus random assignment of different layers to nearby initial regions. Typical computation time in Matlab is about 20 seconds for a 200×100 image with 3 layers. Sample esults are shown in Fig. 4.7 and Fig. 4.8.

4.4.2 Results with the MRF model

The extraction results shown in Fig. 4.9 and Fig. 4.10 demonstrate the effectiveness of the proposed multi-layer MRF GOC model for this type of task. Computation times vary from

$\sim$ 20s to $\sim$ 1000s for images of size $N = 10^4$. The key factor is the number of layers, with the minimum time corresponding to $\ell = 2$, the maximum to $\ell = 6$.

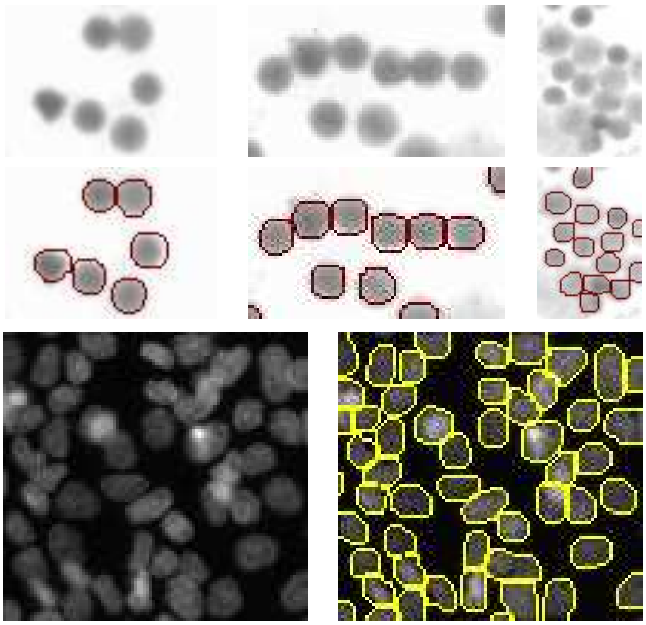


Figure 4.9: *Extraction of cells from light microscope images using the multi-layer MRF GOC model.*

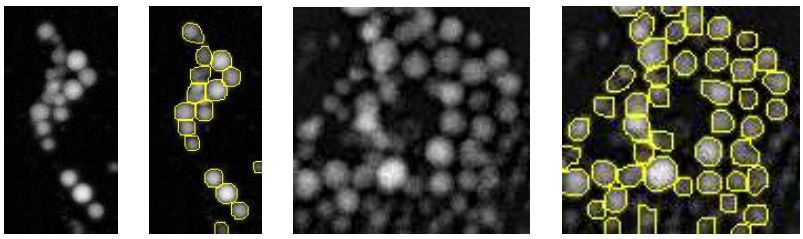


Figure 4.10: *Extraction of lipid drops from light microscope images using the multi-layer MRF GOC model.*

IN THIS CHAPTER:

5.

Linear registration of 2D and 3D objects

We consider the estimation of linear transformations aligning a known 2D or 3D object and its distorted observation. The classical way to solve this registration problem is to find correspondences between the two images and then compute the transformation parameters from these landmarks. Unlike traditional approaches, our method works without landmark extraction and feature correspondences. Here we present how to find a linear transformation as the solution of either a polynomial or a linear system of equations without establishing correspondences.

The basic idea is to set up a system of nonlinear equations whose solution directly provides the parameters of the aligning transformation. Each equation is generated by integrating a nonlinear function over the object's domains. Thus the number of equations is determined by the number of adopted nonlinear functions yielding a flexible mechanism to generate sufficiently many equations. An alternative formulation of the method leads to a linear system of equations constructed by fitting Gaussian densities to the shapes which preserve the effect of the unknown transformation.

The advantages of the proposed solutions are that they are fast, easy to implement, have linear time complexity, work without landmark correspondences and are independent of the magnitude of transformation.

5.1 Introduction

Registration is a crucial step when images of different views or sensors of an object need to be compared or combined. In a general setting, one is looking for a transformation which aligns two images such that one image (called the *observation*, or moving image) becomes similar to the second one (called the *template*, or model image). Due to the large number of possible transformations, there is a huge variability of the object signature. In fact, each *observation* is an element of the orbit of the transformations applied to the *template*. Hence the problem is inherently *ill-defined* unless this variability is taken into account.

Several techniques have been proposed to address the affine registration problem. By thresholding the magnitude of Fourier Transform of the images Zhang *et al.* [204] construct affine invariant features, which are insensitive to noise, in order to establish point correspondence. Several Fourier domain based methods [125, 133] represent images in a coordinate system in which the affine transformation is reduced to an anisotropic scaling factor, which can be computed using cross correlation methods. Govindu and Shekar [102] develop a framework that uses the statistical distribution of geometric properties of image contours to estimate the relevant transformation parameters. Main advantages of these methods is that they do not need point correspondences across views and images may also differ by the overall level of illumination. A novel one-element voxel attribute, the distance-intensity (DI) is defined in [95]. This feature encodes spatial information at a global level, and the distance of the voxel to its closest object boundary, together with the original intensity information. Then the registration is obtained by exploiting mutual information as a similarity measure on the DI feature space. For matching 2D feature points, [117] reduces the general affine case to the orthogonal case by using the means and covariance matrices of the point sets, then the rotation is computed as the roots of a low-degree complex coefficients polynomial. Another direct approach [169] extends the given pattern to a set of affine covariant versions, each carrying slightly different information, and then extract features for registration from each of them separately. The transformation is parameterized at different scales, using a decomposition of the deformation vector field over a sequence of nested (multiresolution) subspaces in [159]. An energy function describing the interactions between the images is then minimized under a set of constraints, ensuring that the transformation maintains the topology in the deformed image. Manay *et al.* [145] explore an optimization framework for computing shape distance and shape matching from integral invariants, which are employed for robustness to high-frequency noise. Shape warping by the computation of an optimal reparameterization allows this method to account for large localized changes such as occlusions and configuration changes. In [116] a method for identifying silhouettes from a given set of Radon projections is presented. The authors study how the Radon transform changes when a given 2D function is subjected to rotation, scaling, translation, and reflection. Using these properties, the parameters of the aligning transformation are expressed in terms of the Radon transform. In [107] a computationally simple solution is proposed to the affine registration of *gray level* images avoiding both the correspondence problem as well as the need for optimization. The original problem was reformulated as an *equivalent linear parame-*

ter estimation one having a *unique* and *exact* solution. However, the method relies on the availability of rich radiometric information which is clearly not available in the binary case.

Herein, we will present our generic framework for recovering linear [9, 49–51] deformations of binary objects without correspondences. The method has many applications in medical image analysis, a few examples will also be presented here.

5.2 Problem statement

Let us denote *template* and *observation* points by $\mathbf{x} = [x_1, x_2, \ldots, x_n]^T \in \mathbb{P}^n$ and $\mathbf{y} = [y_1, y_2, \ldots, y_n]^T \in \mathbb{P}^n$ respectively. The projective space $\mathbb{P}^n$ allows simple notation for affine transforms and assumes using of homogeneous coordinates. Since affine transformations never alter the third (homogeneous) coordinate of a point, which is therefore always equal to 1, we, for simplicity, and without loss of generality, liberally interchange between projective and Euclidean space, keeping the simplest notation. The identity relation between the shapes is then as follows

$$\mathbf{y} = \mathbf{A}\mathbf{x} \quad \Leftrightarrow \quad \mathbf{x} = \mathbf{A}^{-1}\mathbf{y} \,, \tag{5.1}$$

where $\mathbf{A} \in \mathbb{R}^{n\times n}$ is the unknown affine transformation that we want to recover. Classical landmark-based approaches would now identify sufficient number of point pairs $\{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^m$ then solve the system of linear equations obtained from Eq. (5.1). However, we are interested in a direct approach without solving the correspondence problem. For that purpose, shapes will be represented by their characteristic function $\mathbb{1} : \mathbb{P}^n \to \{0, 1\}$, where 0 and 1 correspond to the background and foreground respectively. If we denote the *template* by $\mathbb{1}_t$ and the *observation* by $\mathbb{1}_o$, the following equality also holds

$$\mathbb{1}_t(\mathbf{x}) = \mathbb{1}_o(\mathbf{A}\mathbf{x}) = \mathbb{1}_o(\mathbf{y}) \,. \tag{5.2}$$

When we can observe some image features (*e.g.* gray-levels of the pixels [108, 109]) that are invariant under the transformation, then we can define an additional relation

$$f(\mathbf{x}) = g(\mathbf{A}\mathbf{x}) = g(\mathbf{y}) \,, \tag{5.3}$$

where $f, g : \mathbb{P}^n \to \mathbb{R}$ are *covariant functions* under the transformation $\mathbf{A}$, defined on those observed features. Furthermore, the above relations are still valid when a function is acting on both sides of Eq. (5.1) and Eq. (5.3) [6, 7, 9, 108, 109]. Indeed, for arbitrary $\omega : \mathbb{P}^n \to \mathbb{R}$ and $\eta : \mathbb{R} \to \mathbb{R}$, we get

$$\omega(\mathbf{y}) = \omega(\mathbf{A}\mathbf{x}) \tag{5.4}$$

$$\eta \circ g(\mathbf{y}) = \eta \circ g(\mathbf{A}\mathbf{x}) = \eta \circ f(\mathbf{x}) \,. \tag{5.5}$$

Starting from either Eq. (5.4) or Eq. (5.5), we can generate as many linearly independent equations as needed by making use of nonlinear ω (resp. η) functions. There is a fundamental difference between the above two equations though: the nonlinear function ω is acting directly on the point coordinates and hence on the unknown parameters of $\mathbf{A}$ resulting in a *nonlinear* system of equations [9]; whereas η is acting on the *covariant functions* f and g allowing for a *linear* system of equations [6, 7, 108, 109].

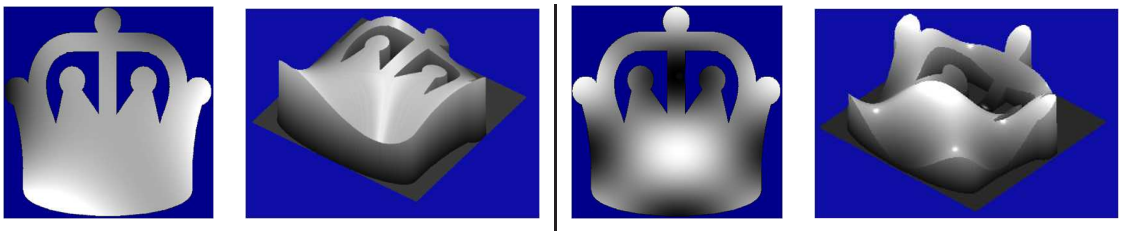


Figure 5.1: *The effect of applying a polynomial (left) and a trigonometric (right) ω function can be interpreted as a consistent colorization or as a volume.*

5.3 Solution via a nonlinear system of equations

Let us start with Eq. (5.4) and construct a system of polynomial equations [9]. For an easier presentation, we will consider 2D shapes where homogeneous coordinates of the *template* and *observation* points are denoted by $\mathbf{x} = [x_1, x_2, 1]^T \in \mathbb{P}^2$ and $\mathbf{y} = [y_1, y_2, 1]^T \in \mathbb{P}^2$ respectively. The unknown affine transformation $\mathbf{A}$ that we want to recover is thus

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{A}^{-1} = \begin{pmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ 0 & 0 & 1 \end{pmatrix}.$$

Note that $\mathbf{A}^{-1}$ exists and it is also an affine transformation since $\mathbf{A}$ is affine. We are interested in a direct solution without solving the correspondence problem. For that purpose, we will take the Lebesgue integral[1] of both sides of the identity relation in Eq. (5.1)

$$\int_{\mathbb{P}^2} \mathbf{x} d\mathbf{x} = \frac{1}{|\mathbf{A}|} \int_{\mathbb{P}^2} \mathbf{A}^{-1}\mathbf{y} d\mathbf{y}, \tag{5.6}$$

where the integral transformation $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$, $d\mathbf{x} = d\mathbf{y}/|\mathbf{A}|$ has been applied. The determinant $|\mathbf{A}|$ is the Jacobian which corresponds to the measure of the transformation. Furthermore, since shapes are represented by their characteristic functions $\mathbb{1}_t$ and $\mathbb{1}_o$ satisfying Eq. (5.2), the Jacobian can then be evaluated by integrating

$$\int_{\mathbb{P}^2} \mathbb{1}_t(\mathbf{x}) d\mathbf{x} = \frac{1}{|\mathbf{A}|} \int_{\mathbb{P}^2} \mathbb{1}_o(\mathbf{y}) d\mathbf{y} \quad \Leftrightarrow \quad |\mathbf{A}| = \frac{\int_{\mathbb{P}^2} \mathbb{1}_o(\mathbf{y}) d\mathbf{y}}{\int_{\mathbb{P}^2} \mathbb{1}_t(\mathbf{x}) d\mathbf{x}}.$$

Since the characteristic functions take only values from $\{0, 1\}$, we can further simplify the above integrals by making use of the relation:

$$\int_{\mathbb{P}^2} \mathbb{1}_t(\mathbf{x}) d\mathbf{x} \equiv \int_{\mathcal{F}_t} d\mathbf{x},$$

[1] Although we write these integrals in $\mathbb{P}^2$, they are equivalent to the corresponding Lebesgue integrals in $\mathbb{R}^2$ (*i.e.* integration is actually performed in the corresponding Cartesian coordinate system). This is because by using homogeneous coordinates, the real plane $\mathbb{R}^2$ is mapped to the $w = 1$ plane in real projective space $\mathbb{P}^2$ and affine transformations will never alter the homogeneous component w. One can therefore safely assume that it is always 1 and ignore it.

where the finite domain $\mathcal{F}_t$ consists of the *template* foreground regions: $\mathcal{F}_t = \{\mathbf{x} \in \mathbb{P}^2 | \mathbb{1}_t(\mathbf{x}) = 1\}$. Similarly, we can restrict the integral of $\mathbb{1}_o(\mathbf{y})$ to the *observation* foreground regions $\mathcal{F}_o$. Therefore evaluating the integrals yields the *area* of the foreground regions. From this point of view, the measure of the transformation $|\mathbf{A}|$ corresponds to the ratio of the *observation* and *template* shapes' area

$$|\mathbf{A}| = \frac{\int_{\mathcal{F}_o} d\mathbf{y}}{\int_{\mathcal{F}_t} d\mathbf{x}}, \tag{5.7}$$

which can be directly computed from the input images. The sign ambiguity of the determinant is also easily eliminated: A negative Jacobian would mean that the transformation is not orientation-preserving (*i.e.* flipping of coordinates is allowed). In practice, however, physical constraints will usually prevent such a transformation hence we can assume that $|\mathbf{A}|$ is always positive.

Now multiplying Eq. (5.6) and Eq. (5.2) yields a finite integral equation:

$$\begin{aligned} \int_{\mathbb{P}^2} \mathbf{x}\mathbb{1}_t(\mathbf{x})d\mathbf{x} &= \frac{1}{|\mathbf{A}|}\int_{\mathbb{P}^2} \mathbf{A}^{-1}\mathbf{y}\mathbb{1}_o(\mathbf{y})d\mathbf{y} \quad \Leftrightarrow \\ \int_{\mathcal{F}_t} \mathbf{x}d\mathbf{x} &= \frac{1}{|\mathbf{A}|}\int_{\mathcal{F}_o} \mathbf{A}^{-1}\mathbf{y}d\mathbf{y}. \end{aligned} \tag{5.8}$$

This equation implies that the finite domains $\mathcal{F}_t$ and $\mathcal{F}_o$ are also related as $\mathcal{F}_o = \mathbf{A}\mathcal{F}_t$, *i.e.* we match the shapes as a whole instead of point correspondences. In fact, Eq. (5.8) is a linear system of two equations for $k = 1, 2$:

$$|\mathbf{A}|\int_{\mathcal{F}_t} x_k d\mathbf{x} = q_{k1}\int_{\mathcal{F}_o} y_1 d\mathbf{y} + q_{k2}\int_{\mathcal{F}_o} y_2 d\mathbf{y} + q_{k3}\int_{\mathcal{F}_o} d\mathbf{y}.$$

It is clear that both sides of the equation as well as the Jacobian can be easily computed from the input shapes. Unfortunately, two equations alone are not enough to solve for 6 unknowns. However, making use of Eq. (5.4), we can generate sufficiently many equations by making use of nonlinear ω functions:

$$\int_{\mathcal{F}_t} \omega(\mathbf{x})d\mathbf{x} = \frac{1}{|\mathbf{A}|}\int_{\mathcal{F}_o} \omega(\mathbf{A}^{-1}\mathbf{y})d\mathbf{y}. \tag{5.9}$$

Intuitively, each ω generates a consistent coloring of the shapes as shown in Fig. 5.1. From a geometric point of view, Eq. (5.8) simply matches the center of mass of the *template* and *observation* while the new equations of Eq. (5.9) match the volumes over the shapes constructed by the nonlinear functions ω (see Fig. 5.1). Since ωs are also applied to the unknowns, the resulting equations will be nonlinear. The simplest nonlinear system is a low order polynomial system thus we aim at choosing ω such that Eq. (5.9) is polynomial. The following proposition states that this is achieved when ω is a polynome.

Theorem 5.3.1 *Let $\omega : \mathbb{P}^n \to \mathbb{P}^n$ and $\mathbf{x} \in \mathbb{P}^n$ ($n \in \mathbb{N}$). If the kth coordinate of $\omega(\mathbf{x})$, denoted by $\omega^k(\mathbf{x}) = p_k$ is a real n-variate polynome, $1 \leq k \leq n$, then applying ω in Eq. (5.9) results in a polynomial system of equations up to a degree of $\deg(p_k)$.*

The proof can be found in Appendix A.2 and in [9].

It is thus clear that the class of $x^n (n \in \mathbb{N}_0)$ functions are a perfect choice for ω. Hence, we obtain the following polynomial equations for $k = 1, 2$ and $n = 1, 2, 3$:

$$|\mathbf{A}| \int x_k^n = \sum_{i=1}^{n} \binom{n}{i} \sum_{j=0}^{i} \binom{i}{j} q_{k1}^{n-i} q_{k2}^{i-j} q_{k3}^{j} \int y_1^{n-i} y_2^{i-j}. \tag{5.10}$$

The system of Eq. (5.10) contains six polynomial equations up to order three which is sufficient to solve for all unknowns. In fact we have two separate systems for $k = 1, 2$:

$$|\mathbf{A}| \int x_k = q_{k1} \int y_1 + q_{k2} \int y_2 + q_{k3} \int 1, \tag{5.11}$$

$$\begin{aligned} |\mathbf{A}| \int x_k^2 &= q_{k1}^2 \int y_1^2 + q_{k2}^2 \int y_2^2 + q_{k3}^2 \int 1 + 2q_{k1}q_{k2} \int y_1 y_2 + 2q_{k1}q_{k3} \int y_1 \\ &\quad + 2q_{k2}q_{k3} \int y_2, \end{aligned} \tag{5.12}$$

$$\begin{aligned} |\mathbf{A}| \int x_k^3 &= q_{k1}^3 \int y_1^3 + q_{k2}^3 \int y_2^3 + q_{k3}^3 \int 1 + 3q_{k1}^2 q_{k2} \int y_1^2 y_2 + 3q_{k1}^2 q_{k3} \int y_1^2 \\ &\quad + 3q_{k2}^2 q_{k3} \int y_2^2 + 3q_{k1}q_{k2}^2 \int y_1 y_2^2 + 3q_{k2}q_{k3}^2 \int y_2 + 3q_{k1}q_{k3}^2 \int y_1 \\ &\quad + 6q_{k1}q_{k2}q_{k3} \int y_1 y_2. \end{aligned} \tag{5.13}$$

The solution is obtained via a direct solver (*e.g.* the *solve* function in Matlab, but other packages are also available, like *PHCpack* [193–195]) in a single pass without any loop or optimization [9]. Although, we have to solve a polynomial system, the complexity of this step is constant and, most importantly, independent of the image size. A demo implementation is also availabel at our website `http://www.inf.u-szeged.hu/~kato/software/affbinregdemo.html`. However, we may get several possible solutions for each unknown q_{ki} due to the cubic polynomial equations. Out of these potential solutions, we can select the right one by dropping the complex roots and selecting the transformation whose determinant matches the Jacobian computed by Eq. (5.7). Note that an exact solution always exists, whenever Eq. (5.2) is satisfied. In practice, however, a solution may not exists due to discretization errors or noise on the point coordinates. We can always check for the existence of a solution by computing the resultant of the system, which is a second order polynome. On the other hand, the solution is not unique (but exists!), when the shape is affine symmetric.

When an overdetermined system is constructed [51], then a *least-squares* solution can be obtained by classical *Levenberg-Marquardt* algorithm [147].

Some registration examples can be seen in Fig. 5.8, where hip prosthesis X-ray image pairs are aligned using a 2D affine transformation. The goal is to fuse post operative follow-up scans of the hip prosthesis to check loosening of the implant. Note that correspondence-based methods are challenged by lack of corner-like landmarks and the nonlinear radiometric distortion between follow-ups. However, segmentation of the implant is straightforward,

hence binary registration is a viable option here. In spite of the inherent modeling error (the physical transformation of the implant is a 3D rigid motion followed by a projection), our method was able to find a precise alignment. This is mainly due to the fact, that images are taken in a standard position of the patient, hence affine transformation is a good approximation.

5.3.1 Registration of 3D objects

The extension of the polynomial equations to 3D objects [49–51] is relatively straightforward. To generate a polynomial system, we select polynomial functions of the form $\omega(\mathbf{x})^{(k)}_{p_1,\ldots,p_n} = x_1^{p_1}\ldots x_n^{p_n}$, denoting the k-th coordinate of the transformed point, where $p_1,\ldots,p_n \in \mathbb{N}_0$. From Eq. (5.9) these functions generate the following polynomial equations (up to order 3):

$$|\mathbf{A}| \int_{\mathcal{F}_t} 1 \; d\mathbf{x} \;=\; \int_{\mathcal{F}_o} 1 \; d\mathbf{y} \;, \tag{5.14}$$

$$|\mathbf{A}| \int_{\mathcal{F}_t} x_a \; d\mathbf{x} \;=\; \sum_{i=1}^{n+1} q_{ai} \int_{\mathcal{F}_o} y_i \; d\mathbf{y} \;, \tag{5.15}$$

$$|\mathbf{A}| \int_{\mathcal{F}_t} x_a x_b \; d\mathbf{x} \;=\; \sum_{i=1}^{n+1}\sum_{j=1}^{n+1} q_{ai}q_{bj} \int_{\mathcal{F}_o} y_i y_j \; d\mathbf{y} \;, \tag{5.16}$$

$$|\mathbf{A}| \int_{\mathcal{F}_t} x_a x_b x_c \; d\mathbf{x} \;= \sum_{i=1}^{n+1}\sum_{j=1}^{n+1}\sum_{k=1}^{n+1} q_{ai}q_{bj}q_{ck} \int_{\mathcal{F}_o} y_i y_j y_k \; d\mathbf{y} \tag{5.17}$$

where $1 \le a,b,c \le n$, $a \le b \le c$, and q_{ij} denote the unknown elements of the 3-dimensional inverse transformation $\mathbf{A}^{-1}$. The formulation of higher order polynomials is analogous.

In general, these functions introduce mixed moments in both sides of the equations, yielding one, non-separable system. In practice, that causes no problems when using an iterative least-squares solution method, but usually it is not efficient to solve such a system analytically We can achieve n separable systems of equations by further limiting the ω functions: out of parameters p_i only one of them is allowed to be non-zero. In this case the k-th system of equations can be written in a more compact form to any order p as

$$\begin{aligned} |\mathbf{A}| \int_{\mathcal{F}_t} x_k^p \; d\mathbf{x} \;=\; & \sum_{i_1=0}^{p} \binom{p}{i_1} \sum_{i_2=0}^{i_1} \binom{i_1}{i_2} \ldots \sum_{i_n=0}^{i_{n-1}} \binom{i_{n-1}}{i_n} \\ & q_{k1}^{p-i_1} \ldots q_{kn}^{i_{n-1}-i_n} q_{k,n+1}^{i_n} \int_{\mathcal{F}_o} y_1^{p-i_1} y_2^{i_1-i_2} \ldots y_n^{i_{n-1}-i_n} \; d\mathbf{y} \;. \end{aligned} \tag{5.18}$$

However, numerical implementation has to be carefully designed. Therefore, both in 2D and 3D we examined two different types of solution methods: iterative least-squares solutions and direct analytical solutions.

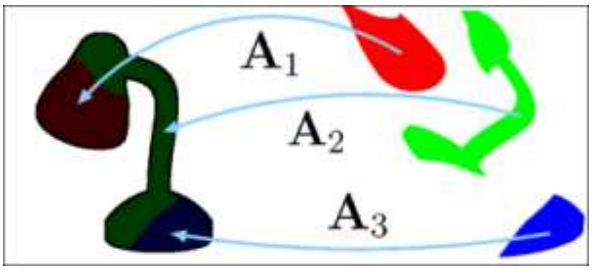


Figure 5.2: *Affine puzzle: reconstructing the complete template object from its deformed parts.*

- In case of a direct method, limited number of equations can be used (according to the degree of freedom of the n-dimensional affine transformation), while an iterative approach allows for an overdetermined system, which may give more stability.

- Direct methods may provide many hundreds or even thousands of possible solutions, many (or even all) of them may be complex thus a solution selection scheme has to be used to produce only one real solution from these. Iterative methods provide a single real solution, but the search may fall into local minima. To avoid such local minima, usually a sophisticated search strategy is necessary.

- Direct methods can provide full affine solutions only, but in case of iterative methods restrictions to lower degree of freedom transformations are easy to impose.

We found that the direct approach gives more stable results, but the iterative one is more precise. It is also possible to combine the two approaches: The direct approach provides the initialization of the iterative one. A demo implementation of our 3D shape registration method is available at `http://www.inf.u-szeged.hu/~kato/software/affbin3dregdemo.html`

Another issue is discretization error, which might be particularly problematic in 3D. For that purpose, we extended our method by investigating the case when the segmentation method is capable of producing *fuzzy object*s instead of a binary result in both 2D and 3D. It has been shown that the information preserved by using fuzzy representation based on area coverage may be successfully utilized to improve precision and accuracy of our equations [49,51]. The result of a series of synthetic tests showed that fuzzy representation yields lower registration errors in average.

Quantitative evaluation on synthetic data of the various algorithmic solutions has been presented in [9, 49, 51] In Fig. 5.9 and Fig. 5.10, some registration results on 3D medical images are shown.

5.4 Affine puzzle

The affine puzzle problem can be formulated as follows: Given a binary image of an object (the *template*) and another binary image (the *observation*) containing the fragments of the *template*, we want to establish the geometric correspondence between these images which reconstructs the complete *template* object from its parts. The overall distortion is a global nonlinear transformation with the following constraint (see Fig. 5.2):

- the object parts are distinct (*i.e.* either disconnected or separated by segmentation),
- all fragments of the *template* are available, but
- each of them is subject to a different affine deformation, and the partitioning of the *template* object is unknown.

A related problem is partial matching of shapes [69,94,170]. Partial matching addresses a particularly challenging setting of classical shape matching, where two shapes are dissimilar in general, but have significant similar parts. In this context, our problem would require to find a partial matching between the *template* and each fragments of the *observation*. Current approaches are usually based on the Laplace-Beltrami framework [170, 178], but classical approaches like the Iterative Closest Point (ICP) [65] algorithm can also be used assuming an appropriate shape representation [69]. Considering the rather high computational complexity of these algorithms, this solution is far from optimal for our problem.

Most of the existing solutions to the puzzle problem [132, 150, 163] consist in matching fragment-pairs to find neighbors, which are then reassembled by a rigid body transformation. Although classical approaches may account for a *template* object by incorporating a set of constraints to improve the overall performance, they are primarily targeted to problems where a *template* is not available, *e.g.* archaeology [150]. On the other hand, there are many applications where a *template* object is available: In industrial applications usually 3D models of manufactured parts can be easily produced. In medical imaging an *atlas* can be used or, by taking advantage of the symmetry of the human body, the intact bone can provide a *template* for bone fracture reduction, as shown in Fig. 5.11. Therefore we address this important setting of the puzzle problem and propose a generic solution which can be applied to both 2D and 3D transformations. The proposed solution [8] consists in constructing and solving a polynomial system of equations similar to Eq. (5.11)–(5.13), which provides all the unknown parameters of the alignment.

5.4.1 Realigning object parts

Given an n dimensional *template* object and an *observation* containing its affine deformed fragments, we want to recover the transformations realigning these shapes into their original position on the *template*. Let us denote the homogeneous point coordinates of the *template*

and *observation* by $\mathbf{x} = [x_1, \ldots, x_n, 1]$ and $\mathbf{y} = [y_1, \ldots, y_n, 1] \in \mathbb{P}^n$. Furthermore, let $\ell \in \mathbb{N}$ denote the number of fragments on the *observation*. The transformation aligning the *observation* with the *template* is a non-linear one, composed of ℓ linear transformations

$$\mathbf{A}_i = \begin{bmatrix} a_{i11} & a_{i12} & \ldots & a_{i1(n+1)} \\ \vdots & & \ddots & \vdots \\ a_{in1} & a_{in2} & \ldots & a_{in(n+1)} \\ 0 & 0 & \ldots & 1 \end{bmatrix} \quad i = 1, \ldots \ell \, . \tag{5.19}$$

Since the *observation* has disjoint parts, we can assume that points of each deformed shape are labeled by the function $\lambda' : \mathbb{P}^n \to \{0, 1, \ldots, \ell\}$, which assigns 0 to the background. Obviously, there is a corresponding *hidden* labeling $\lambda : \mathbb{P}^n \to \{0, 1, \ldots, \ell\}$ which assigns the label i to the *template* points corresponding to the i^{th} shape. Our goal is to recover the affine matrices $\{\mathbf{A}_i\}_{i=1}^{\ell}$. The main challenges are that neither the partitioning (*i.e.* the hidden labeling λ) of the *template* nor correspondences between the shapes are known.

Let us first establish a solution for the i^{th} shape. The *template* and *observation* domains are denoted by $\mathcal{D}_i = \{\mathbf{x} \in \mathbb{P}^n | \lambda(\mathbf{x}) = i\}$ and $\mathcal{D}'_i = \{\mathbf{y} \in \mathbb{P}^n | \lambda'(\mathbf{y}) = i\}$, respectively. Note that $\mathcal{D}'_i$ is known but $\mathcal{D}_i$ is unknown. The points of these domains are related by the unknown transformation $\mathbf{A}_i$ via the identity relation of Eq. (5.1). As before, applying a function $\omega : \mathbb{P}^n \to \mathbb{R}$ and integrating out individual point correspondences yields

$$\int_{\mathcal{D}_i} \omega(\mathbf{x}) d\mathbf{x} = |\mathbf{A}_i| \int_{\mathcal{D}'_i} \omega(\mathbf{A}_i \mathbf{y}) d\mathbf{y}, \tag{5.20}$$

Based on Eq. (5.20), we can construct as many equations as needed by making use of a set of nonlinear functions $\{\omega_j\}_{j=1}^{m}$, $m \geq n(n+1)$ [8, 9].

We have established relations between the i^{th} shape-pair, but we know neither the correspondence between the shapes nor the partitioning $\mathcal{D}_i$ of the *template*. Would these information available, a pairwise alignment could be recovered by any standard binary registration method. Unfortunately, that would require to solve a partial matching problem [69] between each *observation* shape and the *template*, which is far from trivial. Therefore we will sum equations for all shape domains $\mathcal{D}_i$ and solve the problem simultaneously, estimating all parameters in one system of equations. Thus Eq. (5.20) becomes

$$\sum_{i=1}^{\ell} \int_{\mathcal{D}_i} \omega_j(\mathbf{x}) d\mathbf{x} = \sum_{i=1}^{\ell} |\mathbf{A}_i| \int_{\mathcal{D}'_i} \omega_j(\mathbf{A}_i \mathbf{y}) d\mathbf{y} \, . \tag{5.21}$$

Let $\mathcal{D} := \cup_{i=1}^{\ell} \mathcal{D}_i$, where $\mathcal{D} = \{\mathbf{x} \in \mathbb{P}^n | \lambda(\mathbf{x}) \neq 0\}$ is the shape domain corresponding to the whole *template*. Therefore the left hand side of the above equation can be written as

$$\sum_{i=1}^{\ell} \int_{\mathcal{D}_i} \omega_j(\mathbf{x}) d\mathbf{x} = \int_{\bigcup_{i=1}^{\ell} \mathcal{D}_i} \omega_j(\mathbf{x}) d\mathbf{x} = \int_{\mathcal{D}} \omega_j(\mathbf{x}) d\mathbf{x} \, , \tag{5.22}$$

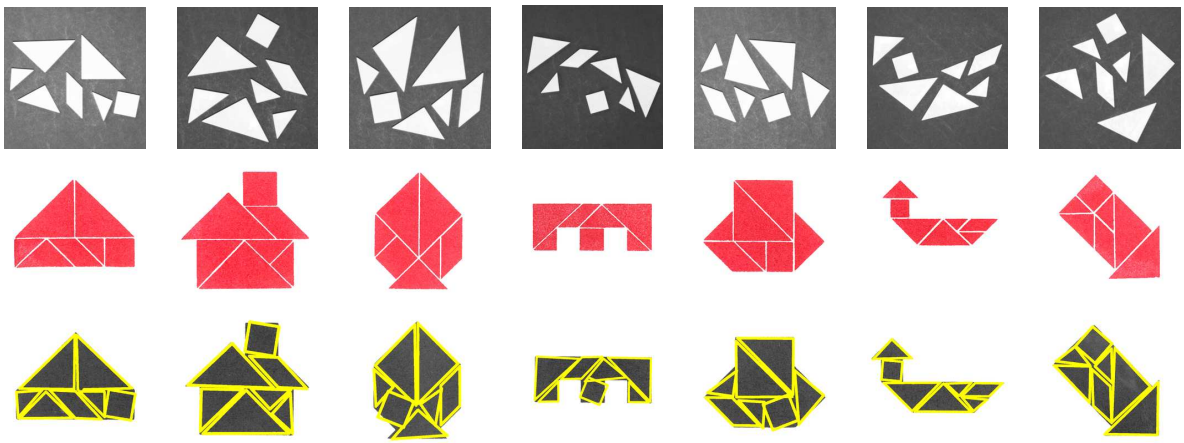


Figure 5.3: *Solutions of the Tangram puzzle (the average alignment runtime of an image was about 50 sec. in Matlab).* **Top:** Observations *are taken by digital camera.* **Middle:** *Solutions, found in the Tangram manual.* **Bottom:** *The scanned* template *silhouettes with overlaid contours of aligned fragments.*

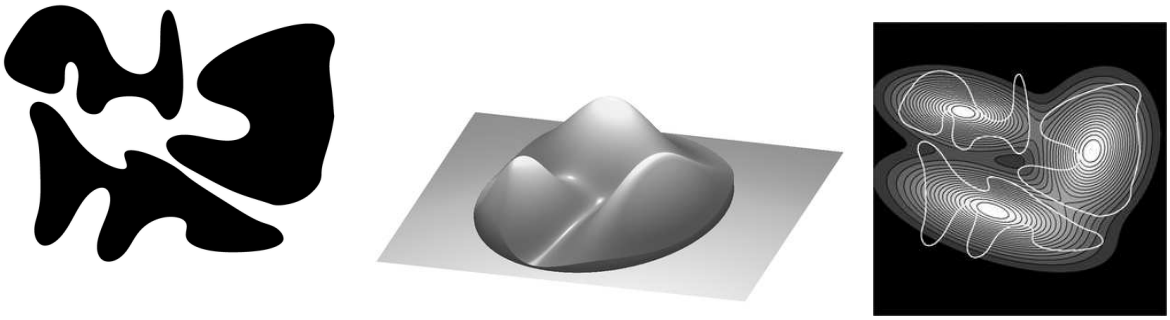
which can be computed directly from the input image without knowing the partitioning $\mathcal{D}_i$. The resulting system of equations has $\ell n(n+1)$ unknowns:

$$\int_{\mathcal{D}} \omega_j(\mathbf{x})d\mathbf{x} = \sum_{i=1}^{\ell} |\mathbf{A}_i| \int_{\mathcal{D}'_i} \omega_j(\mathbf{A}_i\mathbf{y})d\mathbf{y}, \quad j = 1, \ldots, m\ . \tag{5.23}$$

The solution of the above system provides all the unknown parameters of the overall deformation. Since each ω_j provides one equation, we need $m \geq \ell n(n+1)$ nonlinear functions to solve for ℓ linear transformations. In practice, $m > \ell n(n+1)$ yielding an over-determined system for which a least squares solution is obtained [8]. The quantitative evaluation of the proposed algorithm on a large synthetic dataset containing 2D and 3D images has been presented in [8]. The results show that the method performs well and robust against segmentation errors. The method has been succesfully applied to 2D real images of a tangram puzzle (see Fig. 5.3) as well as to volumetric medical images used for surgical planning (see Fig. 5.11) [8].

5.5 Solution via a linear system of equations

Let us now consider Eq. (5.5) and construct a system of linear equations [6, 7]. The crucial step of this approach is to construct a pair of *covariant functions* satisfying Eq. (5.3). Once these functions are established, we can adopt the direct method from [108, 109] to solve for the unknown affine transformation given in its inhomogeneous form $(\mathbf{A}, \mathbf{t})$. When graylevel images are considered, the image functions themself serve as appropriate covariant functions [108, 109]. Unfortunately, the construction of such functions for binary images is quite



(a) (b) (c)

Figure 5.4: *Gaussian PDFs fitted over a compound shape yield a consistent coloring. (a) Original shape; (b) 3D plot of the Gaussian PDFs over the elliptic domain with $r = 2$; (c) Gaussian densities as a grayscale image. The white contour shows shape boundaries.*

challenging due to the lack of radiometric information: These functions must be based on the only available geometric information [6, 7].

5.5.1 Construction of covariant functions

Since we know that the *template* and *observation* are identical up to an affine transformation (this is stated in Eq. (5.1)), we do not need to represent shapes. Therefore we can safely consider the points of the *template* as a sample from a bivariate normally distributed random variable denoted by $X \sim N(\mu, \mathbf{\Sigma})$ with probability density function (PDF) [6]

$$p(\mathbf{x}) = \frac{1}{2\pi\sqrt{|\mathbf{\Sigma}|}} \exp\Big(-\frac{1}{2}(\mathbf{x}-\mu)^T \mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)\Big) .$$

Applying any linear transformation to X results also in a bivariate normal random variable $Y = \mathbf{A}X + \mathbf{t}$ with parameters

$$X \overset{(\mathbf{A},\mathbf{t})}{\mapsto} Y \sim N(\mu', \mathbf{\Sigma}') = N(\mathbf{A}\mu + \mathbf{t}, \mathbf{A}\mathbf{\Sigma}\mathbf{A}^T) . \tag{5.24}$$

The parameters of the probability densities $N(\mu, \mathbf{\Sigma})$ and $N(\mu', \mathbf{\Sigma}')$ can be readily estimated as the sample means and covariances of the point coordinates, while $|\mathbf{A}|$ can be expressed from $\mathbf{A}\mathbf{\Sigma}\mathbf{A}^T = \mathbf{\Sigma}'$ as

$$|\mathbf{A}||\mathbf{\Sigma}||\mathbf{A}^T| = |\mathbf{\Sigma}'| , \quad \text{hence} \quad |\mathbf{A}| = \sqrt{|\mathbf{\Sigma}'|/|\mathbf{\Sigma}|} . \tag{5.25}$$

From a geometric point of view, the mean values μ and μ' represent the center of mass of the *template* and *observation* respectively, while $\mathbf{\Sigma}$ and $\mathbf{\Sigma}'$ capture the orientation and eccentricity of the shapes. Fig. 5.4 shows a compound shape and the fitted Gaussian densities of each component.

Now let us have a closer look at the relationship between $p(\mathbf{x})$ and $s(\mathbf{y})$, the PDF of Y. It follows from Eq. (5.24), that $\mathbf{\Sigma}'^{-1} = \mathbf{A}^{-T}\mathbf{\Sigma}^{-1}\mathbf{A}^{-1}$, furthermore

$$(\mathbf{y} - \mu') = (\mathbf{A}\mathbf{x} + \mathbf{t} - (\mathbf{A}\mu + \mathbf{t})) = (\mathbf{A}\mathbf{x} - \mathbf{A}\mu) \; .$$

We thus get

$$\begin{aligned} s(\mathbf{y}) &= \frac{\exp\Big(-\frac{1}{2}(\mathbf{A}\mathbf{x} - \mathbf{A}\mu)^T\mathbf{A}^{-T}\mathbf{\Sigma}^{-1}\mathbf{A}^{-1}(\mathbf{A}\mathbf{x} - \mathbf{A}\mu)\Big)}{2\pi\sqrt{|\mathbf{A}||\mathbf{\Sigma}|\mathbf{A}^T|}} \\ &= \frac{\exp\Big(-\frac{1}{2}(\mathbf{x} - \mu)^T\mathbf{A}^T\mathbf{A}^{-T}\mathbf{\Sigma}^{-1}\mathbf{A}^{-1}\mathbf{A}(\mathbf{x} - \mu)\Big)}{|\mathbf{A}|2\pi\sqrt{|\mathbf{\Sigma}|}} = \frac{p(\mathbf{x})}{|\mathbf{A}|} \; . \end{aligned}$$

Finally, substituting back $|\mathbf{A}|$ from Eq. (5.25), we get

$$\sqrt{|\mathbf{\Sigma}|}p(\mathbf{x}) = \sqrt{|\mathbf{\Sigma}'|}s(\mathbf{y}) \; . \tag{5.26}$$

It is well known, that the normalizing constant $1/(2\pi\sqrt{|\mathbf{\Sigma}|})$ in the density functions ensures that the integral of the PDF evaluates to 1. It is also the maximum value of the density function, which is inversely proportional to the area of the shape. This dependence on the shape size may cause numerical instabilities hence we define our *covariant functions* $\mathcal{P}, \mathcal{S} : \mathbb{R}^2 \to \mathbb{R}$ as the unnormalized densities

$$\begin{aligned} \mathcal{P}(\mathbf{x}) &= 2\pi\sqrt{|\mathbf{\Sigma}|}p(\mathbf{x}) = \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^T\mathbf{\Sigma}^{-1}(\mathbf{x} - \mu)\right) \; , \\ \mathcal{S}(\mathbf{y}) &= 2\pi\sqrt{|\mathbf{\Sigma}'|}s(\mathbf{y}) = \exp\left(-\frac{1}{2}(\mathbf{y} - \mu')^T\mathbf{\Sigma}'^{-1}(\mathbf{y} - \mu')\right) \; . \end{aligned} \tag{5.27}$$

Since the covariance matrices and mean vectors can be computed from the images, both $\mathcal{P}$ and $\mathcal{S}$ is obtained directly from the input shapes and they are *covariant* satisfying Eq. (5.3):

$$\mathcal{P}(\mathbf{x}) = \mathcal{S}(\mathbf{A}\mathbf{x} + \mathbf{t}) = \mathcal{S}(\mathbf{y}) \; . \tag{5.28}$$

Hence, we only make use of the statistics of the foreground pixels in order to compute a *covariant function pair*. We remark that we don't represent the shape itself, instead we use this function pair to construct our system of linear equations.

5.5.2 Linear estimation of affine parameters

Multiplying Eq. (5.1) and Eq. (5.28), and integrating out individual point correspondences, we get the following equation:

$$\int_{\mathcal{F}_t} \mathbf{x}\mathcal{P}(\mathbf{x})d\mathbf{x} = |\mathbf{A}|^{-1} \int_{\mathcal{F}_o} \mathbf{A}^{-1}(\mathbf{y} - \mathbf{t})\mathcal{S}(\mathbf{y})d\mathbf{y} \; ,$$

Adopting a set of nonlinear functions $\eta : \mathbb{R} \to \mathbb{R}$, we can easily generate new equations according to Eq. (5.5):

$$\int_{\mathcal{F}_t} \mathbf{x}\eta(\mathcal{P}(\mathbf{x}))d\mathbf{x} = |\mathbf{A}|^{-1} \int_{\mathcal{F}_o} \mathbf{A}^{-1}(\mathbf{y} - \mathbf{t})\eta(\mathcal{S}(\mathbf{y}))d\mathbf{y} \; . \tag{5.29}$$

If q_{ki} denotes the elements of $\mathbf{A}^{-1}$ and $-\mathbf{A}^{-1}\mathbf{t}$

$$\mathbf{A}^{-1} = \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix} \quad \text{and} \quad -\mathbf{A}^{-1}\mathbf{t} = \begin{bmatrix} q_{13} \\ q_{23} \end{bmatrix} \; ,$$

we can expand the integrals yielding the following linear system

$$|\mathbf{A}| \int_{\mathcal{F}_t} x_k \eta(\mathcal{P}(\mathbf{x}))d\mathbf{x} =$$

$$\sum_{i=1}^{2} q_{ki} \int_{\mathcal{F}_o} y_i \eta(\mathcal{S}(\mathbf{y}))d\mathbf{y} \; + \; q_{k3} \int_{\mathcal{F}_o} \eta(\mathcal{S}(\mathbf{y}))d\mathbf{y}, \; k = 1, 2 \; .$$

Adopting a set of linearly independent functions $\{\eta_i\}_{i=1}^{\ell}$, we can rewrite the system in matrix form

$$\begin{bmatrix} \int_{\mathcal{F}_o} y_1\eta_1(\mathcal{S}(\mathbf{y})) & \int_{\mathcal{F}_o} y_2\eta_1(\mathcal{S}(\mathbf{y})) & \int_{\mathcal{F}_o} \eta_1(\mathcal{S}(\mathbf{y})) \\ \vdots & \vdots & \vdots \\ \int_{\mathcal{F}_o} y_1\eta_\ell(\mathcal{S}(\mathbf{y})) & \int_{\mathcal{F}_o} y_2\eta_\ell(\mathcal{S}(\mathbf{y})) & \int_{\mathcal{F}_o} \eta_\ell(\mathcal{S}(\mathbf{y})) \end{bmatrix} \begin{bmatrix} q_{k1} \\ q_{k2} \\ q_{k3} \end{bmatrix}$$

$$= |\mathbf{A}| \begin{bmatrix} \int_{\mathcal{F}_t} x_k\eta_1(\mathcal{P}(\mathbf{x})) \\ \vdots \\ \int_{\mathcal{F}_t} x_k\eta_\ell(\mathcal{P}(\mathbf{x})) \end{bmatrix}, \quad k = 1, 2 \; . \tag{5.30}$$

The solution of this linear system provides the parameters of the registration. If $\ell > 3$ then the system is over-determined and the solution is obtained as a least squares solution. Note that independently of the number of systems, the coefficient matrix on the left hand side of Eq. (5.30) need to be computed only once. Hence the complexity of the algorithm depends linearly on the size of the shapes.

5.5.3 Choosing the integration domain

A trivial choice for the domains in our integral equation Eq. (5.29) is the foreground regions $\mathcal{F}_t$ and $\mathcal{F}_o$ [6]. Since the parameters of the transformation are estimated by integrating over the segmented domains, this approach works well as long as we have a near-perfect segmentation. Unfortunately, this is rarely encountered in reality [7]. Therefore a clear disadvantage of this approach is that any segmentation error will inherently result in erroneous integrals causing misalignment. However, image analysis often deals with the matching of objects composed of several parts, yielding a group of disjoint shapes when segmented. The topology of such compound shapes will not change under the action of the affine group. In such

cases, we use a more robust method to define corresponding integration domains making use of the statistics of compound shapes instead of their segmentation.

Let us assume that the *template* consists of $m \geq 2$ disjoint shapes. This is the typical output of classical region-based segmentation algorithms, where the labelling of the different regions results in disjoint shapes, but similar results can be achieved by detecting the connected components of a compound object. In both cases, the input of our method will be a pair of labeled images, where each component on the *template* has exactly one corresponding shape on the *observation*, *i.e.* there exists a bijective mapping between the *template* and *observation* components under the transformation $(\mathbf{A}, \mathbf{t})$. For each pair of corresponding components, we can establish *covariant functions* $\mathcal{P}_i$ and $\mathcal{S}_i$ similar to Eq. (5.27):

$$\begin{aligned} \mathcal{P}_i(\mathbf{x}) &= \exp\Big(-\frac{1}{2}(\mathbf{x}-\mu_i)^T\mathbf{\Sigma}_i^{-1}(\mathbf{x}-\mu_i)\Big) \\ \mathcal{S}_i(\mathbf{y}) &= \exp\Big(-\frac{1}{2}(\mathbf{y}-\mu_i')^T\mathbf{\Sigma}_i'^{-1}(\mathbf{y}-\mu_i')\Big)\ , \end{aligned}$$

where $\mathbf{\Sigma}_i$,$\mathbf{\Sigma}_i'$ and μ_i,μ_i' are the covariance matrices and mean vectors of the i^{th} shape on the *template* and *observation*, respectively. This yields an equation similar to Eq. (5.29) for each $1 \leq i \leq m$. If the correspondence between components would be known then we could simply construct a system of m equations and solve for the unknowns. As such a matching is not known, we will sum these relations yielding

$$\sum_{i=1}^{m}\int_{\mathcal{F}_t}\mathbf{x}\eta(\mathcal{P}_i(\mathbf{x}))d\mathbf{x} = \int_{\mathcal{F}_t}\mathbf{x}\sum_{i=1}^{m}\eta(\mathcal{P}_i(\mathbf{x}))d\mathbf{x} =$$

$$|\mathbf{A}|^{-1}\int_{\mathcal{F}_o}\mathbf{A}^{-1}(\mathbf{y}-\mathbf{t})\sum_{i=1}^{m}\eta(\mathcal{S}_i(\mathbf{y}))d\mathbf{y}\ . \tag{5.31}$$

Instead of establishing $\mathcal{P}_i$ and $\mathcal{S}_i$ and then constructing Eq. (5.31), we can sum these functions [7] resulting another *covariant function pair* P and S. Hence, we have an alternative way to define our *covariant functions* [7]:

$$P = \sum_{i=0}^{m}\mathcal{P}_i\ , \quad \text{and} \quad S = \sum_{i=0}^{m}\mathcal{S}_i\ ,$$

where $\mathcal{P}_0$ and $\mathcal{S}_0$ corresponds to the the overall shape (*i.e.* the whole foreground region) of the *template* and the *observation*, respectively. Note that these sums are mixtures of unnormalized Gaussian densities which can also be interpreted as a consistent coloring of the template and observation respectively (see Fig. 5.4(b)). By consistent coloring, we mean that these functions preserve the effect of the unknown transformation. Furthermore, these functions can be constructed exactly and uniquely from the object points alone without any knowledge about the aligning transformation. As a result, we can transform the original binary images into graylevel ones, where corresponding pixels have exactly the same gray

value. Nevertheless, P and S can also be used in Eq. (5.29), however, we recommend to use Eq. (5.31), since it can be numerically more efficient to compute.

The next step is to get rid of the segmentation domains $\mathcal{F}_t$ and $\mathcal{F}_o$. Our goal is to to select appropriate domains $\mathcal{D}_t$ and $\mathcal{D}_o$ satisfying the following properties

1. they are related by the unknown transformation $\mathbf{A}\mathcal{D}_t + \mathbf{t} = \mathcal{D}_o$
2. the integrands are rich enough (*i.e.* have characteristic pattern) within the selected domains.

The key idea is using the statistics of the whole *template* and *observation* objects. Indeed, the overall shape (*i.e.* the whole foreground region) of the *template* and *observation* also gives rise to a pair of covariant Gaussian densities $p(\mathbf{x})$ and $s(\mathbf{y})$. Since the equidensity contours of these PDFs are ellipsoids centered at the mean, it is natural to chose a corresponding pair of these ellipses as the integration domain. Simplifying Eq. (5.26), we get the well known Mahalanobis-distance which defines a metric invariant under the unknown transformation $(\mathbf{A}, \mathbf{t})$:

$$(\mathbf{x} - \mu)^T \mathbf{\Sigma}^{-1} (\mathbf{x} - \mu) = (\mathbf{y} - \mu')^T \mathbf{\Sigma}'^{-1} (\mathbf{y} - \mu') \ .$$

Corresponding domains can then be obtained by selecting points whose Mahalanobis-distance are less than r^2 from the mean:

$$\mathcal{D}_t = \{\mathbf{x} \in \mathbb{R}^2 | (\mathbf{x} - \mu)^T \mathbf{\Sigma}^{-1} (\mathbf{x} - \mu) \leq r^2\} \tag{5.32}$$

$$\mathcal{D}_o = \{\mathbf{y} \in \mathbb{R}^2 | (\mathbf{y} - \mu')^T \mathbf{\Sigma}'^{-1} (\mathbf{y} - \mu') \leq r^2\} \tag{5.33}$$

To satisfy property 2), we may choose an ellipse according to the *two sigma rule* (*i.e.* $r = 2$), which guarantees that about 95% of values are within the enclosed ellipsoid (see Fig. 5.4). Experiments show that good alignments can be achieved by settings ranging from $r = 1$ to $r = 3$.

In summary, all we need to construct a system of linear equations are the means and covariances of the input shapes. Based on these statistics, we can select the integration domains and construct appropriate covariant functions yielding the system Eq. (5.34) similar to Eq. (5.30).

$$\begin{bmatrix} \int_{\mathcal{D}_o} y_1 \sum_{i=1}^{m} \eta_1(\mathcal{S}_i(\mathbf{y})) & \int_{\mathcal{D}_o} y_2 \sum_{i=1}^{m} \eta_1(\mathcal{S}_i(\mathbf{y})) & \int_{\mathcal{D}_o} \sum_{i=1}^{m} \eta_1(\mathcal{S}_i(\mathbf{y})) \\ \vdots & \vdots & \vdots \\ \int_{\mathcal{D}_o} y_1 \sum_{i=1}^{m} \eta_\ell(\mathcal{S}_i(\mathbf{y})) & \int_{\mathcal{D}_o} y_2 \sum_{i=1}^{m} \eta_\ell(\mathcal{S}_i(\mathbf{y})) & \int_{\mathcal{D}_o} \sum_{i=1}^{m} \eta_\ell(\mathcal{S}_i(\mathbf{y})) \end{bmatrix} \begin{bmatrix} q_{k1} \\ q_{k2} \\ q_{k3} \end{bmatrix}$$

$$= |\mathbf{A}| \begin{bmatrix} \int_{\mathcal{D}_t} x_k \sum_{i=1}^{m} \eta_1(\mathcal{P}_i(\mathbf{x})) \\ \vdots \\ \int_{\mathcal{D}_t} x_k \sum_{i=1}^{m} \eta_\ell(\mathcal{P}_i(\mathbf{x})) \end{bmatrix}, \ k = 1, 2 \ . \tag{5.34}$$

	Runtime (sec.)	ϵ (pixel)	δ (%)
Polynomial	0.98	0.51	0.15
Linear	1.5	5.42	2.6
Mult. covar. functions	0.33	0.54	0.19

Table 5.1: *Registration results on a benchmark dataset of synthetic shapes.*

Figure 5.5: *Alignment of hip prosthesis X-ray images using a polynomial system of equations with ω functions $\{x, x^2, x^3\}$. Registration results are shown as an overlayed contour on the second image.*

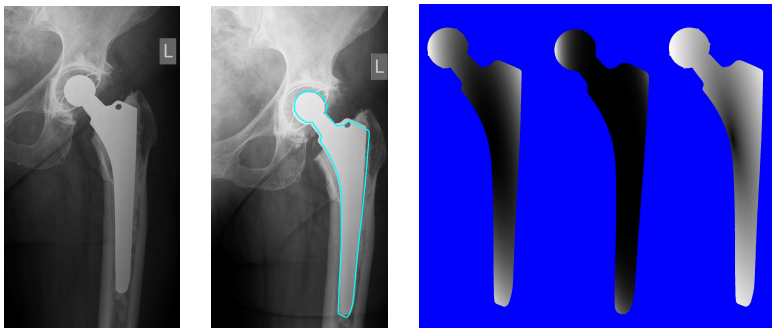


Figure 5.6: *Alignment of a hip prosthesis X-ray image using a linear system of equations with ω functions $\{x, x^3, x^{1/3}\}$ (corresponding colorizations are shown on the right). Registration result is shown as an overlayed contour on the second image.*

Figure 5.7: *Alignment of a traffic sign images using a linear system of equations with multiple shape parts. The first image shows the elliptic integration domain with the compound covariant function fitted over the* template*. Registration results are shown as an overlayed contour on the second image.*

5.6 Discussion

Based on the Matlab implementation of the polynomial (Section 5.3 and linear (Section 5.5) methods, we give a brief quantitativ evaluation on a set of more than 1000 synthetically generated observations for 60 different shapes. The applied transformations were randomly composed of $0°, 10°, \ldots, 350°$ rotations; $0, 0.4, \ldots, 1.2$ shearings; $0.5, 0.7, \ldots, 1.9$ scalings, and $-20, 0, 20$ translations along both axes. The resulting images are of size $\approx 1400 \times 1400$. For evaluation, we have computed two error measures: the Dice coefficient as $\delta = \frac{|R \triangle O|}{|R|+|O|} \cdot 100\%$, where $\triangle$ denotes symmetric difference, while T, R and O are the pixels of the *template*, *registered* object and *observation* respectively; and $\epsilon = \frac{1}{|T|} \sum_{\mathbf{p} \in T} \|(\mathbf{A} - \tilde{\mathbf{A}})\mathbf{p}\|$, which measures the average distance between the true $\mathbf{A}$ and the estimated $\tilde{\mathbf{A}}$ transformation. These results are shown in Table 5.1. Based on these numbers, it is clear that the polynomial solution provides rather good alignments at the price of $\approx$ 1sec. CPU time. The linear system based on a single pair of covariant functions given in Eq. (5.30) works well when there are no segmentation errors, but deteriorates quickly when pixels are missing. On the other hand, the linear system with multiple pairs of covariant functions given in Eq. (5.34) clearly outperforms the polynomial solution in terms of CPU time as well as in robustness: even for 90% missing pixels [7], this method still provides acceptable alignments while the polynomial system fails over 50% [9]. We also remark, that -like any other area based method- both approaches are quite sensitive to occlusions as it yields large errors in the system of equations.

In Fig. 5.5 and Fig. 5.6, we show registration results on real X-ray images. These results also confirm the higher precision of the polynomial system. While the multiple covariant function approach cannot be applied on these images since we only have a single shape,

Figure 5.8: *Fusion of hip prosthesis X-ray image pairs by registering follow up images using a 2D affine transformation (typical CPU time is around* 1 *sec. in Matlab).*

Fig. 5.7 shows the alignment of traffic signes where -due to the compound shape of these signs- the multiple covariant function approach works pretty well.

A more detailed evaluation of these methods can be found in [6, 7, 9–11, 49–51]

5.7 Medical applications

Fast rigid-body registration of bone structures is important in image guided surgical planning in execution for registering pre-operative volumes to intra-operative ones. Zhang *et al.* give an overview of surface based registration techniques [203] and propose a 15 times faster method than standard Iterative Closest Point (ICP) methods. However, it still takes around one minute to register vertebrae models segmented from high resolution CT images. If the segmentation is available, our method could be used instead of ICP to get the result faster.

Affine registration is also applied for creation of brain atlases, at least as preliminary step before an elastic or non-linear part [104]. Automatic initial placement of deformable organ models can also benefit from fast linear registrations [186]. By collecting and transforming a set of images to a common reference frame, a probabilistic atlas can be produced for various organs. Using the same registration method, this probabilistic atlas data can be mapped to the space of the study to be segmented. Here precise alignment of all anatomical structures is not crucial, the focus is on fast execution.

Herein we present some medical applications of our linear registration framework.

5.7.1 Fusion of hip prosthesis X-ray images

Hip replacement [84, 111] is a surgical procedure in which the hip joint is replaced by a prosthetic implant. In the short post-operative time, infection is a major concern. An inflammatory process may cause bone resorption and subsequent loosening or fracture, often requiring revision surgery. In current practice, clinicians assess loosening by inspecting a

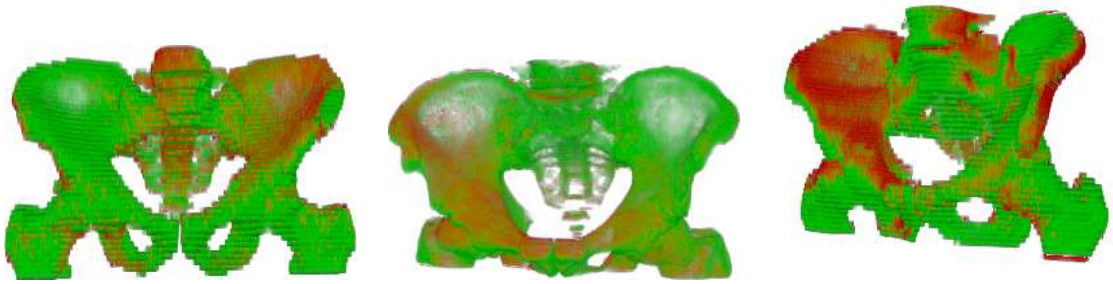


Figure 5.9: *Registration of pelvic CT data: superimposed registered 3D bone models (typical CPU time is around* 0.25 *sec for 1 megavoxel objects using our Java demo program). The first two cases show good alignment. Even the third one provides a good approximation of the true alignment.*

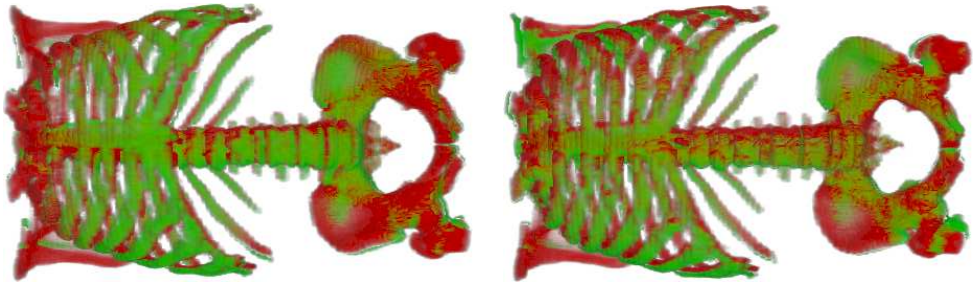


Figure 5.10: *Registration of thoracic CT data: superimposed registered 3D bone models. Perfect alignment is not possible due to the relative movements of the bone structure. Affine alignment results are used as a good starting point for* e.g. *lymph node detection.*

number of post-operative X-ray images of the patient's hip joint, taken over a period of time. Obviously, such an analysis requires the registration of X-ray images. Even visual inspection can benefit from registration as clinically significant prosthesis movement can be very small.

There are two main challenges in registering hip X-ray images: One is the highly non-linear radiometric distortion [87] which makes any greylevel-based method unstable. Fortunately, the segmentation of the prosthetic implant is quite straightforward [162] so shape registration is a valid alternative here. The second problem is that the true transformation is a projective one which depends also on the position of the implant in 3D space. Indeed, there is a rigid-body transformation in 3D space between the implants, which becomes a projective mapping between the X-ray images. The affine assumption is a good approximation here, as the X-ray images are taken in a well defined *standard position* of the patient's leg. For diagnosis, the area around the implant (especially the bottom part of it) is the most important for the physician. It is where the registration must be the most precise. Based on such an alignment, we can *e.g.* visualize the fused follow-up images for evaluation by an expert (see Fig. 5.8).

5.7.2 Registration of pelvic and thoracic CT volumes

The application of the affine solution to 3D objects has been presented in [49–51]: Typical medical applications include the alignment of pelvic and thoracic CT volumes based on segmented bony structures.

CT image pairs delineating the pelvic area were acquired at different times. The pairs of images were from the same patient. The spatial resolution of the CT studies were around $0.6 - 0.8$ mm in-slice. The slice distance was 5 mm in 11 cases, 2.5 mm in 4 cases. We also got three CT thoracic studies of the same person acquired by a PET-CT scanner. Here the in-slice resolution was 0.9766 mm and the slice distance was 3.27 mm. We used the full thoracic region and the extracted pelvis region also.

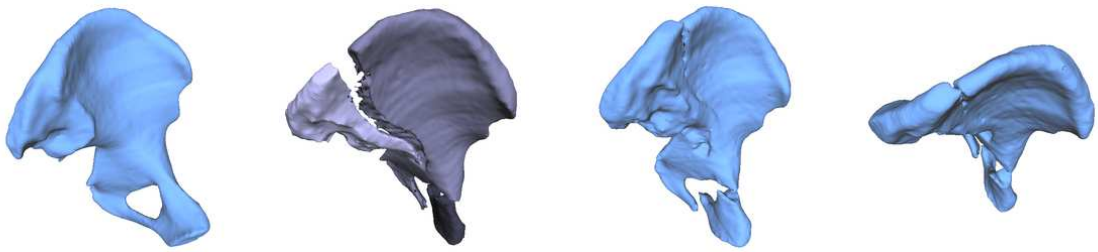
Our algorithm was implemented in Matlab 7.7 and was run on a desktop computer using Intel Core2 Duo processor at 2.4 GHz. The construction of the system of equations took around half a second, the optimization around 0.2 second. For example, the average computing time of a Mutual Information-based registration method was around 2 minutes for the smaller, and 6 minutes for the larger pelvic CT studies [50]. For the studies of the thorax, it usually took around 8 minutes to finish. This clearly shows the computational superiority of our approach.

The main challenges for *pelvic CT images* are poor image resolution, substantial segmentation errors, and slightly different placement of the femoral head and lower portion of the spine. These CT experiments (see Fig. 5.9 showed [50] that even when the physical deformation does not correspond exactly to the affine required, the results are good or at least acceptable.

For *thoracic images* the rigidity criterion no longer holds. Besides the femoral heads, the spine and the scapula can be moved with respect to each other. Nevertheless, rigid registration of such images are also common and important for *e.g.* lymphoma detections and changes over time using PET-CT scanners. PET images delineate the uptake of the contrast agent in organs (lymph nodes), while the CT modality can be used for registration and morphological localization. Here non-rigid registrations are discouraged since these could change the size of the organs. Sample registration results of our algorithm are shown in Fig. 5.10.

5.7.3 Bone fracture reduction

Complex bone fracture reduction frequently requires surgical care, especially when angulation or displacement of bone fragments are large. Since the input data is typically a volume CT image, bone fragment repositioning has to be performed in 3D space which requires an expensive special 3D haptic device and quite a lot of manual work. Therefore automatic bone fracture reduction can save considerable time, providing experts with a rough alignment which can be manually fine-tuned according to anatomic requirements. Since surgical planning involves the biomechanical analysis of the bone with implants, only rigid-body



template *observation* *realigned* bone fragments

Figure 5.11: *Bone fracture reduction (CPU time in Matlab was 15 sec. for these 1 megavoxel CT volumes). The* template *is obtained by mirroring the intact bone.*

transformations are allowed. In [85], a classical ICP algorithm is used to realign fractures. Winkelbach *et al.* [198] proposed an approach for estimating the relative transformations between fragments of a broken cylindrical structure by using well known surface registration techniques, like 2D depth correlation and the ICP algorithm. In [166], registration is solved by using quadrature filter phase difference to estimate local displacements.

In [8], we formulated the problem as an *affine puzzle* and found that our algorithm is able to solve bone fracture reduction on large volumetric medical images within a couple of seconds (see Fig. 5.11).

IN THIS CHAPTER:

6.

Nonlinear alignment of 2D shapes

In this chapter, we extend our framework to estimate the parameters of a general diffeomorphism that aligns a known shape and its distorted observation. Classical registration methods first establish correspondences between the shapes and then compute the transformation parameters from these landmarks. Herein, we trace back the problem to the solution of a system of nonlinear equations which directly gives the parameters of the aligning transformation. The proposed method provides a generic framework to recover any diffeomorphic deformation without established correspondences. It is easy to implement, not sensitive to the strength of the deformation, and robust against segmentation errors. The method has been applied to several commonly used transformation models. The performance of the proposed framework has been demonstrated on large synthetic datasets as well as in the context of various applications.

6.1 Introduction

When registering a pair of objects, first we have to characterize the possible deformations. From this point of view, registration techniques can be classified into two main categories: physical model-based and parametric or functional representation [118]. Herein, we deal with the latter representation, which typically originate from interpolation and approximation theory. Broadly used classes of such deformations are polynomial and spline-based transformations *e.g.* thin plate splines (TPS) [67, 202]. TPS models are typically based on interpolation: the control points of thin plate spline models are placed at extracted point matches and they usually include various regularizations, such as the bending energy [67]. Polynomial deformations are governed by fewer parameters and are acting *globally* on the shapes, hence regularization is usually not needed. Moreover, many non-polynomial transformation can be approximated by a polynomial one *e.g.* via a Taylor expansion [12].

In this chapter, we extend the framework presented in Chapter 5 to estimate nonlinear diffeomorphic transformations without establishing correspondences or restricting the strength of the deformation. The basic idea is to set up a system of nonlinear equations by integrating a set of nonlinear functions over the image domains and then solve it by classical *Levenberg-Marquardt algorithm* [147]. If perfect graylevel images would be available *without any radiometric distortion*, then the estimation of an aligning homeomorphism could be traced back to the solution of a linear system of equations [93]. In real applications, however, such a strict requirement cannot be satisfied. Herein, we will show that registration can be solved without making use of any intensity information. The main contribution is a unifying framework, which provides the registration of planar shapes under various diffeomorphic deformations (*e.g.* planar homography, polynomial or thin plate spline transformations). We have conducted a comprehensive test on a large set of synthetic images to demonstrate the performance and robustness of the proposed approach. The method has been successfully applied to a variety of real problems, *e.g.* alignment of hip prosthesis X-ray images, registering traffic signs and handwritten characters, or visual inspection of printed signs on hoses manufactured in automotive industry.

6.1.1 State of the art

While registration of grayscale or color images is well studied [53, 131, 134, 139, 148, 160, 180], the alignment of binary shapes [9, 60, 70, 124, 184, 185, 197] received less attention. Most of the current approaches are restricted to the affine group [9, 182, 184, 201]. In [9], Domokos and Kato showed that it is possible to trace back the affine matching problem to an exactly solvable polynomial system of equations. Moments and invariants also provide an efficient tool for recovering linear deformations [184]. A geometric, variational framework is introduced in [201], which uses active contours to simultaneously segment and register images. The method [201] is applied to medical image registration, where 2D and 3D rigid body transformations are considered. Another statistics-based algorithm is proposed in [182] for registration of edge-detected images, which utilizes edge pixel matching to determine

the ”best” translations. Then a statistical procedure, namely the McNemar test, is used to determine which candidate solutions are not significantly worst than the best ones. This allows for the construction of confidence regions in the registration parameters. Note that this method is limited to solving for 2D translations only [182].

In this paper, we are interested in nonlinear alignment of shapes, which is a more challenging problem. The most common nonlinear registration methods are based on point correspondences [60, 105, 200]. Although there are robust keypoint detectors like SIFT [142] or SURF [59], these are relying on rich intensity patterns thus their use is limited in binary registration. *Landmark-based* nonlinear shape matching has been addressed by Belongie *et al.* [60]. The method first searches for point correspondences between the two objects, then estimates the transformation using a generic thin plate spline model. The point matches are established using a novel similarity metric, called *shape context*, which consists in constructing a log-polar histogram of surrounding edge pixels. The advantage compared to traditional landmark based approaches is that landmarks need not be salient points and radiometric information is not involved. Basically the method can be regarded as matching two points sets, each of them being a dense sample from the corresponding shape's boundary. Obviously, there is no guarantee that point pairs are exactly corresponding because of the sampling procedure. However, having a dense sample will certainly keep mismatch error at a minimum. The correspondences are simply established by solving a linear assignment problem, which requires time consuming optimization methods. For example, the complexity of the Hungarian method adopted in [60] is $\mathcal{O}(N^3)$.

An important class of nonlinear transformations is the plane to plane homography which aligns two images of the same planar object taken from different views. Lepetit and Fua proposed a method [139] for keypoint recognition on grayscale images. The main idea is to find keypoints during a training phase where a projectively different image set of target object is used. Although the recognition of keypoints becomes very fast, the training phase is very time consuming. In [134], a Fourier domain based approach is presented using intensities for computing the image-to-image transformation. Images are transformed into the Fourier domain where the transformation parameters are computed using cross correlation methods. In [124], planar homography is computed in the Fourier domain from a starting affine estimation using the shape contours. In [197], the concept of characteristic line is employed to show some useful properties of a planar homography matrix, which relate with Euler angles of the planar pattern.

Stochastic models with iterative optimization techniques are also quite popular in this domain: In [105], Guo *et al.* propose a method to register shapes which underwent diffeomorphic distortions, where simulated annealing is used to estimate point correspondences between the boundary points of the shapes. A Brownian motion model in the group of diffeomorphisms has been introduced in [160]. The authors exploit a prior for warps based on a simple invariance principle under warping. An estimation based on this prior guarantees an invertible, source-destination symmetric, and warp-invariant warp. The maximum-likelihood warp is then computed via a PDE scheme. [180] uses a Markov Random Field model to solve the registration problem. The deformation is described by a field of discrete

variables, representing displacements of (blocks of) pixels. Exact maximum a posteriori inference is intractable hence a linear programming relaxation technique is used. In [189], the registration problem is formulated as probabilistic inference using a generative model and the expectation-maximization algorithm. The authors define a data-driven technique which makes use of shape features. This gives a hybrid algorithm which combines generative and discriminative models. The measure of similarity is defined in terms of the amount of transformation required. The shapes are represented by sparse-point or continuous-contour representations depending on the form of the data. Klein *et al.* presented a stochastic gradient descent optimization method with adaptive step size prediction [131]. This method employs a stochastic subsampling technique to accelerate the optimization process. The selection mechanism for the method's free parameters takes into account the chosen similarity measure, the transformation model, and the image content, in order to estimate proper values for the most important settings.

Bronsetin *et al.* studied some fundamental problems in the analysis of non-rigid deformable shapes [70]. In particular, a novel similarity criteria for shape comparison and its extension to partial similarity has been proposed. They showed that the correspondence problem is also solvable using their similarity metric. In [200], Wörz and Rohr proposed a novel approximation approach to registrate elastic deformations. This landmark-based method uses Gaussian elastic body splines. Other methods use variational techniques [148]. We note that these methods has a rather high computational demand. In [185], a non-rigid registration algorithms is proposed based on L_2 norm and information-theory.

Another common approach is to approximate a nonlinear deformation via piecewise linear transformations: In [53], a novel framework to fuse local rigid or affine components into a global invertible transformation, called Log-Euclidean polyaffine, has been presented. A simple algorithm is proposed to compute efficiently such transformations and their inverses on regular grids.

6.2 Registration framework

In the general case, we want to recover the parameters of an arbitrary $\varphi : \mathbb{R}^2 \to \mathbb{R}^2$ *diffeomorphism* which aligns a pair of shapes. Let us denote the point coordinates of the *template* and *observation* by $\mathbf{x} = [x_1, x_2]^T \in \mathbb{R}^2$ and $\mathbf{y} = [y_1, y_2]^T \in \mathbb{R}^2$ respectively. The following identity relation is assumed between the point coordinates of the shapes:

$$\mathbf{y} = \varphi(\mathbf{x}) \quad \Leftrightarrow \quad \mathbf{x} = \varphi^{-1}(\mathbf{y}), \tag{6.1}$$

where $\varphi^{-1} : \mathbb{R}^2 \to \mathbb{R}^2$ is the corresponding inverse transformation. Note that φ^{-1} always exists since a diffeomorphism is a bijective function such that both the function and its inverse have continuous mixed partial derivatives. Suppose that shapes are represented by their characteristic function $\mathbb{1} : \mathbb{R}^2 \to \{0, 1\}$, where 0 and 1 correspond to the background and foreground respectively. If we denote the *template* by $\mathbb{1}_t$ and the *observation* by $\mathbb{1}_o$, the

following equality also holds

$$\mathbb{1}_o(\mathbf{y}) = \mathbb{1}_o(\varphi(\mathbf{x})) = \mathbb{1}_t(\mathbf{x}), \tag{6.2}$$

since $\mathbf{x}$ and $\mathbf{y}$ are corresponding point coordinates.

Classical landmark based approaches would now set up a system of equations from Eq. (6.1) using point correspondences. However, we are interested in a direct approach without solving the correspondence problem. As a consequence, we cannot directly use Eq. (6.1)–(6.2) because we do not have established point pairs. However, we can multiply these equations and then integrate out individual point correspondences yielding

$$\int_{\mathbb{R}^2} \mathbf{y}\mathbb{1}_o(\mathbf{y})d\mathbf{y} = \int_{\mathbb{R}^2} \varphi(\mathbf{x})\mathbb{1}_t(\mathbf{x})\left|J_\varphi(\mathbf{x})\right| d\mathbf{x}, \tag{6.3}$$

where the integral transformation $\mathbf{y} = \varphi(\mathbf{x})$, $d\mathbf{y} = \left|J_\varphi(\mathbf{x})\right| d\mathbf{x}$ has been applied. The Jacobian determinant $|J_\varphi| : \mathbb{R}^2 \to \mathbb{R}$

$$|J_\varphi(\mathbf{x})| = \begin{vmatrix} \frac{\partial\varphi_1}{\partial x_1} & \frac{\partial\varphi_1}{\partial x_2} \\ \frac{\partial\varphi_2}{\partial x_1} & \frac{\partial\varphi_2}{\partial x_2} \end{vmatrix} \tag{6.4}$$

gives the measure of the transformation at each point. Note that in the case of affine (*i.e.* linear) transformations, the partial derivatives of the distortion are constants, hence the Jacobian is also constant and the transformation measure can be simply computed as the ratio of the shape areas. This property has been explored in [9]. Herein, however, the transformation is nonlinear causing the Jacobian to become a non-constant function of the coordinates.

Since multiplying with the characteristic functions essentially restricts the integral domains to the foreground regions $\mathcal{F}_t = \{\mathbf{x} \in \mathbb{R}^2 | \mathbb{1}_t(\mathbf{x}) = 1\}$ and $\mathcal{F}_o = \{\mathbf{y} \in \mathbb{R}^2 | \mathbb{1}_o(\mathbf{y}) = 1\}$, we obtain the following finite integral equation:

$$\int_{\mathcal{F}_o} \mathbf{y}d\mathbf{y} = \int_{\mathcal{F}_t} \varphi(\mathbf{x})\left|J_\varphi(\mathbf{x})\right| d\mathbf{x}. \tag{6.5}$$

The diffeomorphism φ can be decomposed as

$$\varphi(\mathbf{x}) = [\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x})]^T, \tag{6.6}$$

where $\varphi_1, \varphi_2 : \mathbb{R}^2 \to \mathbb{R}$ are coordinate functions. Hence Eq. (6.5), which is in *vector form*, can be decomposed into a system of two equations using these coordinate functions:

$$\int_{\mathcal{F}_o} y_i d\mathbf{y} = \int_{\mathcal{F}_t} \varphi_i(\mathbf{x})\left|J_\varphi(\mathbf{x})\right| d\mathbf{x}, \quad i = 1, 2. \tag{6.7}$$

The parameters of φ are the unknowns of these equations. Usually, φ has more than two unknown parameters therefore a system of two equations is not enough to recover φ.

6.2.1 Construction of the system of equations

First of all, let us notice that the identity relation in Eq. (6.1) remains valid when a function $\omega : \mathbb{R}^2 \to \mathbb{R}$ is acting on both sides of the equation [9, 43, 44]. Indeed, for a properly chosen ω

$$\omega(\mathbf{y}) = \omega(\varphi(\mathbf{x})) \quad \Leftrightarrow \quad \omega(\mathbf{x}) = \omega(\varphi^{-1}(\mathbf{y})). \tag{6.8}$$

Thus the following integral equation is obtained from Eq. (6.5)

$$\int_{\mathcal{F}_o} \omega(\mathbf{y}) d\mathbf{y} = \int_{\mathcal{F}_t} \omega(\varphi(\mathbf{x})) \left|J_\varphi(\mathbf{x})\right| d\mathbf{x}. \tag{6.9}$$

The basic idea of the proposed framework is to generate sufficiently many equations using a set of nonlinear ω functions. Let the number of parameters of φ denoted by k and let $\{\omega_i\}_{i=1}^{\ell}, \omega_i : \mathbb{R}^2 \to \mathbb{R}$ denote the set of adopted nonlinear functions. In order to solve for all unknowns, we need at least k equations, hence $\ell \geq k$. We thus obtain the following system of equations

$$\int_{\mathcal{F}_o} \omega_i(\mathbf{y}) d\mathbf{y} = \int_{\mathcal{F}_t} \omega_i(\varphi(\mathbf{x})) \left|J_\varphi(\mathbf{x})\right| d\mathbf{x}, \qquad i = 1, \ldots, \ell, \tag{6.10}$$

where each ω_i function provides one new equation. Note that the generated equations provide no new information, they simply impose additional constraints. Note also that these equations need to be algebraically independent. While this condition is difficult to verify in practice, it is also clear that linear independence of ω_i functions -which is easier to verify- is crucial, as linear dependency would result in algebraically dependent equations. Therefore in practice, we always use a set of *nonlinear* ω_i functions. The solution of the system gives the parameters of the aligning transformation. Intuitively, each ω_i generates a consistent coloring of the shapes as shown in Fig. 5.1. From a geometric point of view, Eq. (6.5) simply matches the center of mass of the *template* and *observation* while the new equations in Eq. (6.9) match the volumes over the shapes constructed by the nonlinear functions ω_i (see Fig. 5.1).

6.2.2 Discussion

6.2.2.1 Relation to moment-based approaches

Although the derivations in the previous section are not moment-based *per se*, it is interesting to analyze how the resulting equation of Eq. (6.10) is related to moments. Image moments and invariants [91] were introduced by Hu [121] for 2D pattern analysis. Since then, they became one of the most popular region-based descriptors because any shape can be reconstructed from its infinite set of moments [92]. Traditional two dimensional $(p+q)^{\text{th}}$ order moments of a function $\varrho : \mathbb{R}^2 \to \mathbb{R}$ are defined as

$$m_{pq} = \int_{\mathbb{R}^2} x_1^p x_2^q \varrho(\mathbf{x}) d\mathbf{x} \ ,$$

where $p, q \in \mathbb{N}_0$. When ϱ is an image function then these moments are also referred to as *image moments*. In the binary case, where objects are represented by their silhouette, ϱ is a characteristic function yielding $m_{pq} = \int_{\mathcal{F}} x_1^p x_2^q d\mathbf{x}$ with $\mathcal{F} = \{\mathbf{x} \in \mathbb{R}^2 : \varrho(\mathbf{x}) = 1\}$. This is often called *shape* or *geometric* moment as it only uses polynomials of the coordinates. Generally, orthogonal moments, such as Legendre [92] or Zernike moments [187], are numerically more stable than regular moments. We remark, however, that orthogonal moments can be expressed by regular moments.

In this sense, we can recognize a 0^{th} order *function moment* of ω_i in the left hand side of Eq. (6.10) (just like any function integral can be regarded as the 0^{th} order moment of the function itself). Similarly to Legendre or Zernike moments, our function moments could also be expressed in terms of *shape moments* whenever the adopted ω_i functions are polynomials. When ω_i is not polynomial then its Taylor expansion results in an approximating polynomial which in turn yields an *infinite* sum of shape moments. The right hand side of Eq. (6.10) is more complex as it includes the product of the unknown transformation $\varphi(\mathbf{x})$ and its Jacobian determinant $|J_\varphi(\mathbf{x})|$ which are not necessarily polynomials. Therefore, independently of the choice of ω_i, it can only be expressed in terms of shape moments by expanding it into a Taylor series.

It is thus clear that our system of equations outlined in Eq. (6.10) cannot be rewritten in terms of a finite set of classical *shape moments*, and hence not even in terms of orthogonal moments. This result corresponds to similar findings reported in [90, 190] in the context of projective invariants. What we propose in this paper is another approach, which –starting from the identity relation in Eq. (6.1)– builds up a framework to generate an arbitrary set of equations.

6.2.2.2 Invariance vs. covariance

Moment invariants [121, 190] are extensively studied as they provide a powerful tool for shape matching. Basically, invariants are functions immune to the action of a particular deformation. There is a well established theory on affine invariants [89, 91], but invariants of higher order deformations are hard to construct. Recently important results on the existence of projective moment invariants [90] as well as on generalized invariants, called *Implicit Moment Invariants* [88, 91], have also been reported. Herein, we are not interested in constructing invariants as, being immune to the deformation, they do not provide constraints on the actual transformation parameters. Instead, we need *covariant* functions that vary with the transformation $\varphi(\mathbf{x})$, hence constraining its parameters. Indeed, *invariance* and *covariance* play a complementary role: While invariants identify a shape regardless of its deformation, covariants identify the actual deformation.

6.2.2.3 Registration vs. matching

There is a fundamental difference between the problem of registration and *shape matching* [91]. In either case, we fix the family of possible transformations. In the case of match-

ing, we need to determine whether two objects are from the same class or not. For that purpose, it is enough to ask whether there exists a transformation which aligns the objects (*i.e.* whether they are on the same orbit of the fixed transformation class), but the aligning transformation is not of interest. However, in the registration problem we always assume that there exists a transformation which aligns the objects and we need to estimate its parameters. This explains why multiple object matching algorithms often make use of *invariants*, ignoring the effect of the unknown transformation, and why *covariance* is used to solve the registration problem. Due to the difficulty in finding appropriate invariants under elastic deformations [88, 91], nonlinear shape matching (or recognition) is often solved by registering a given *observation*, representing the deformed shape to be recognized, to the *templates* stored in a database [60]. A similarity metric is then constructed using the strength of the deformation (*e.g.* bending energy) and the shape is recognized as the *template* with the minimal distortion.

6.3 Choice of ω functions

Given $\varphi(\mathbf{x})$ and its Jacobian $|J_\varphi(\mathbf{x})|$ of a particular deformation model, the parameters of the aligning transformation are obtained as a solution of the system of equations Eq. (6.10). For constructing these equations, we need an appropriate set of functions $\{\omega_i\}_{i=1}^{\ell}$. Theoretically, any nonlinear function satisfying Eq. (6.8) could be applied. In practice, however there are two important considerations. First, our equations are always corrupted by errors arising from imperfect data (*e.g.* segmentation and discretization errors). Therefore the solution is obtained via least-squares minimization of the *algebraic error*. Since both sides of these equations contain an integral of the corresponding ω_i function, the characteristics of ω_i clearly influence the overall error. In particular, we expect an equal contribution from each equation in order to guarantee an unbiased error measure. Second, iterative least-squares minimization algorithms, like the *Levenberg-Marquardt algorithm* [147], require the evaluation of the equations at every iteration step. Thus the time complexity of the algorithm is considerably decreased if the integrals can be precomputed, hence avoiding scanning the image pixels at every iteration.

6.3.1 Normalization

The *algebraic error* of the system Eq. (6.10) is obtained as the sum of squared errors :

$$\sum_{i=1}^{\ell} \left(\int_{\hat{\varphi}(\mathcal{F}_t)} \omega_i(\mathbf{y}) d\mathbf{y} - \int_{\mathcal{F}_o} \omega_i(\mathbf{y}) d\mathbf{y} \right)^2,$$

where $\hat{\varphi}$ is the estimated transformation. On the other hand, the *geometric error* is measured as the absolute difference between the registered shapes:

$$|G| = |\hat{\varphi}(\mathcal{F}_t) \triangle \mathcal{F}_o|,$$

where $\triangle$ is the symmetric difference. Let $G_1 = \hat{\varphi}(\mathcal{F}_t)\backslash\mathcal{F}_o$ and $G_2 = \mathcal{F}_o\backslash\hat{\varphi}(\mathcal{F}_t)$, yielding $G = G_1 \cup G_2$ and $G_1 \cap G_2 = \emptyset$. Since $\int_{\hat{\varphi}(\mathcal{F}_t)\cap\mathcal{F}_o} (\omega_i(\mathbf{y}) - \omega_i(\mathbf{y}))\, d\mathbf{y} = 0$, the algebraic error can be expressed as

$$\sum_{i=1}^{\ell} \left(\int_{G_1} \omega_i(\mathbf{y}) d\mathbf{y} - \int_{G_2} \omega_i(\mathbf{y}) d\mathbf{y} \right)^2. \tag{6.11}$$

The i^{th} equation contributes to the error by the difference of the integrals of ω_i over the non-overlapping domains G_1 and G_2. Thus the magnitude of the contributed value depends not only on the geometric error G but also on the values ω_i takes over these domains. Large variations in the range of different ω_i functions yield an uneven contribution of different equations which leads to a biased algebraic error or, in extreme cases, to numerical instability.

A usual remedy is to normalize the coordinates of both shapes into the unit square ($[-0.5, 0.5] \times [-0.5, 0.5]$ in our experiments), and to choose ω_i with a range limited to a similar interval (*e.g.* $[-1, 1]$). This is achieved by dividing the integrals in Eq. (6.10) with an appropriate constant corresponding to the maximal magnitude of the integral. Since the integral of a given ω_i depends on the integration domain (*i.e.* the actual transformed shape), a trivial upper bound would be the infinite integral $\int_{\mathbb{R}^2} |\omega_i|$. Unfortunately, this integral may not be computed or yields an infinite value thus making this kind of normalization unfeasible. Therefore we need to find a finite domain which contains all intermediate shapes during the minimization process. We found experimentally (see Fig. 6.1), that the transformations occurring during the least-squares minimization process do not transform the shapes out of a circle with center in the origin and a radius $\frac{\sqrt{2}}{2}$ (*i.e.* the circumscribed circle of the unit square). We thus adopt the following constant

$$N_i = \int_{\|\mathbf{x}\| \le \frac{\sqrt{2}}{2}} |\omega_i(\mathbf{x})| d\mathbf{x}, \tag{6.12}$$

and the normalized version of Eq. (6.10) becomes

$$\frac{\int_{\mathcal{F}_o} \omega_i(\mathbf{y}) d\mathbf{y}}{N_i} = \frac{\int_{\mathcal{F}_t} \omega_i(\varphi(\mathbf{x}))\, |J_\varphi(\mathbf{x})|\, d\mathbf{x}}{N_i}, \quad i = 1, \ldots, \ell. \tag{6.13}$$

6.3.2 Computational efficiency

The *Levenberg-Marquardt algorithm* requires the evaluation of the equations at every iteration step. Unfortunately, the integrands on the right hand side of Eq. (6.13) include the unknowns implying that we have to recompute these integrals at each iteration, yielding a time complexity of $\mathcal{O}(k(N+M))$, where k is the number of iterations (typically around 1000 in our experiments), while N and M are the number of the foreground pixels of the *template* and *observation* respectively. If we could eliminate the unknowns from the integrands then the integrals could be precomputed and the runtime of the solver would become independent

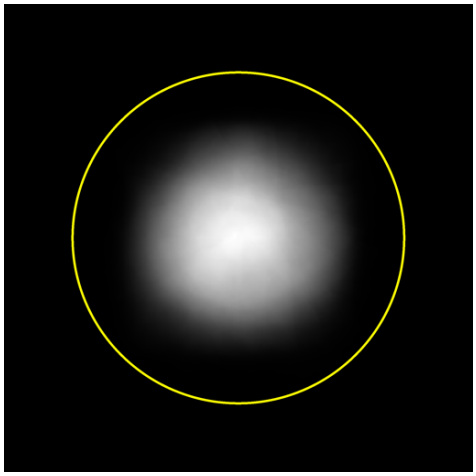


Figure 6.1: *Coverage of transformed shapes of* ≈ 1500 *synthetic observations during the minimization process. Pixel values represent the number of intermediate shapes that included a particular pixel. For reference, we also show the circle with radius* $\frac{\sqrt{2}}{2}$ *used for normalization.*

from the number of foreground pixels $M + N$. We will show that this can be achieved by applying polynomial ω_i functions in Eq. (6.13).

Let us suppose that $\varphi(\mathbf{x}) = \sum_{i=1}^{n} a_i \phi_i(\mathbf{x})$, where $a_i \in \mathbb{R}$, and $\phi_i : \mathbb{R}^2 \to \mathbb{R}^2$ are basis functions. Note that polynomial or thin plate spline deformations are of the above form while other diffeomorphisms can be approximated by the first few terms of their Taylor expansion [93] yielding the above representation. Furthermore, let us denote $\mathbf{a} = [a_1, \ldots, a_n]$ and $\phi(\mathbf{x}) = [\phi_1(\mathbf{x}), \ldots, \phi_n(\mathbf{x})]$.

Definition 6.3.1 *When a function* $f : \mathbb{R}^2 \to \mathbb{R}^2$ *is such that*

$$f\Big(\sum_{i=1}^{n} a_i \phi_i(\mathbf{x})\Big) = \sum_{j=1}^{m} g_j(\mathbf{a}) h_j(\phi(\mathbf{x})),$$

where $m \in \mathbb{N}$ *and* $g_j : \mathbb{R}^n \to \mathbb{R}$, $h_j : \mathbb{R}^{2n} \to \mathbb{R}^2$ *for* $1 \leq j \leq m$, *then we say* f *is* separable *with respect to* $\mathbf{a}$ *and* $\phi(\mathbf{x})$.

The following theorem states that applying polynomial ω_i functions in Eq. (6.13) results in a regular nonlinear equation with respect to the unknowns $a_1, \ldots, a_n$ instead of an integral equation.

Theorem 6.3.1 *When $f : \mathbb{R}^2 \to \mathbb{R}^2$ is polynomial, then the following equality holds:*

$$\int_{\mathcal{F}} f(\varphi(\mathbf{x}))\,|J_\varphi(\mathbf{x})|\,d\mathbf{x} = \sum_{i=1}^{m} g_i(\mathbf{a}) \int_{\mathcal{F}} h_i(\phi(\mathbf{x}))d\mathbf{x}, \tag{6.14}$$

where $m \in \mathbb{N}$ and $g_i : \mathbb{R}^n \to \mathbb{R}$, $h_i : \mathbb{R}^{2n} \to \mathbb{R}^2$ for $1 \leq i \leq m$.

The proof can be found in Appendix A.3 and in [12]. As a consequence, choosing a polynomial ω_i function allows us to eliminate the unknowns $\mathbf{a}$ from the integrand. Hence $\int_{\mathcal{F}} h_i(\phi(\mathbf{x}))d\mathbf{x}$ has to be computed only once and the time complexity of the solver becomes independent of the size of the input images.

6.3.3 Solution and complexity

The obtained system of equation is solved by iterative least squares minimization using the *Levenberg-Marquardt algorithm* (LM) [147]. The time complexity of the algorithm is $\mathcal{O}(N + M)$ whenever we adopt a polynomial set $\{\omega_i\}_{i=1}^{\ell}$. Note, that LM finds a local minimum. However, our numerical experiments show that the solution found by LM is quite close to the geometrically correct one. A theoretical analysis would be far too complex, but intuitively we can argue as follows: To avoid geometrically wrong local minima, proper normalization is crucial. As explained in Section 6.3.1, the equations need to be balanced and shapes must be normalized into the unit square. This guarantees, that initially shapes are overlapping making the identity transform a good initialization, while balanced equations eliminate undesirable bias during iterations caused by large coefficients in some equations. Finally, we have to remark that deformations with higher degree of freedom (*e.g.* TPS) may have many geometrically correct solutions (*i.e.* many transformation may produce an almost perfect alignment due to the fact that deformations are only visible around the boundary of the shapes). Therefore, although the parameter space is of higher dimension, LM also has a higher chance to find a local minima close to one of these correct solutions.

6.4 Modeling deformation fields

It is a quite common assumption in image registration, that the deformation field is smooth and invertible, especially when the resulting deformation field is further analyzed (*e.g.* in deformation-based morphometry or construction of shape models). Diffeomorphisms provide a convenient mathematical framework to describe such deformations. Various parametric models of diffeomorphic deformations have been proposed in the literature [202]. These are either based on a physical model (*e.g.* planar homography) or on a general parameterization using different basis functions (*e.g.* thin plate spline, B-spline). Herein, we will focus on some broadly used class of deformations, but our framework can be applied to other

nonlinear transformations as well (see Section 6.6.6, for instance). Essentially, all we need to apply our framework to a particular deformation model are the formulas of the adopted diffeomorphism $\varphi(\mathbf{x})$ and its Jacobian $|J_\varphi(\mathbf{x})|$.

6.4.1 Planar homography

Perspective images of planar scenes are usual in perception of man made environments. In such cases, a planar scene and its image are related by a plane to plane homography, also known as a plane projective transformation. Estimating the parameters of such transformations is a fundamental problem in computer vision with various applications.

Let us denote the homogeneous coordinates of the *template* and *observation* by $\mathbf{x}' = [x'_1, x'_2, x'_3]^T \in \mathbb{P}^2$ and $\mathbf{y}' = [y'_1, y'_2, y'_3]^T \in \mathbb{P}^2$, respectively. Planar homography is then a linear transformation in the projective plane $\mathbb{P}^2$

$$\mathbf{y}' = \mathbf{H}\mathbf{x}' \quad \Leftrightarrow \quad \mathbf{x}' = \mathbf{H}^{-1}\mathbf{y}' \,, \tag{6.15}$$

where $\mathbf{H} = \{H_{ij}\}$ is the unknown 3×3 transformation matrix that we want to recover. Note that $\mathbf{H}$ has only 8 degree of freedom thus one of its 9 elements can be fixed. Herein we will set $H_{33} = 1$. Although H_{33} could be 0 or small in general, the coordinates of the input shapes are normalized before matching into $[-0.5; 0.5] \times [-0.5; 0.5]$ with center of mass being the origin. Thus if H_{33} would be 0 then $\mathbf{H}$ would map the origin $[0, 0, x'_3]^T$ of the *template* into $[H_{13}x'_3, H_{23}x'_3, 0]^T$ on the *observation* (*i.e.* to infinity yielding an ellipse to become a parabola), which is quite unlikely to be observed in a real image pair. Similarly, if H_{33} is very small, then the origin is mapped to a distant point implying extreme distortion which is again unlikely in practice. These are close to degenerate situations for which a numerically stable solution may not exists anyway.

As usual, the inhomogeneous coordinates $\mathbf{y} = [y_1, y_2]^T \in \mathbb{R}^2$ of a homogeneous point $\mathbf{y}'$ are obtained by projective division

$$\begin{aligned} y_1 &= \frac{y'_1}{y'_3} = \frac{H_{11}x_1 + H_{12}x_2 + H_{13}}{H_{31}x_1 + H_{32}x_2 + 1} \equiv \chi_1(\mathbf{x}) \\ y_2 &= \frac{y'_2}{y'_3} = \frac{H_{21}x_1 + H_{22}x_2 + H_{23}}{H_{31}x_1 + H_{32}x_2 + 1} \equiv \chi_2(\mathbf{x}) \,, \end{aligned} \tag{6.16}$$

where $\chi_i : \mathbb{R}^2 \to \mathbb{R}$. Indeed, planar homography is a linear transformation in the projective plane $\mathbb{P}^2$, but it becomes nonlinear within the Euclidean plane $\mathbb{R}^2$. The nonlinear transformation corresponding to $\mathbf{H}$ is denoted by $\chi : \mathbb{R}^2 \to \mathbb{R}^2$, $\chi(\mathbf{x}) = [\chi_1(\mathbf{x}), \chi_2(\mathbf{x})]^T$ and the Jacobian determinant $|J_\chi| : \mathbb{R}^2 \to \mathbb{R}$ is given by

$$|J_\chi(\mathbf{x})| = \begin{vmatrix} \frac{\partial \chi_1}{\partial x_1} & \frac{\partial \chi_1}{\partial x_2} \\ \frac{\partial \chi_2}{\partial x_1} & \frac{\partial \chi_2}{\partial x_2} \end{vmatrix} = \frac{|\mathbf{H}|}{(H_{31}x_1 + H_{32}x_2 + 1)^3} \,. \tag{6.17}$$

6.4.2 Thin plate spline

Thin plate splines (TPS) [67, 101, 202] are widely used to approximate non-rigid deformations using radial basis functions. Given a set of control points $\mathbf{c}_k \in \mathbb{R}^2$ and associated mapping coefficients $a_{ij}, w_{ki} \in \mathbb{R}$ with $i = 1, 2$, $j = 1, 2, 3$ and $k = 1, \ldots, K$, the TPS interpolating points $\mathbf{c}_k$ is given by [202]

$$\varsigma_i(\mathbf{x}) = a_{i1}x_1 + a_{i2}x_2 + a_{i3} + \sum_{k=1}^{K} w_{ki} Q(||\mathbf{c}_k - \mathbf{x}||) \ , \tag{6.18}$$

where $Q : \mathbb{R} \to \mathbb{R}$ is the *radial basis function*

$$Q(r) = r^2 \log r^2 \ .$$

Note that parameters include 6 global affine parameters a_{ij} and $2K$ local coefficients w_{ki} for the control points. In classical correspondence-based approaches control points are placed in extracted point matches, *i.e.* we know the exact mapping at the control points and mappings of other points are *interpolated* using TPS. In our approach, however, TPS can be regarded as a parametric model to *approximate* the underlying free-form deformation. The parameters of this model are then estimated by our method. In order to capture deformations everywhere, we place the radial basis functions (*i.e.* control points) on a uniform grid. Obviously, a finer grid allows to recover finer details of the deformation field at the price of more equations.

The physical interpretation of Eq. (6.18) is a thin plate deforming under point loads acting in the control points. Additional constraints are that the sum of the loads applied to the plate as well as moments with respect to both axes should be 0. These are needed to ensure that the plate would not move or rotate under the imposition of the loads, thus remaining stationery [202]:

$$\sum_{k=1}^{K} w_{ki} = 0 \quad \text{and} \quad \sum_{k=1}^{K} c_{kj} w_{ki} = 0, \quad i, j = 1, 2 \ . \tag{6.19}$$

Another interpretation of the above constraints is that the plate at infinity behaves according to the affine term. Let $\varsigma : \mathbb{R}^2 \to \mathbb{R}^2, \varsigma(\mathbf{x}) = [\varsigma_1(\mathbf{x}), \varsigma_2(\mathbf{x})]^T$ a TPS map with $6 + 2K$ parameters. The Jacobian $|J_\varsigma(\mathbf{x})|$ of the transformation ς is composed of the following partial derivatives ($i, j = 1, 2$)

$$\frac{\partial \varsigma_i}{\partial x_j} = a_{ij} - \sum_{k=1}^{K} 2 w_{ki} (c_{kj} - x_j)(1 + \log(||\mathbf{c}_k - \mathbf{x}||^2)) \ . \tag{6.20}$$

6.5 Experimental results

The proposed method ha been tested on various synthetic and real datasets. The performance of the algorithm has also been compared to two other nonlinear registration methods: *Shape*

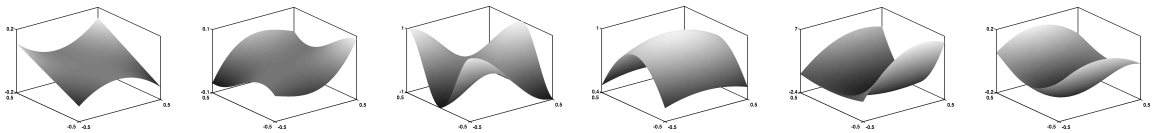
Context [60] which has been developed for general nonlinear registration and *homest* [141] which implements a classical algorithm for homography estimation. The proposed algorithm has been implemented in Matlab R2008 and all tests have been ran on a Pentium IV 3.2 GHz under Linux operating system. The demo implementation of our method is available for download at `http://www.inf.u-szeged.hu/~kato/software/`.

Registration results were quantitatively evaluated using two kind of error measures. The first one (δ) is the absolute difference of the registered shapes, while ϵ measures the distance between the true φ and the estimated $\hat{\varphi}$ transformation:

$$\delta = \frac{|F_r \triangle F_o|}{|F_r| + |F_o|} \cdot 100\%, \quad \epsilon = \frac{1}{|F_t|} \sum_{\mathbf{x} \in F_t} \|\varphi(\mathbf{x}) - \hat{\varphi}(\mathbf{x})\|,$$

where F_t, F_o, and F_r denote the set of foreground pixels of the *template*, *observation*, and the *registered* template respectively.

Intuitively, ϵ shows the average transformation error per pixel. Note that this measure can only be evaluated on synthetic images where the applied transformation is known while δ can always be computed. On the other hand, ϵ gives a better characterization of the transformation error as it directly evaluates the mistransformation. δ sees only the percentage of non-overlapping area between the *observation* and *registered* shape. Hence the value of δ depends also on the compactness, topology, and segmentation error of the shapes.



Eq. (6.21) Eq. (6.22) Eq. (6.23) Eq. (6.24) Eq. (6.25) Eq. (6.26)

Figure 6.2: *Plots of tested $\{\omega_i\}$ function sets.*

6.5.1 Comparison of various ω functions

According to our theoretical results presented in Section 6.3, we expect that the precision of the recovered transformation parameters is independent of the choice of the $\{\omega_i\}$ set as long as equations are properly normalized. To verify these findings, we evaluated the registration quality of various $\{\omega_i\}$ function sets. We considered power and trigonometric functions as well as polynomials, a total of 6 different sets (see Fig. 6.2):

1. Power functions

$$\omega_i(\mathbf{x}) = x_1^{n_i} x_2^{m_i} \tag{6.21}$$

with $(n_i, m_i) \in \{(0,0), (1,0), (0,1), (1,1), (2,0), (0,2), (2,1), (1,2), (2,2), (3,0), (0,3), (3,1), (1,3)\}$

2. Rotated power functions

$$\omega_i(\mathbf{x}) = (x_1 \cos\alpha_i - x_2 \sin\alpha_i)^{n_i}(x_1 \sin\alpha_i + x_2 \cos\alpha_i)^{m_i} \tag{6.22}$$

with $\alpha_i \in \left\{0, \frac{\pi}{6}, \frac{\pi}{3}\right\}$ and $(n_i, m_i) \in \{(1,2), (2,1), (1,3), (3,1)\}$

3. Mixed trigonometric functions

$$\omega_i(\mathbf{x}) = \sin(n_i x_1 \pi)\cos(m_i x_2 \pi) \tag{6.23}$$

with $(n_i, m_i) \in \{(1,2), (2,1), (2,2), (1,3), (3,1), (2,3), (3,2), (3,3), (1,4), (4,1), (2,4), (4,2)\}$

4. Trigonometric functions

$$\omega_i(\mathbf{x}) = Q_i(n_i x_1) R_i(m_i x_2) \tag{6.24}$$

with $Q_i(x), R_i(x) \in \{\sin(x), \cos(x)\}$ and $(n_i, m_i) \in \{(1,1), (1,2), (2,1)\}$

5. Polynomials

$$\omega_i(\mathbf{x}) = P_{n_i}(x_1) P_{m_i}(x_2) \tag{6.25}$$

with $(n_i, m_i) \in \{(1,2), (2,1), (1,3), (3,1), (2,3), (3,2), (1,4), (4,1), (2,4), (4,2), (3,4), (4,3)\}$ composed of the following random polynomials:

$$\begin{aligned} P_1(x) &= 2x^2 - x - 1 \\ P_2(x) &= 2x^3 - x^2 \\ P_3(x) &= x^3 - 30x^2 + 3x + 2 \\ P_4(x) &= 3x^5 - x^2 + 5x - 1 \end{aligned}$$

6. Polynomials

$$\omega_i(\mathbf{x}) = L_{n_i}(x_1) L_{m_i}(x_2) \tag{6.26}$$

with $(n_i, m_i) \in \{(2,3), (3,2), (2,4), (4,2), (3,4), (4,3), (2,5), (5,2), (3,5), (5,3), (4,5), (5,4)\}$ composed of the following Legendre polynomials:

$$\begin{aligned} L_2(x) &= \frac{1}{2}\left(3x^2 - 1\right) \\ L_3(x) &= \frac{1}{2}\left(5x^3 - 3x\right) \\ L_4(x) &= \frac{1}{8}\left(35x^4 - 30x^2 + 3\right) \\ L_5(x) &= \frac{1}{8}\left(63x^5 - 70x^3 + 15x\right) \end{aligned}$$

The quantitative evaluation of the above function sets are summarized in Table 6.1. Basically, all median δ error measures are between $0.1 - 0.2$. Although the mean values have a slightly bigger variance, this is mainly caused by a few outliers rather than a systematic error. It is thus fair to say that the considered ω_i functions perform equally well, which confirms our theoretical results.

The question is therefore naturally arising: Which $\{\omega_i\}$ set should be used? Or in more general: What properties should the $\{\omega_i\}$ set have? From a theoretical point of view, there are only trivial restrictions on the applied functions: Obviously, ω_i must be an integrable function over the finite domains $\mathcal{F}_o$ and $\mathcal{F}_t$. The functions have to be rich enough, *i.e.* they have to produce a varying surface over the shape domain (*e.g.* see Fig. 6.2). For example, the constant function $\omega(x) \equiv c$ is clearly wrong as it makes Eq. (6.8) always true independently of the underlying deformation. From a practical point of view, the picture is different: First of all, we have to solve numerically a system of integral equations. According to Theorem 6.3.1, we can reduce this problem to the solution of a nonlinear system of equations when the ω_i functions are polynomial. The empirical results presented in this section show, that registration quality is almost unaffected by the choice of ω_i functions but computational efficiency is clearly increased for a polynomial $\{\omega_i\}$ set. Therefore we recommend to use low order polynomials for computational efficiency. In our experiments, we have used the set 11), unless otherwise noted.

$\{\omega_i\}$ set	δ(%)			ϵ(pixel)		
	m	μ	σ	m	μ	σ
11)	0.09	0.53	3.38	0.08	3.03	22.36
22)	0.11	1.01	5.01	0.10	4.40	24.14
33)	0.21	12.28	19.61	0.19	20.14	41.73
44)	0.12	1.52	6.25	0.11	6.02	25.79
55)	0.10	0.80	4.75	0.08	3.27	18.60
66)	0.10	0.99	4.84	0.08	4.17	20.78

Table 6.1: *Quantitative comparison of various $\{\omega_i\}$ function sets. m, μ, and σ denote the median, mean, and deviation.*

6.5.2 Quantitative evaluation on synthetic data

Herein, we will focus on planar homography. Synthetic tests with other deformation models can be found in [43]. Our benchmark dataset contains 37 different shapes and their transformed versions, a total of ≈ 1500 images of size 256×256. The applied plane projective transformations were randomly composed of $0.5, \ldots, 1.5$ scalings; $-\frac{\pi}{4}, \ldots, \frac{\pi}{4}$ rotations along the three axes; $-1, \ldots, 1$ translations along both x and y axis and $0.5, \ldots, 2.5$ along the z axis; and a random focal length chosen from $[0.5, 1.5]$. Note that these are projective transformations mapping a *template* shape from a plane placed in the 3D Euclidean space to the xy plane. Some typical examples of these images can be seen in Fig. 6.3, while a summary of

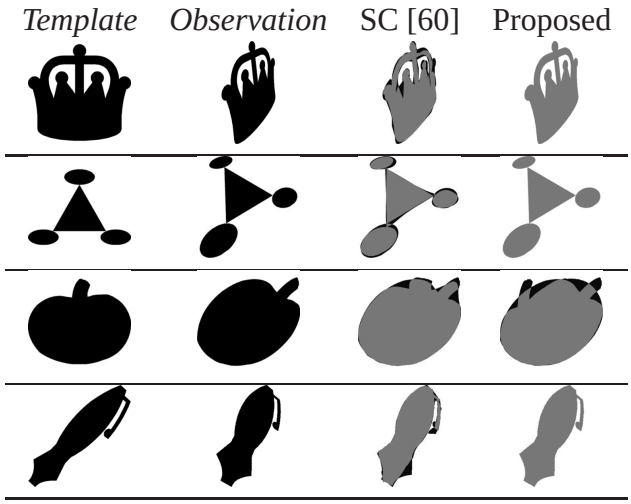


Figure 6.3: Planar homographies: *Example images from the synthetic data set and registration results obtained by* Shape Context *[60] and the proposed method. The* observation *and the registered* template *were overlaid, overlapping pixels are depicted in gray whereas non-overlapping ones are shown in black.*

	Runtime (sec.)		δ (%)		ϵ (pixel)
	SC	P	SC	P	P
m	98.72	16.04	2.69	0.09	0.08
μ	102.78	27.04	4.41	0.54	2.97
σ	28.26	45.34	4.79	3.42	22.04

Table 6.2: *Comparative tests of the proposed method on the synthetic dataset for recovering a* planar homography. *SC* – Shape Context *[60]; P – proposed method.* m*,* μ*, and* σ *denote the median, mean, and deviation.*

registration results is presented in Table 6.2. We have also compared the performance of our method to that of *Shape Context* [60]. For testing, we used the program provided by the authors and set its parameters empirically to their optimal value ($beta\_init = 30$, $n\_iter = 30$, annealing rate $r = 1$). We remark that the program's only output is the registered shape, hence ϵ could not be computed.

6.5.2.1 Robustness

In practice, segmentation never produces perfect shapes. Therefore we have also evaluated the robustness of the proposed approach against segmentation errors. Besides using various kind of real images inherently subject to such errors, we have also conducted a systematic test on simulated data: In the first testcase, $5\%, \ldots, 20\%$ of the foreground pixels has been removed from the *observation* before registration. In the second case, we occluded continu-

(a) missing pixels (b) occlusion (c) disocclusion (d) boundary error

Figure 6.4: *Sample observations with various degradations.*

ous square-shaped regions of size equal to $1\%, \ldots, 10\%$ of the shape, while in the third case we disoccluded a similar region. Finally, we randomly added or removed squares uniformly around the boundary of a total size $1\%, \ldots, 10\%$ of the shape. Note that we do not include cases where erroneous foreground regions appear as disconnected regions, because such false regions can be efficiently removed by appropriate morphological filtering. We therefore concentrate on cases where segmentation errors cannot be filtered out. See samples of these errors in Fig. 6.4.

Table 6.3 shows that our method is quite robust whenever errors are uniformly distributed (first and fourth testcases) over the whole shape. However, it becomes less stable in case of larger localized errors, like occlusion and disocclusion. This is a usual behavior of area-based methods because they are relying on quantities obtained by integrating over the whole object area. Thus large missing parts would drastically change these quantities resulting in false registrations. Nevertheless, in many application areas one can take images under controlled conditions which guarantees that observations are not occluded (*e.g.* medical imaging, industrial inspection). Note also that *Shape Context* [60] is consistently outperformed by our method except in the cases of occlusion and disocclusion.

6.6 Applications

Herein, we will demonstrate the relevance of our approach in various application domains using two common models: planar homography and thin plate spline as well as application-specific deformation models.

6.6.1 Matching traffic signs

Nowadays, modern cars include many safety systems. Automatic traffic sign recognition is a major challenge of such intelligent systems, where one of the key tasks is the matching of a projectively distorted sign with a *template*. Herein, we have used classical thresholding and some morphological operations for segmentation but automatic detection/segmentation is also possible [164]. Fig. 6.5 show some registration results. Each image pair was taken from different signboards. The main challenges were strong deformations, segmentation errors and variations in the style of the depicted objects. For example, the *observations* in Fig. 6.5(b) and Fig. 6.5(c) do not contain exactly the same shape as the object on the

(a) missing pixels		5%	10%	15%	20%
Shape Context [60]	m	21.85	24.91	26.38	27.2
	σ	5.97	6.14	6.37	6.56
Proposed method	m	2.98	5.69	8.51	11.57
	σ	4.13	5.23	6.09	6.74
(b) size of occlusion		1%	2.5%	5%	10%
Shape Context [60]	m	3.03	3.55	4.55	6.79
	σ	4.79	4.79	5.09	7.03
Proposed method	m	1.41	3.40	6.19	11.27
	σ	3.49	4.18	5.09	6.6
(c) size of disocclusion		1%	2.5%	5%	10%
Shape Context [60]	m	3.63	4.52	6.25	9.28
	σ	5.19	5.61	6.84	7.78
Proposed method	m	1.93	4.54	8.28	13.62
	σ	4.31	5.13	6.16	7.09
(d) size of boundary error		1%	5%	10%	20%
Shape Context [60]	m	2.86	3.78	4.68	6.92
	σ	4.72	4.83	5.04	5.92
Proposed method	m	0.54	1.67	2.67	4.03
	σ	3.28	3.5	3.9	4.47

Table 6.3: *Median (m) and standard deviation (σ) of δ error (%) vs. various type of segmentation errors as shown in Fig. 6.4.*

template. In particular, the *STOP* sign in Fig. 6.5(c) uses different fonts. In spite of these difficulties, our method was able to recover a quite accurate transformation (the average δ error was $12,66\%$ on these images).

6.6.2 Aligning hip prosthesis X-ray images

In Section 5.7.1, the problem of assessing hip prosthesis loosing has already been addressed. Clinicians assess loosening by inspecting a number of follow-up X-ray images, where a crucial task is the registration of X-ray images as shown in Fig. 6.6. Herein, we show registration results using a planar homography deformation model. Since the X-ray images are always taken in a well defined *standard position* of the patient's leg, this is a good approximation here. Some of these results are presented in Fig. 6.6.

6.6.2.1 Comparison with correspondence-based homography estimation

Since the grayscale versions of the images were available, it was possible to compare our method to a feature-correspondence based solution. For that purpose, we have used *homest* [141],

	Runtime (sec.)			δ (%)		
	m	μ	σ	m	μ	σ
Shape Context [60]	35.02	34.43	7.58	7.86	9.40	4.71
Proposed method	10.00	9.81	1.47	7.66	8.93	4.22

Table 6.4: *Comparative results on* 2000 *image pairs from the MNIST database.* m*,* μ*, and* σ *stand for the median, mean, and standard deviation.*

which implements a kind of "gold standard" algorithm composed of [112, 205]. The point correspondences has been extracted by the *SIFT* [142] method. As input, we provided the masked signboard region for traffic sign matching and the prosthesis region for medical registration. Furthermore, we have also extracted point correspondences established by *Shape Context* [60]. Here, the input was the binary mask itself used for *SIFT* as well as for our method. Although the *SIFT* parameter called `distRatio`, controlling the number of the extracted correspondences, has been manually fine-tuned, we could not get reliable results due to the lack of rich radiometric features. Fig. 6.6 shows two results on X-ray images while on traffic signs (see Fig. 6.5), *SIFT* could not find enough correspondences in about half of the cases. As for *Shape Context*-based correspondences, we got somewhat better alignments (an average δ of 33.47% for the traffic sign images).

6.6.3 Matching handwritten characters

The performance of our method has been evaluated on aligning handwritten digits from the *MNIST* dataset [138]. A standard approach in matching characters is to align the observation (to be recognized) with each of the digit templates, and recognize it as the template with the lowest deformation. A similar approach is used in [60] which can be applied in our case too. Herein, we will concentrate on the alignment of these characters. Since this is a free-form deformation, we used the *thin plate spline* model with 25 control points placed on a regular grid over the input shapes. The model has $2 \cdot 25 + 6 = 56$ parameters. The equations were generated using the function set Eq. (6.21) with parameters $0 \leq m_i, n_i \leq 8$, $m_i + n_i \leq 8$ resulting in an overdetermined system of 81 equations. The experiment consisted of ≈ 2000 test cases, some example images and registration results are shown in Fig. 6.7. Moreover, we also compared our results to Shape Context [60], which also uses a thin plate spline model but control points are placed on corresponding contour points. Comparative results in Table 6.4 show that our method provides slightly better matches within $1/3$ of CPU time.

6.6.4 Fusion of MRI and TRUS prostate images

Countries in Europe and USA have been following prostate cancer screening programs since the last 15 years [52]. A patient with abnormal findings is generally advised for a prostate biopsy to diagnose the benign or malignant lesions. During needle biopsy, the most common

appearance of malignant lesions in Transrectal Ultrasound (TRUS) is hypoechoic. The accuracy of sonographic finding of hypoechoic prostate cancer lesions is typically 43% [73]. In contrast, Magnetic Resonance Imaging (MRI) has a negative predictive value of $80\% - 84\%$ for significant cancer and the accuracy of MRI to diagnose prostate cancer is approximately $72\% - 76\%$ [196]. Therefore, MRI may serve as a triage test for men deemed to be at risk of prostate cancer and may reduce the number of re-biopsies while at the same time provide more useful information for those who are sent for biopsy. Consequently, fusion of pre-biopsy MR images onto interoperative TRUS images might increase the overall biopsy accuracy [122].

An essential part of this procedure is the alignment of the segmented prostate regions in the two modalities. Since the prostate may undergo deformations due to the insertion of the endorectal probe through the rectum during the MR imaging as well as inflation of the endorectal balloon, nonlinear registration is needed for the multimodal alignment. Due to the rather different content of these modalities, radiometric information cannot be used reliably. Fortunately, the segmentation of the prostate is available in both modalities, which is efficiently obtained by an Active Appearance Model [98]. Placing the control points of a TPS on a uniform grid over the prostate shapes, they can be directly aligned without established correspondences. Fig. 6.9 shows some examples of aligned prostate images obtained by this method. In [41], we have improved the generic non-linear registration framework of [12] by establishing a few prostate-specific point correspondences and regularizing the overall deformation. The point correspondences under the influence of which the thin-plate bends are established on the prostate contours by a method based on matching the shape-context ([61]) representations of contour points using Bhattacharyya distance ([155]). The approximation and regularization of the bending energy of the thin-plate splines are added to the set of non-linear TPS equations and are jointly minimized for a solution. Fig. 6.8 shows some registration results on multimodal prostate images.

6.6.5 Elastic registration of 3D lung CT volumes

Lung alignment is a crucial task in lung cancer diagnosis [72]. During the treatment, changes in the tumor size are determined by comparing *follow-up* PET/CT scans which are taken at regular intervals depending on the treatment and the size of the tumor. Due to respiratory motion, the lung is subject to a nonlinear deformation between such *follow-up*s, hence it is hard to automatically find correspondences. A common practice is to determine corresponding regions by hand, but this makes the procedure time consuming and the obtained alignments may not be accurate enough for measuring changes.

Our algorithm has been successfully applied [46, 47] to align 3D lung CT scans. As usual in elastic medical imaging, the adopted parametric model is a 3D Thin plate splines (TPS) [67, 202] $\varsigma : \mathbb{R}^3 \to \mathbb{R}^3$ which can also be decomposed as three coordinate functions $\varsigma(\mathbf{x}) = [\varsigma_1(\mathbf{x}), \varsigma_2(\mathbf{x}), \varsigma_3(\mathbf{x})]^T$. Given a set of control points $c_k \in \mathbb{R}^3$ and associated mapping coefficients $a_{ij}, w_{ki} \in \mathbb{R}$ with $i = 1, \ldots, 3, j = 1, \ldots, 4$ and $k = 1, \ldots, K$, the TPS

functions are

$$\varsigma_i(\mathbf{x}) = a_{i1}x_1 + a_{i2}x_2 + a_{i3}x_3 + a_{i4} + \sum_{k=1}^{K} w_{ki}Q(\|c_k - \mathbf{x}\|) \tag{6.27}$$

where $Q : \mathbb{R} \to \mathbb{R}$ is the radial basis function, which has the following form in 3D [67]:

$$Q(r) = |r|.$$

The number of the necessary parameters are $N = 3(K+4)$ consisting of 12 affine parameters a_{ij} and 3 coefficients w_{ki} for each of the K control points c_k.

As for the prostate registration problem, we also included a bending energy regularization to ensure the proper alignment of the inner structures. Some registration results are presented in Fig. 6.11, where we also show the achieved alignment on grayscale slices of the original lung CT images. For these slices, the original and transformed images were combined as an 8×8 checkerboard pattern.

6.6.6 Industrial inspection

An important step in hose manufacturing for automotive industry is to print various signs on the hose surface in order to facilitate installation (see Fig. 6.10). The quality control of this process involves visual inspection of the printed signs. In an automated inspection system, this can be implemented by comparing images of the printed sign to its *template*, which requires the alignment of the *template* and *observation* shapes. The main challenges are segmentation errors and complex distortions. The physical model of the contact printing procedure is as follows:

1. The stamp (basically a planar template of the sign) is positioned on the hose surface. This can be described by a 2D rotation and scaling $\mathbf{S} : \mathbb{R}^2 \to \mathbb{R}^2$ of the *template*.

2. Then the stamp is pressed onto the surface, modeled as a transformation $\gamma : \mathbb{R}^2 \to \mathbb{R}^3$ that maps the *template*'s plane to a cylinder with radius r:

$$\gamma(\mathbf{x}) = \left[r \sin \frac{x_1}{r}, x_2, -r \cos \frac{x_1}{r}\right]^T.$$

3. Finally, a picture is taken with a camera, which is described by a classical projective transformation $\mathbf{P} : \mathbb{R}^3 \to \mathbb{R}^2$ with an unknown camera matrix.

Thus the transformation

$$\varphi(\mathbf{x}) = (\mathbf{P} \circ \gamma \circ \mathbf{S})(\mathbf{x}) \tag{6.28}$$

acting between a planar *template* and its distorted *observation* has 11 parameters: $\mathbf{S}$ has 3 parameters, γ has one (r), and $\mathbf{P}$ has 7 (six extrinsic parameters and the focal length). The Jacobian $|J_\varphi|$ is straightforward to compute, although yields a lengthy formula that we omit here due to lack of space. Equations were generated by the function set Eq. (6.22) with parameters using all combinations for $\alpha_i \in \left\{0, \frac{\pi}{6}, \frac{\pi}{3}\right\}$ and $(n_i, m_i) \in \{(1,2),(2,1),(1,3),(3,1)\}$ yielding a system of 12 equations. The method has been tested on more than 150 real images and it proved to be efficient in spite of segmentation errors and severe distortions.

Figure 6.5: *Registration results on traffic signs. The* templates *are in the first row, then the results obtained by* SIFT *[142]*+homest *[141] (second row), where the images show point correspondences between the images found by* SIFT *[142] in the third row. The results obtained by* Shape Context *[60]*+homest *[141] (fourth row) and the proposed method in the last row. The contours of the registered images are overlaid.*

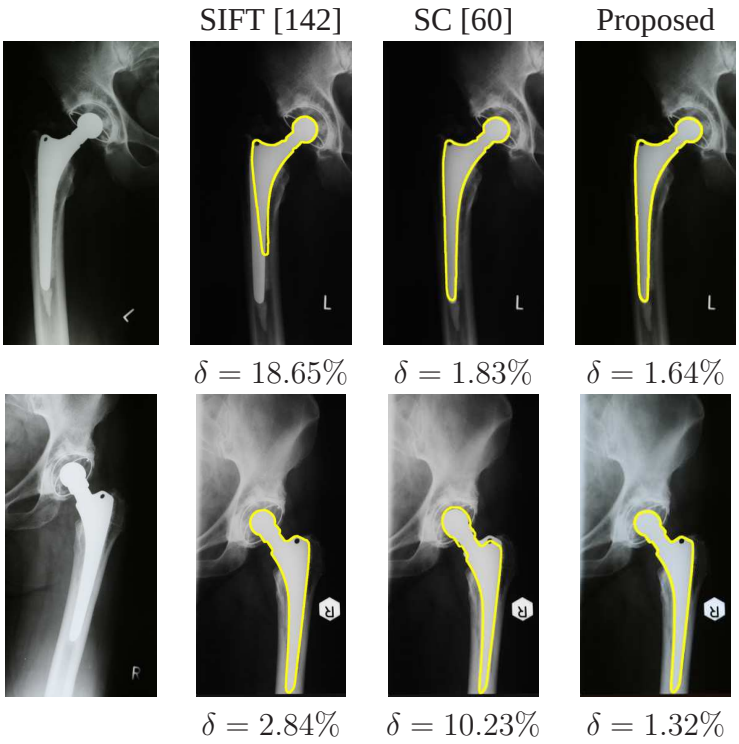


Figure 6.6: *Registration results on hip prosthesis X-ray images. The overlaid contours show the aligned contours of the corresponding images on the left. Images in the second column show the registration results obtained by* SIFT *[142]*+homest *[141], in the third column the results of* Shape Context *[60]*+homest *[141], while the last column contains the results of the proposed method.*

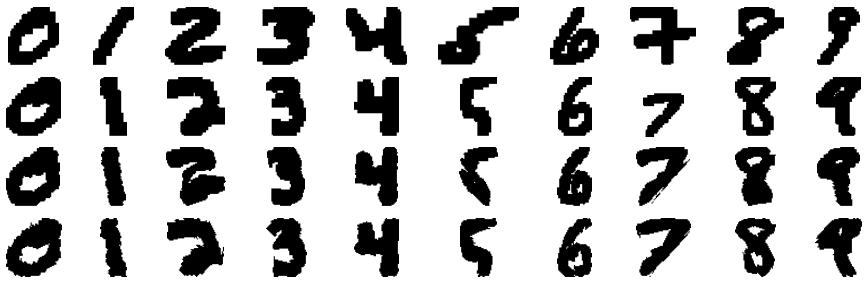


Figure 6.7: *Sample images from the MNIST dataset and registration results using a thin plate spline model. First and second rows show the images used as* templates *and* observations *while the* 3^{rd} *and* 4^{th} *rows show the registration results obtained by Shape Context [60] and the proposed method, respectively.*

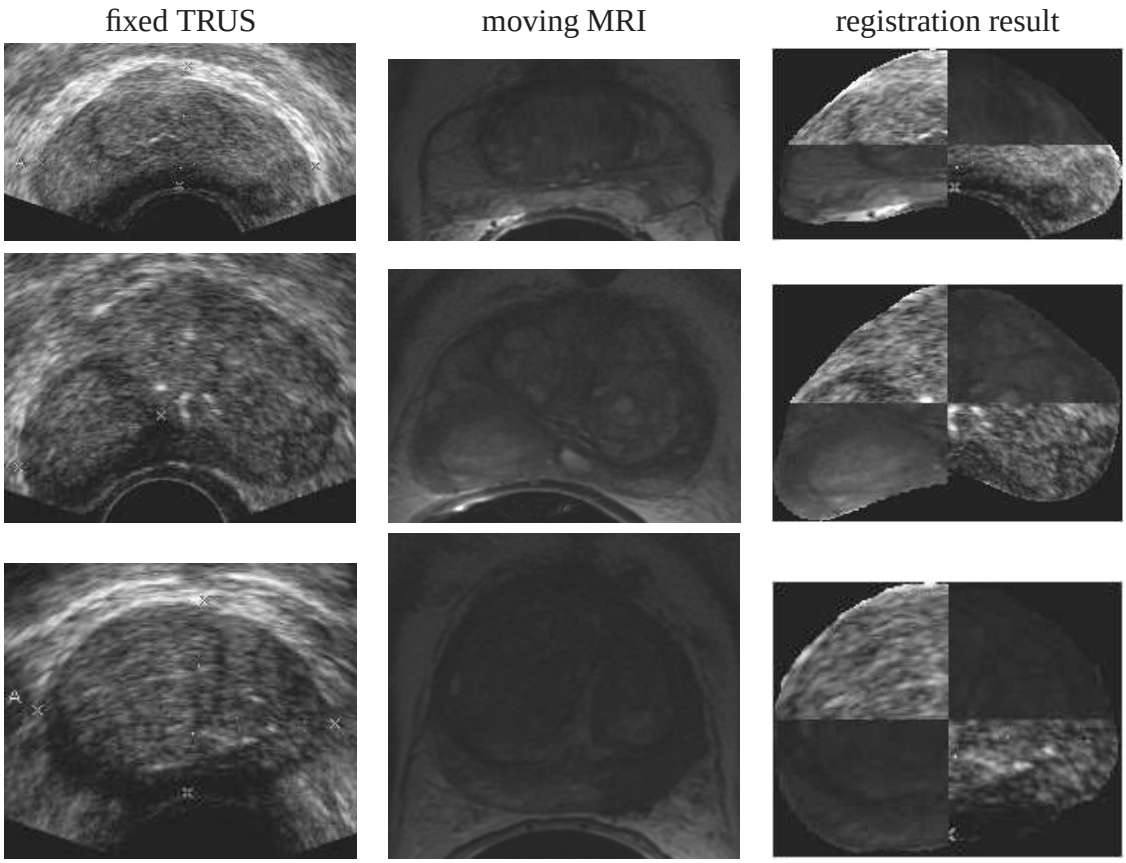


Figure 6.8: *MRI-TRUST multimodal prostate registration results. Registration result is shown as a checkerborard of TRUS and transformed MR images to show the alignment of the inner structures.*

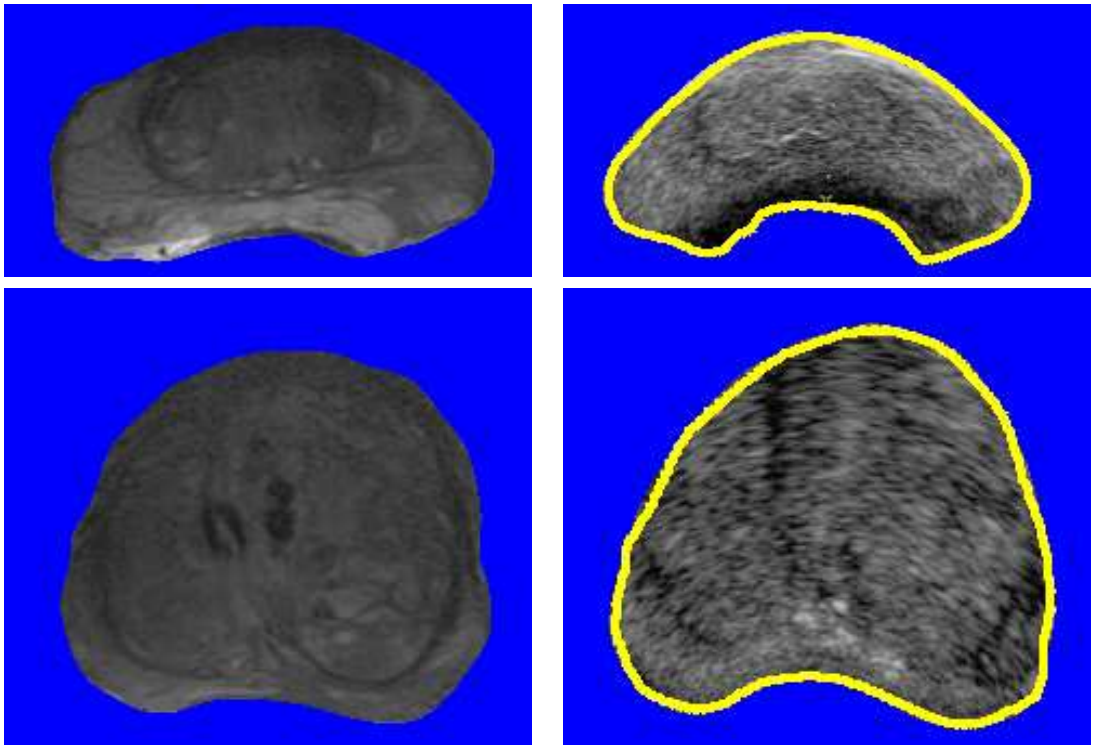


Figure 6.9: *Alignment of MRI (left) and US (right) prostate images using a TPS deformation model. The contours of the registered MRI images are overlaid on the US images.* δ *errors are 2.12% (first row) and 1.88% (second row).*

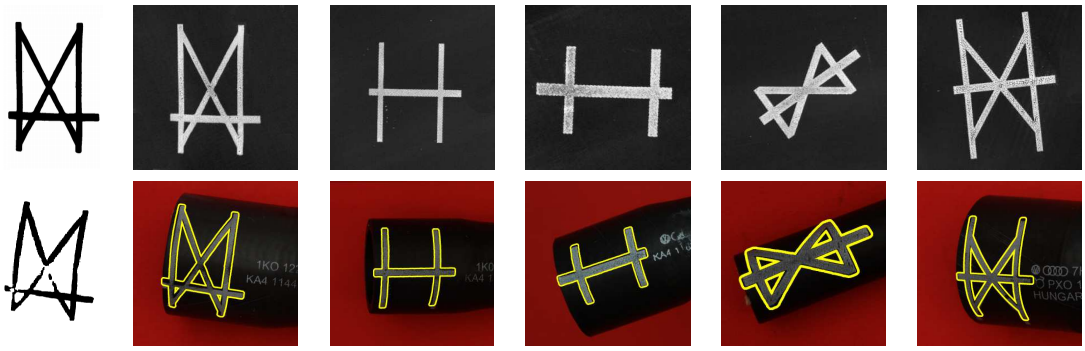


Figure 6.10: *Registration results of printed signs.* **Top:** *planar* templates. **Bottom:** *the corresponding* observations *with the overlaid contour of the registration results. The first image pair shows the segmented regions used for registration. Note the typical segmentation errors. (Images provided by* ContiTech Fluid Automotive Hungária Ltd.*)*

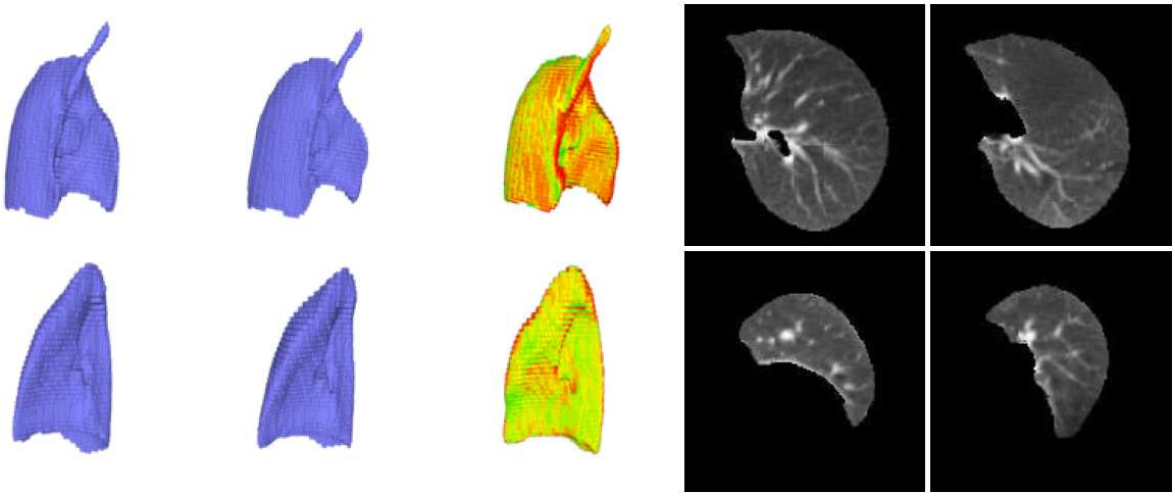


Figure 6.11: *Alignment of lung CT volumes and the combined slices of the original and the transformed images as an 8x8 checkerboard pattern. Segmented 3D lung images were generated by the* InterView Fusion software of Mediso Ltd..

Conclusion

In this dissertation, we have summarized our main contributions to MRF image segmentation and shape alignment:

A novel hierarchical MRF model and its application to supervised and unsupervised satellite image segmentation has been proposed. A new annealing schedule for Simulated Annealing: Multi-temperature annealing allows to assign different temperatures to different cliques during the minimization of the energy of a MRF model. The convergence of the new algorithm has also been proved toward a global optimum.

Probabilistic models for multi-cue segmentation and the 'gas of circles' shape prior. In particular, monogrid and multilayer probabilistic models for color-, texture-, and motion-based segmentation and associated parameter estimation techniques. RJMCMC sampler has been generalized for multi-variate Gaussian mixtures and has been used for fully automatic color image segmentation. The methods have been applied to motion segmentation of video frames and change detection in aerial imagery.

A unified correspondence-less framework for the geometric alignment of 2D and 3D objects. The framework is able to recover a wide range of deformations such as affine, projective, and thin plate splines. Independently of the particular transformation, it relies on the solution of a system on non-linear equations which can be easily constructed by integrating non-linear functions over the shape's domains. Successful applications include various problems in medical image analysis as well as industrial inspection.

Summary of new scientific results

My new scientific results, where my contribution was essential, will be summarized in three thesis points. The first one being my results presented also in my PhD dissertation, while the rest has been achieved after obtaining my PhD degree. Relevant publications and related Chapters of the dissertation are listed at the end of each thesis point.

1. **Multi-resolution and hierarchical Markov models for image segmentation**

   New probabilistic models and optimization methods were developed for supervised and unsupervised gray-level image segmentation. *See Chapter 1*.

   (i) I have proposed a novel hierarchical MRF model and its application to satellite image segmentation. *Related publications*: [14, 20–24, 32].

   (ii) I have developed a new annealing schedule for Simulated Annealing: Multi-temperature annealing allows to assign different temperatures to different cliques during the minimization of the energy of a MRF model. I have proved the convergence of the new algorithm toward a global optimum. *Related publications*: [24, 32].

   (iii) I have shown how to estimate the hierarchical model parameters and applied it to land coverage segmentation on satellite images. *Related publications*: [25, 32, 36–38].

2. **Probabilistic models for multi-cue segmentation and the 'gas of circles' shape prior.**

   Besides gray-levels, there are many cues that one can take as observation for the segmentation process: color, motion, different texture features, etc. Moreover, many application-specific restrictions may apply to the shape of extracted regions. To deal with segmentation problems where coherent regions are defined in terms of such complex features, I have proposed new probabilistic data models and shape priors as well as associated parameter estimation methods.

   (i) One way to combine various features is to design a joint probability distribution which is able to represent the union of the complex observation. I have shown that this approach works well when the combined features are of similar nature (*e.g.* define a multivariate Gaussian density). I have developed a monogrid MRF model which is able to combine color and texture features in order to improve the quality of unsupervised segmentations. I have introduced a novel Reversible Jump Markov Chain Monte Carlo sampling method which is able to identify multi-dimensional Gaussian mixtures. This technique has been successfully applied to fully automatic color image segmentation. *See Chapter 2. Related publications*: [15, 16, 26, 28, 32].

(ii) I have proposed a new multilayer MRF model which is able to segment an image based on multiple cues (*e.g.* color, texture, or motion), which are not necessarily representable as a simple joint distribution. The method has been successfully applied to motion segmentation (a crucial step in *e.g.* MPEG coding) as well as change detection in aerial images. *See Chapter 2. Related publications*: [1, 2, 27, 29–32].

(iii) Higher order active contour (HOAC) models integrate shape knowledge via the inclusion of explicit long-range dependencies between region boundary points. It is possible to set the parameters of the HOAC model to favor regions consisting of any number of approximately circular connected components with some specified radius. This yields the 'gas of circles' HOACs. Starting from the equivalent phase field formulation of the model, I have developed a probabilistic Markov model: the 'gas of circles' MRF. The proposed methods has been successfully applied to extract tree crowns in aerial images for forestry resource management. *See Chapter 3. Related publications*: [5, 13].

(iv) In biomedical image interpretation, a major limitation of the 'gas of circles' model is that touching or overlapping objects cannot be represented. To overcomes these limitations, I have proposed an alternative representation while maintaining computational efficiency: the multi-layer 'gas of circles' model. Both continuous phase-field and discrete MRF models have been developed and successfully applied to various segmentation tasks in microscope image analysis. *See Chapter 4. Related publications*: [42, 45]

3. **Correspondence-less alignment of 2D and 3D visual objects**.

I have proposed a general framework for recovering diffeomorphic deformations of 2D and 3D shapes. The fundamental difference compared to classical image registration algorithms is that this model works without any landmark, feature detection, or correspondences by adopting a novel idea where the transformation is obtained as a solution of a system of non-linear equations.

(i) I have developed a generic framework for recovering linear deformations of 2D and 3D binary objects without correspondences. The basic idea is to set up a system of nonlinear equations whose solution directly provides the parameters of the aligning transformation. Each equation is generated by integrating a nonlinear function over the object's domains. Thus the number of equations is determined by the number of adopted nonlinear functions yielding a flexible mechanism to generate sufficiently many equations. I have shown that power functions always yield a polynomial system. I have given an alternative formulation of the method yielding a linear system of equations constructed by fitting Gaussian densities to the shapes which preserve the effect of the unknown transformation. The method has many applications in medical image analysis. *See Chapter 5. Related publications*: [6–11, 17, 49–51].

(ii) I have developed a substantial extension of the affine registration framework to solve the estimation of a broad range of nonlinear diffeomorphic transformations without establishing correspondences or restricting the strength of the deformation. In particular, I have explicitly shown how to construct a system of equations to recover deformations like planar homography, polynomial and thin plate splines, but other diffeomorphic transformations are also relatively easy to adopt. I have formulated a theorem stating that using power functions and a parametric transformation model in the form of a linear combination of some basis functions, then the resulting system consists of plain non-linear equations. Using the proposed method, numerous registration problems have been solved in many important application areas ranging from medical image analysis to industrial inspection. *See Chapter 6. Related publications*: [12, 18, 19, 39–41, 43, 44, 46–48]

Demo implementations of some of my methods are also available from `http://www.inf.u-szeged.hu/~kato/software/` as follows:

- *Supervised Image Segmentation Using Markov Random Fields* : This is the sample implementation of a Markov random field based supervised image segmentation algorithm for simple gray-level imagery.
- *Supervised Color Image Segmentation in a Markovian Framework*: Implementation of a supervised Markov random field based color image segmentation algorithm.
- *Affine Registration of Planar Shapes*: JAVA code with a direct solver (only runs under Windows).
- *Affine Registration of 3D Objects*: JAVA code with multi-threading (≈ 0.2sec. CPU time for megavoxel volumes).
- *Nonlinear Shape Registration without Correspondences*: Implements planar homography, extension to other nonlinear deformations is relatively easy.

IN THIS CHAPTER:

A.

Proof of theorems

The appendix contains the technical details of the proofs of various theoretical contributions appearing in this dissertation. The first such result is the convergence of the multi-temperature annealing algorithm (Theorem 1.2.1). The next two results related to affine and elastic registrations are Theorem 5.3.1 and Theorem 6.3.1, which state the conditions for reducing the system of integral equations to a system of plain polynomial equations.

A.1 Proof of the multi-temperature annealing theorem

We follow the proof of the annealing theorem given by Geman and Geman in [97]. Essentially, we can apply the same proof, only a slight modification is needed [24].

A.1.1 Notations

We recall a few notations: $\mathcal{S} = \{s_1, \ldots, s_N\}$ denotes the set of sites, $\Lambda = \{0, 1, \ldots, L-1\}$ is a common state space and $\omega, \eta, \eta' \ldots \in \Omega$ denote configurations, where $\Omega = \Lambda^N$ is finite. The sites are updated in the order $\{n_1, n_2, \ldots\} \subset \mathcal{S}$. The generated configurations constitute an inhomogeneous Markov chain $\{X(k), k = 0, 1, 2, \ldots\}$, where $X(0)$ is the initial configuration. The transition $X(k-1) \to X(k)$ is controlled by the Gibbs distribution $\pi_{T(k,C)}$ according to the transition matrix at time k:

$$P_{\omega,\eta}(k) == \begin{cases} \pi_{T(k,C)}(X_{n_k} = \eta_{n_k} \mid X_s = \eta_s, s \neq n_k), & \text{if } \eta = \omega|_{\omega_{n_k}=\lambda} \text{ for some } \lambda \in \Lambda \\ 0, & \text{otherwise} \end{cases} \tag{A.1}$$

$\pi_{T(k,C)}(\omega)$ denotes the Gibbs distribution at iteration k

$$\pi_{T(k,C)}(\omega) = \frac{\exp(-U(\omega) \oslash T(k,C))}{Z} \tag{A.2}$$

$$\text{with } U(\omega) \oslash T(k,C) = \sum_{C \in \mathcal{C}} \frac{V_C(\omega)}{T(k,C)}\,. \tag{A.3}$$

The local characteristics of the above distribution are denoted by:

$$\pi_{T(k,C)}(X_s = \omega_s \mid X_r = \omega_r, s \neq r) = \frac{1}{Z_s} \exp\left(- \sum_{C \in \mathcal{C}: s \in C} \frac{V_C(\omega)}{T(k,C)} \right) \tag{A.4}$$

$$\text{with } Z_s = \sum_{\lambda \in \Lambda} \exp\left(- \sum_{C \in \mathcal{C}: s \in C} \frac{V_C(\omega|_{\omega_s=\lambda})}{T(k,C)} \right) \tag{A.5}$$

The decomposition of $U(\omega) - U(\eta)$ for arbitrary ω and η, $\omega \neq \eta$ is given by:

$$U(\omega) - U(\eta) = \sum_{C \in \mathcal{C}} (V_C(\omega) - V_C(\eta)). \tag{A.6}$$

Denoting respectively by $\Sigma^+(\omega, \eta)$ and $\Sigma^-(\omega, \eta)$ the sum over the positive and negative cliques, we get:

$$\begin{aligned} &\sum_{C \in \mathcal{C}} (V_C(\omega) - V_C(\eta)) \\ &= \underbrace{\sum_{C \in \mathcal{C}: (V_C(\omega) - V_C(\eta)) < 0} (V_C(\omega) - V_C(\eta))}_{\Sigma^-(\omega,\eta)} + \underbrace{\sum_{C \in \mathcal{C}: (V_C(\omega) - V_C(\eta)) \geq 0} (V_C(\omega) - V_C(\eta))}_{\Sigma^+(\omega,\eta)}. \end{aligned} \tag{A.7}$$

Furthermore, let

$$U^{sup} = \max_{\omega\in\Omega} U(\omega), \tag{A.8}$$

$$U^{inf} = \min_{\omega\in\Omega} U(\omega), \tag{A.9}$$

$$\text{and } \Delta = U^{sup} - U^{inf}. \tag{A.10}$$

and define Σ^+_Δ as the minimum of positive sums:

$$\Sigma^+_\Delta = \min_{\substack{\omega'\in\Omega_{sup}\\ \omega''\in\Omega_{opt}}} \Sigma^+(\omega',\omega''). \tag{A.11}$$

Obviously $\Delta \le \Sigma^+_\Delta$.

Given any starting distribution μ_0, the distribution of $X(k)$ is given by the vector $\mu_0\prod_{i=1}^k P(i)$:

$$P_{\mu_0}(X(k)=\omega) = \left(\mu_0\prod_{i=1}^k P(i)\right)\Bigg|_\omega \tag{A.12}$$

$$= \sum_\eta P(X(k)=\omega|X(0)=\eta)\mu_0(\eta) \tag{A.13}$$

We use the following notation for transitions: $\forall l<k$ and $\omega,\eta\in\Omega$:

$$P(k,\omega|l,\eta) = P(X(k)=\omega|X(l)=\eta),$$

and for any distribution μ on Ω:

$$P(k,\omega|l,\mu) = \sum_\eta P(X(k)=\omega|X(l)=\eta)\mu(\eta).$$

Sometimes, we use this notation as $P(k,\cdot|l,\mu)$, where "$\cdot$" means *any* configuration from Ω. Finally, let $\|\mu-\nu\|$ denotes the following distance between two distributions on Ω:

$$\|\mu-\nu\| = \sum_\omega |\mu(\omega)-\nu(\omega)|\,.$$

It is clear, that $\lim_{n\to\infty}\mu_n = \mu$ in distribution (i.e. $\forall\omega:\ \mu_n(\omega)\to\mu(\omega)$) if and only if $\|\mu_n-\mu\|\to 0$.

A.1.2 Proof of the theorem

First, we state two lemmas which imply Theorem 1.2.1:

Lemma A.1.1 *For every $k_0 = 0, 1, 2 \ldots$:*

$$\lim_{k\to\infty} \sup_{\omega,\eta',\eta''} |P(X(k) = \omega | X(k_0) = \eta') - P(X(k) = \omega | X(k_0) = \eta'')| = 0. \qquad \text{(A.14)}$$

Proof:
Fix $k_0 = 0, 1, 2, \ldots$, define $K_l = k_0 + l\kappa$, $l = 0, 1, 2, \ldots$, where κ is the number of transitions necessary for a full sweep of $\mathcal{S}$ (for every $k = 0, 1, 2, \ldots$: $\mathcal{S} \subseteq \{n_{k+1}, n_{k+2}, \ldots, n_{k+\kappa}\}$). Let $\delta(k)$ be the smallest probability among the local characteristics:

$$\delta(k) = \inf_{\substack{1\le i\le N \\ \omega\in\Omega}} \pi_{T(k,C)}(X_{s_i} = \omega_{s_i} | X_{s_j} = \omega_{s_j}, j \neq i).$$

A lower bound for $\delta(k)$ is given by:

$$\delta(k) \geq \frac{\exp(-U^{sup} \oslash T(k,C))}{L \exp(-U^{inf} \oslash T(k,C))} = \frac{\exp(-\Delta \oslash T(k,C))}{L} \geq \frac{1}{L} \exp(-\Sigma_\Delta^+ \oslash T(k,C)$$

$$\geq \frac{1}{L} \exp(-\Sigma_\Delta^+ / T_k^{inf}) \ ,$$

where $L = \mid \Lambda \mid$ is the number of possible states at a site. Now fix l and define m_i as the time of the last replacement of site s_i before $K_l + 1$ (that is before the l^{th} full sweep):

$$\forall i\colon 1 \leq i \leq N : m_i = \sup\{k : k \leq K_l, n_k = s_i\}.$$

Without loss of generality, we can assume that $m_1 > m_2 \cdots > m_N$ (otherwise relabel the sites). Then:

$$P(X(K_l) = \omega | X(K_{l-1}) = \omega')$$
$$= P(X_{s_1}(m_1) = \omega_{s_1}, X_{s_2}(m_2) = \omega_{s_2}, \ldots, X_{s_N}(m_N) = \omega_{s_N} | X(K_{l-1}) = \omega')$$
$$= \prod_{i=1}^{N-1} P(X_{s_i}(m_i) = \omega_{s_i} | X_{s_{i+1}}(m_{i+1}) = \omega_{s_{i+1}}, \ldots, X_{s_N}(m_N) = \omega_{s_N}, X(K_{l-1}) = \omega')$$
$$\geq \prod_{i=1}^{N} \delta(m_i) \geq L^{-N} \prod_{i=1}^{N} \exp(-\Delta / T_{m_i}^{inf}) \geq L^{-N} \exp\left(-\frac{\Sigma_\Delta^+ N}{T_{k_0+l\kappa}^{inf}}\right) \qquad \text{(A.15)}$$

since $m_i \leq K_l = k_0 + l\kappa, i = 1, 2 \ldots, N$ and T_k^{inf} is decreasing. If $k_0 + l\kappa$ is sufficiently large then $T_{k_0+l\kappa}^{inf} \geq N\Sigma_\Delta^+ / \ln(k_0 + l\kappa)$ according to condition 1.2.1/ii and Eq. (A.15) can be continued as:

$$P(X(K_l) = \omega | X(K_{l-1}) = \omega') \geq L^{-N} \exp\left(-\frac{\Sigma_\Delta^+ N}{N\Sigma_\Delta^+ / \ln(k_0 + l\kappa)}\right) = L^{-N}(k_0 + l\kappa)^{-1}.$$

Hence, for a sufficiently small constant Γ ($0 < \Gamma \leq 1$), we can assume that

$$\inf_{\omega,\omega'} P(X(K_l) = \omega | X(K_{l-1}) = \omega') \geq \frac{\Gamma L^{-N}}{k_0 + l\kappa} \qquad \text{(A.16)}$$

for every $k_0 = 0, 1, 2, \ldots$ and $l = 1, 2, \ldots$, keeping in mind that K_l depends on k_0.

Consider now the limit given in Eq. (A.14) and for each $k > k_0$, define $K^{sup}(k) = \sup\{l : K_l < k\}$ (the last sweep before the k^{th} transition) so that $\lim_{k\to\infty} K^{sup}(k) = \infty$. Fix $k > K_1$:

$$\begin{aligned}
&\sup_{\omega,\eta',\eta''} |P(X(k) = \omega|X(0) = \eta') - P(X(k) = \omega|X(0) = \eta'')| \\
&= \sup_{\omega} \left(\sup_{\eta} P(X(k) = \omega|X(0) = \eta) - \inf_{\eta} P(X(k) = \omega|X(0) = \eta) \right) \\
&= \sup_{\omega} \left(\sup_{\eta} \sum_{\omega'} P(X(k) = \omega|X(K_1) = \omega') P(X(K_1) = \omega'|X(0) = \eta) \right. \\
&\quad \left. - \inf_{\eta} \sum_{\omega'} P(X(k) = \omega|X(K_1) = \omega') P(X(K_1) = \omega'|X(0) = \eta) \right) \\
&\doteq \sup_{\omega} Q(k, \omega).
\end{aligned}$$

Furthermore, for each $\omega \in \Omega$:

$$\begin{aligned}
&\sup_{\eta} \sum_{\omega'} P(X(k) = \omega|X(K_1) = \omega') P(X(K_1) = \omega'|X(0) = \eta) \\
&\leq \sup_{\mu} \sum_{\omega'} P(X(k) = \omega|X(K_1) = \omega') \mu(\omega'),
\end{aligned}$$

where μ is any probability measure on Ω. Using Eq. (A.16), we get:

$$\mu(\omega') \geq \frac{\Gamma L^{-N}}{k_0 + l\kappa}.$$

Suppose that $P(X(k) = \omega|X(K_1) = \omega')$ is maximized at $\omega' = \omega^{sup}$ and minimized at $\omega' = \omega^{inf}$. Then we get:

$$\begin{aligned}
&\sup_{\mu} \sum_{\omega'} P(X(k) = \omega|X(K_1) = \omega') \mu(\omega') \leq \\
&\left(1 - (L^N - 1)\frac{\Gamma L^{-N}}{k_0 + l\kappa}\right) P(X(k) = \omega|X(K_1) = \omega^{sup}) \\
&+ \frac{\Gamma L^{-N}}{k_0 + l\kappa} \underbrace{\sum_{\omega' \neq \omega^{sup}} P(X(k) = \omega|X(K_1) = \omega')}_{P(X(k)=\omega|X(K_1)=\omega^{inf}) + \sum_{\omega' \neq \omega^{sup}, \omega^{inf}} P(X(k)=\omega|X(K_1)=\omega')},
\end{aligned}$$

and in a similar way:

$$\inf_{\mu} \sum_{\omega'} P(X(k) = \omega|X(K_1) = \omega') \mu(\omega') \geq$$

$$\left(1-(L^N-1)\frac{\Gamma L^{-N}}{k_0+l\kappa}\right)P(X(k)=\omega|X(K_1)=\omega^{inf})$$

$$+\frac{\Gamma L^{-N}}{k_0+l\kappa}\underbrace{\sum_{\omega'\neq\omega^{inf}}P(X(k)=\omega|X(K_1)=\omega')}_{P(X(k)=\omega|X(K_1)=\omega^{sup})+\sum_{\omega'\neq\omega^{sup},\omega^{inf}}P(X(k)=\omega|X(K_1)=\omega')}.$$

Then, it is clear that

$$Q(k,\omega)\leq\left(1-\frac{\Gamma}{k_0+l\kappa}\right)\left(P(X(k)=\omega|X(K_1)=\omega^{sup})-P(X(k)=\omega|X(K_1)=\omega^{inf})\right),$$

hence:

$$\sup_{\omega,\eta',\eta''}|P(X(k)=\omega|X(0)=\eta')-P(X(k)=\omega|X(0)=\eta'')|\leq$$

$$\left(1-\frac{\Gamma}{k_0+l\kappa}\right)\sup_{\omega,\eta',\eta''}|P(X(k)=\omega|X(K_1)=\eta')-P(X(k)=\omega|X(K_1)=\eta'')|\leq$$

$$\left(1-\frac{\Gamma}{k_0+l\kappa}\right)\left(\left(1-\frac{\Gamma}{k_0+l\kappa}\right)\sup_{\omega,\eta',\eta''}\left|P(X(k)=\omega|X(K_2)=\eta')-P(X(k)=\omega|X(K_2)=\eta'')\right|\right)$$

Proceeding this way, we have the following bound:

$$\leq\prod_{k=1}^{K^{sup}(k)}\left(1-\frac{\Gamma}{k_0+l\kappa}\right)\sup_{\omega,\eta',\eta''}\left|P(X(k)=\omega|X(K_{K^{sup}(k)})=\eta')\right.$$

$$\left.-P(X(k)=\omega|X(K_{K^{sup}(k)})=\eta'')\right|$$

and finally, since the the possible maximal value of the supremum is 1:

$$\sup_{\omega,\eta',\eta''}|P(X(k)=\omega|X(0)=\eta')-P(X(k)=\omega|X(0)=\eta'')|\leq\prod_{k=1}^{K^{sup}(k)}\left(1-\frac{\Gamma}{k_0+l\kappa}\right).$$

It is then sufficient to show that

$$\lim_{m\to\infty}\prod_{k=1}^{m}\left(1-\frac{\Gamma}{k_0+l\kappa}\right)=0.$$

which is a well known consequence of the divergence of the series

$$\sum_l(k_0+l\kappa)^{-1}$$

for all k_0 and κ. This completes the proof. **Q.E.D.**

Lemma A.1.2

$$\lim_{k_0\to\infty}\sup_{k\geq k_0}\parallel P(k,\cdot|k_0,\pi_0)-\pi_0\parallel=0. \tag{A.17}$$

Proof:
In the following, let $P_{k_0,k}(\cdot)$ stand for $P(k,\cdot|k_0,\pi_0)$, so that for any $k\geq k_0>0$:

$$P_{k_0,k}(\omega)=\sum_{\eta}P(X(k)=\omega|X(k_0)=\eta)\pi_0(\eta).$$

First, we show that for any $k>k_0\geq 0$:

$$\|P_{k_0,k}-\pi_{T(k,C)}\|\leq\|P_{k_0,k-1}-\pi_{T(k,C)}\|. \tag{A.18}$$

We can assume for convenience that $n_k=s_1$. Then

$$\|P_{k_0,k}-\pi_{T(k,C)}\|=$$

$$\sum_{(\omega_{s_1},\ldots\omega_{s_N})}\Big|\pi_{T(k,C)}(X_{s_1}=\omega_{s_1}|X_s=\omega_s,s\neq s_1)P_{k_0,k-1}(X_s=\omega_s,s\neq s_1)$$

$$-\pi_{T(k,C)}(X_s=\omega_s,s\in\mathcal{S})\Big|$$

$$=\sum_{(\omega_{s_2},\ldots\omega_{s_N})}\left(\sum_{\omega_{s_1}\in\Lambda}\pi_{T(k,C)}(X_{s_1}=\omega_{s_1}|X_s=\omega_s,s\neq s_1)\,|P_{k_0,k-1}(X_s=\omega_s,s\neq s_1)\right.$$

$$\left.-\pi_{T(k,C)}(X_s=\omega_s,s\neq s_1)\Big|\right)$$

$$=\sum_{(\omega_{s_2},\ldots\omega_{s_N})}\Big|P_{k_0,k-1}(X_s=\omega_s,s\neq s_1)-\pi_{T(k,C)}(X_s=\omega_s,s\neq s_1)\Big|$$

$$=\sum_{(\omega_{s_2},\ldots\omega_{s_N})}\left|\sum_{\omega_{s_1}}(P_{k_0,k-1}(X_s=\omega_s,s\in\mathcal{S})-\pi_{T(k,C)}(X_s=\omega_s,s\in\mathcal{S}))\right|$$

$$\leq\sum_{(\omega_{s_1},\ldots\omega_{s_N})}\Big|P_{k_0,k-1}(X_s=\omega_s,s\in\mathcal{S})-\pi_{T(k,C)}(X_s=\omega_s,s\in\mathcal{S})\Big|$$

$$=\|P_{k_0,k-1}-\pi_{T(k,C)}\|.$$

Second, we prove that $\pi_{T(k,C)}$ converges to π_0 (the uniform distribution on Ω_{opt}):

$$\lim_{k\to\infty}\|\pi_0-\pi_{T(k,C)}\|=0.$$

To see this, let $|\Omega_{opt}|$ be the number of globally optimal configurations. Then

$$\lim_{k\to\infty}\pi_{T(k,C)}(\omega)$$

$$= \lim_{k\to\infty} \frac{\exp(-U(\omega) \oslash T(k,C))}{\sum_{\omega'\in\Omega_{opt}} \exp(-U(\omega') \oslash T(k,C)) + \sum_{\omega'\notin\Omega_{opt}} \exp(-U(\omega') \oslash T(k,C))}$$

$$= \lim_{k\to\infty} \frac{\exp(-(U(\omega) - U^{inf}) \oslash T(k,C))}{|\Omega_{opt}| + \sum_{\omega'\notin\Omega_{opt}} \exp(-(U(\omega) - U^{inf}) \oslash T(k,C))} = \begin{cases} 0 & \omega \notin \Omega_{opt} \\ \frac{1}{|\Omega_{opt}|} & \omega \in \Omega_{opt} \end{cases} \tag{A.19}$$

The above equation is true if $(U(\omega) - U^{inf}) \oslash T(k,C) \geq 0$. Let us rewrite this inequality as

$$\sum_{C\in\mathcal{C}} \frac{V_C(\omega) - V_C(\omega')}{T(k,C)} \geq 0 \tag{A.20}$$

where ω' is any globally optimal configuration (i.e. $\omega' \in \Omega_{opt}$). While $V_C(\omega) - V_C(\omega')$ may be negative, $U(\omega) - U^{inf}$ is always positive or zero. We denote by $\Sigma(\omega)$ the energy difference in Eq. (A.20) without the temperature. Obviously, it is non-negative:

$$\Sigma(\omega) = \sum_{C\in\mathcal{C}} V_C(\omega) - V_C(\omega') = U(\omega) - U^{inf} \geq 0$$

Then, let us decompose $\Sigma(\omega)$ according to Eq. (1.14):

$$\Sigma(\omega) = \Sigma^+(\omega,\omega') + \Sigma^-(\omega,\omega').$$

From which:

$$\Sigma^+(\omega,\omega') = \Sigma(\omega) - \Sigma^-(\omega,\omega').$$

Now, we consider Eq. (A.20):

$$\sum_{C\in\mathcal{C}} \frac{V_C(\omega) - V_C(\omega')}{T(k,C)} = \Sigma^-(\omega,\omega') \oslash T(k,C) + \Sigma^+(\omega,\omega') \oslash T(k,C)$$

$$\geq \Sigma^-(\omega,\omega')/T_k^{inf} + \Sigma^+(\omega,\omega')/T_k^{sup} = \frac{\Sigma^-(\omega,\omega')\cdot T_k^{sup} + \Sigma^+(\omega,\omega')\cdot T_k^{inf}}{T_k^{inf} T_k^{sup}} \geq 0$$

Furthermore:

$$\Sigma^-(\omega,\omega')\cdot T_k^{sup} + \Sigma^+(\omega,\omega')\cdot T_k^{inf} = \Sigma^-(\omega,\omega')\cdot T_k^{sup} + (\Sigma(\omega) - \Sigma^-(\omega,\omega'))T_k^{inf}$$

Therefore:

$$\Sigma^-(\omega,\omega')(T_k^{sup} - T_k^{inf}) - \Sigma(\omega)\cdot T_k^{inf} \geq 0$$

Dividing by $\Sigma^-(\omega,\omega')$ which is negative, we get:

$$T_k^{sup} - T_k^{inf} \leq \frac{\Sigma(\omega)}{\mid \Sigma^-(\omega,\omega') \mid} T_k^{inf}$$

Which is true due to condition 1.2.1/iii of the theorem.

Finally, we can prove that

$$\sum_{k=1}^{\infty} \left\| \pi_{T(k,C)} - \pi_{T(k+1,C)} \right\| < \infty \tag{A.21}$$

since

$$\sum_{k=1}^{\infty} \left\| \pi_{T(k,C)} - \pi_{T(k+1,C)} \right\| = \sum_{\omega} \sum_{k=1}^{\infty} \left| \pi_{T(k,C)}(\omega) - \pi_{T(k+1,C)}(\omega) \right|$$

and since

$$\forall \omega : \pi_{T(k,C)}(\omega) \longrightarrow \pi_0(\omega),$$

it is enough to show that $\pi_T(\omega)$ is monotonous for every ω. However it is clear from Eq. (A.19) that

- if $\omega \notin \Omega_{opt}$ then $\pi_T(\omega)$ is strictly increasing for $0 < T \leq \epsilon$ for some sufficiently small ϵ,
- if $\omega \in \Omega_{opt}$ then $\pi_T(\omega)$ is strictly decreasing for all $T > 0$.

Fix $k > k_0 \geq 0$. From Eq. (A.18) and Eq. (A.21), we obtain:

$$\begin{aligned}
\|P_{k_0,k} - \pi_0\| &\leq \|P_{k_0,k} - \pi_{T(k,C)}\| + \|\pi_{T(k,C)} - \pi_0\| \\
&\leq \|P_{k_0,k-1} - \pi_{T(k,C)}\| + \|\pi_{T(k,C)} - \pi_0\| \text{ by Eq. (A.18)} \\
&\leq \|P_{k_0,k-1} - \pi_{T(k-1,C)}\| + \|\pi_{T(k-1,C)} - \pi_{T(k,C)}\| + \|\pi_{T(k,C)} - \pi_0\| \\
&\leq \|P_{k_0,k-2} - \pi_{T(k-2,C)}\| + \|\pi_{T(k-2,C)} - \pi_{T(k-1,C)}\| + \|\pi_{T(k-1,C)} - \pi_{T(k,C)}\| + \|\pi_{T(k,C)} - \pi_0\| \\
&\leq \cdots \leq \|P_{k_0,k_0} - \pi_{T(k_0,C)}\| + \sum_{l=k_0}^{k-1} \|\pi_{T(l,C)} - \pi_{T(l+1,C)}\| + \|\pi_{T(k,C)} - \pi_0\|.
\end{aligned}$$

On the other hand,

$$P_{k_0,k_0} = \pi_0$$

and

$$\lim_{k \to \infty} \|\pi_{T(k,C)} - \pi_0\| = 0.$$

Then we have

$$\limsup_{k_0 \to \infty} \sup_{k \geq k_0} \|P_{k_0,k} - \pi_0\| \leq \limsup_{k_0 \to \infty} \sup_{k > k_0} \sum_{l=k_0}^{k-1} \|\pi_{T(l,C)} - \pi_{T(l+1,C)}\|$$

$$= \limsup_{k_0 \to \infty} \sum_{l=k_0}^{\infty} \|\pi_{T(l,C)} - \pi_{T(l+1,C)}\| = 0$$

The last term is 0 by (A.21) which completes the proof. **Q.E.D.**

Theorem 1.2.1 (Multi-Temperature Annealing) *Assume that there exists an integer $\kappa \geq N$ such that for every $k = 0, 1, 2, \ldots$, $\mathcal{S} \subseteq \{n_{k+1}, n_{k+2}, \ldots, n_{k+\kappa}\}$. For all $C \in \mathcal{C}$, let $T(k, C)$ be any decreasing sequence of temperatures in k for which*

(i) $\lim_{k\to\infty} T(k, C) = 0$.
Let us denote respectively by T_k^{inf} and T_k^{sup} the maximum and minimum of the temperature function at k ($\forall C \in \mathcal{C}$: $T_k^{inf} \leq T(k, C) \leq T_k^{sup}$).

(ii) For all $k \geq k_0$, for some integer $k_0 \geq 2$: $T_k^{inf} \geq N\Sigma_\Delta^+ / \ln(k)$.

(iii) If $\Sigma^-(\omega, \omega') \neq 0$ for some $\omega \in \Omega \setminus \Omega_{opt}$, $\omega' \in \Omega_{opt}$ then a further condition must be imposed:
For all k: $\frac{T_k^{sup} - T_k^{inf}}{T_k^{inf}} \leq R$ with

$$R = \min_{\substack{\omega \in \Omega \setminus \Omega_{opt} \\ \omega' \in \Omega_{opt} \\ \Sigma^-(\omega,\omega') \neq 0}} \frac{U(\omega) - U^{inf}}{\mid \Sigma^-(\omega, \omega') \mid}$$

Then for any starting configuration $\eta \in \Omega$ and for every $\omega \in \Omega$:

$$\lim_{k\to\infty} P(X(k) = \omega \mid X(0) = \eta) = \pi_0(\omega). \tag{A.22}$$

Proof:
Using the above mentioned lemmas, we can easily prove the annealing theorem:

$$\limsup_{k\to\infty} \|P(X(k) = \cdot | X(0) = \eta) - \pi_0\| = \limsup_{k_0\to\infty} \limsup_{\substack{k\to\infty \\ k\geq k_0}} \| \sum_{\eta'} P(k, \cdot | k_0, \eta') P(k_0, \eta' | 0, \eta) - \pi_0 \|$$

$$\leq \limsup_{k_0\to\infty} \limsup_{\substack{k\to\infty \\ k\geq k_0}} \| \sum_{\eta'} P(k, \cdot | k_0, \eta') P(k_0, \eta' | 0, \eta) - P(k, \cdot | k_0, \pi_0) \|$$

$$+ \limsup_{k_0\to\infty} \limsup_{\substack{k\to\infty \\ k\geq k_0}} \|P(k, \cdot | k_0, \pi_0) - \pi_0\| \ .$$

The last term is 0 by Lemma A.1.2. Moreover, $P(k_0, \cdot | 0, \eta)$ and π_0 have total mass 1, thus:

$$\| \sum_{\eta'} P(k, \cdot | k_0, \eta') P(k_0, \eta' | 0, \eta) - P(k, \cdot | k_0, \pi_0) \|$$

$$= \sum_{\omega} \sup_{\eta''} | \sum_{\eta'} (P(k, \omega | k_0, \eta') - P(k, \omega | k_0, \eta'')) (P(k_0, \eta' | 0, \eta) - \pi_0(\eta')) |$$

$$\leq 2 \sum_{\omega} \sup_{\eta', \eta''} |P(k, \omega | k_0, \eta') - P(k, \omega | k_0, \eta'')| \ .$$

Finally,

$$\limsup_{k \to \infty} \| P(X(k) = \cdot | X(0) = \eta) - \pi_0 \|$$

$$\leq 2 \sum_{\omega} \limsup_{k_0 \to \infty} \limsup_{\substack{k \to \infty \\ k \geq k_0}} \sup_{\eta', \eta''} |P(k, \omega | k_0, \eta') - P(k, \omega | k_0, \eta'')| = 0$$

The last term is 0 by Lemma A.1.1 which completes the proof of the annealing theorem. **Q.E.D.**

A.2 Proof of Theorem 5.3.1

Let $1 \leq k \leq n$ arbitrary and fixed. We assume that $\omega^k(\mathbf{x})$ is polynomial, *i.e.* there exists an n-variate real polynome p_k with $\deg(p_k) \geq 1$, such that

$$\omega^k(\mathbf{x}) = p_k(x_1, \ldots, x_n) = \sum_{i=1}^{u_k} s_i (\mathbf{A}_1^{-1}\mathbf{y})^{\alpha_{i1}} \ldots (\mathbf{A}_n^{-1}\mathbf{y})^{\alpha_{in}}, \tag{A.23}$$

where $u_k = \begin{pmatrix} \deg(p_k) + n \\ \deg(p_k) \end{pmatrix}$, and $\mathbf{A}_j^{-1}$ denotes the jth row of $\mathbf{A}^{-1}$. One term of Eq. (A.23) can be expanded by making use of the *Multinomial theorem* [151]. For a given i and for all $1 \leq j \leq n$, we get

$$(\mathbf{A}_j^{-1}\mathbf{y})^{\alpha_{ij}} = (q_{j1}y_1 + \ldots + q_{jn}y_n + q_{j(n+1)})^{\alpha_{ij}}$$

$$= \sum_{\substack{\beta_{ij1}, \ldots, \beta_{ij(n+1)} \in \mathbb{N}_0 \\ \beta_{ij1} + \ldots + \beta_{ij(n+1)} = \alpha_{ij}}} \frac{\alpha_{ij}!}{\beta_{ij1}! \ldots \beta_{ij(n+1)}!} q_{j1}^{\beta_{ij1}} \ldots q_{jn}^{\beta_{ijn}} q_{j(n+1)}^{\beta_{ij(n+1)}} y_1^{\beta_{ij1}} \ldots y_n^{\beta_{ijn}},$$

hence we get an $(n+1)$-variate real polynome. In fact, we should compute the sum of the product of n pieces of $(n+1)$ - variate polynoms in Eq. (A.23). Let $m = n(n+1)$ and consider these products as m-variate polynoms. Furthermore, the sum of m-variate polynoms is also an m-variate polynome. Integrating and using this observation we can rewrite Eq. (A.23) as

$$\int \sum_{i=1}^{u_k} s_i (\mathbf{A}_1^{-1}\mathbf{y})^{\alpha_{i1}} \ldots (\mathbf{A}_n^{-1}\mathbf{y})^{\alpha_{in}} \equiv \int \sum_{i=1}^{v_k} t_i q_1^{\gamma_{i1}} \ldots q_m^{\gamma_{im}} y_1^{\delta_{i1}} \ldots y_n^{\delta_{in}},$$

where $v_k = \begin{pmatrix} \deg(p_k) + m \\ \deg(p_k) \end{pmatrix}$. It is obvious from the above equation that the system of equation has a degree of up to $\deg(p_k)$. Furthermore, by making use of the basic properties of the Lebesgue integral, we get

$$\int \sum_{i=1}^{v_k} t_i q_1^{\gamma_{i1}} \ldots q_m^{\gamma_{im}} y_1^{\delta_{i1}} \ldots y_n^{\delta_{in}} = \sum_{i=1}^{v_k} \int t_i q_1^{\gamma_{i1}} \ldots q_m^{\gamma_{im}} y_1^{\delta_{i1}} \ldots y_n^{\delta_{in}}$$

$$= \sum_{i=1}^{v_k} t_i q_1^{\gamma_{i1}} \ldots q_m^{\gamma_{im}} \int y_1^{\delta_{i1}} \ldots y_n^{\delta_{in}} = \sum_{i=1}^{v_k} w_i q_1^{\gamma_{i1}} \ldots q_m^{\gamma_{im}}.$$

The last term is indeed a real polynome r_k with variables $q_1, \ldots, q_m$ yielding

$$\int \omega^k(\mathbf{x}) \equiv r_k(q_1, \ldots, q_m).$$

Hence the system of equations is polynomial which completes the proof.

A.3 Proof of Theorem 6.3.1

The statement of Theorem 6.3.1 follows from the next three lemmas [12].

Lemma A.3.1 *If f_1 and f_2 are separable with respect to $(\mathbf{a},\phi(\mathbf{x}))$, then the function $F(\mathbf{x}) = f_1(\mathbf{x})f_2(\mathbf{x})$ is also separable.*

Proof:
Since both f_1 and f_2 are separable, there exist two sets of functions $g_i^{(1)}, g_j^{(2)} : \mathbb{R}^n \to \mathbb{R}$ and $h_i^{(1)}, h_j^{(2)} : \mathbb{R}^{2n} \to \mathbb{R}^2$ for $1 \leq i \leq s$ and $1 \leq j \leq t$ such that

$$\begin{aligned} F(\mathbf{x}) &= \sum_{i=1}^{s} g_i^{(1)}(\mathbf{a}) h_i^{(1)}(\phi(\mathbf{x})) \sum_{j=1}^{t} g_j^{(2)}(\mathbf{a}) h_j^{(2)}(\phi(\mathbf{x})) \\ &= \sum_{i=1}^{s}\sum_{j=1}^{t} g_i^{(1)}(\mathbf{a}) g_j^{(2)}(\mathbf{a}) h_i^{(1)}(\phi(\mathbf{x})) h_j^{(2)}(\phi(\mathbf{x})). \end{aligned}$$

Making use of the notations $g_l = g_i^{(1)} g_j^{(2)}$ and $h_l = h_i^{(1)} h_j^{(2)}$ with $l = (i-1)t + j$, we get

$$F(\mathbf{x}) = \sum_{l=1}^{st} g_l(\mathbf{a}) h_l(\phi(\mathbf{x})),$$

which completes the proof. **Q.E.D.**

Lemma A.3.2 *If $\varphi(\mathbf{x}) = \sum_{i=1}^{n} a_i \phi_i(\mathbf{x})$, then $|J_\varphi(\mathbf{x})|$ is separable with respect to $(\mathbf{a},\phi(\mathbf{x}))$.*

Proof:
Let us denote the components of the basis functions as $\phi_i(\mathbf{x}) = [\phi_{i1}(\mathbf{x}), \phi_{i2}(\mathbf{x})]$. The partial derivatives $\partial_l \varphi_k$ $(k, l = 1, 2)$ of $\varphi(\mathbf{x})$ are then given by

$$\frac{\partial \varphi_k}{\partial x_l} = \sum_{i=1}^{n} a_i \partial_l \phi_{ik}(\mathbf{x}), \quad k, l = 1, 2$$

from which the Jacobian determinant of Eq. (6.4) can be written as

$$\begin{aligned} |J_\varphi(\mathbf{x})| &= \Big(\sum_{i=1}^{n} a_i \partial_1 \phi_{i1}(\mathbf{x})\Big)\Big(\sum_{j=1}^{n} a_j \partial_2 \phi_{j2}(\mathbf{x})\Big) \\ &\quad -\Big(\sum_{i=1}^{n} a_i \partial_2 \phi_{i1}(\mathbf{x})\Big)\Big(\sum_{j=1}^{n} a_j \partial_1 \phi_{j2}(\mathbf{x})\Big) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j (\partial_1 \phi_{i1}(\mathbf{x}) \partial_2 \phi_{j2}(\mathbf{x}) \\ &\quad -\partial_2 \phi_{i1}(\mathbf{x}) \partial_1 \phi_{j2}(\mathbf{x})). \end{aligned}$$

Setting $g_l(\mathbf{a}) = a_i a_j$ and $h_l(\phi(\mathbf{x})) = \partial_1\phi_{i1}(\mathbf{x})\partial_2\phi_{j2}(\mathbf{x}) - \partial_2\phi_{i1}(\mathbf{x})\partial_1\phi_{j2}(\mathbf{x})$ with $l = (i-1)n + j$, we get

$$|J_\varphi(\mathbf{x})| = \sum_{l=1}^{n^2} g_l(\mathbf{a}) h_l(\phi(\mathbf{x})).$$

The Jacobian is thus separable, which completes the proof. **Q.E.D.**

Lemma A.3.3 *If* $\varphi(\mathbf{x}) = \sum_{i=1}^n a_i\phi_i(\mathbf{x})$ *and* $p(\mathbf{x}) = [p_1(x_1, x_2), p_2(x_1, x_2)]$*, where* $p_1, p_2 \in \mathbb{R}[x_1, x_2]$ *are polynomials with* $\deg(p_1) = d_1$ *and* $\deg(p_2) = d_2$*, then* $p(\varphi(\mathbf{x}))$ *is separable with respect to* $(\mathbf{a}$*,*$\phi(\mathbf{x}))$*.*

Proof:
Let us consider the first component p_1 of

$$p\Big(\sum_{i=1}^n a_i\phi_i(\mathbf{x})\Big) = \Big[p_1\Big(\sum_{i=1}^n a_i\phi_i(\mathbf{x})\Big), p_2\Big(\sum_{i=1}^n a_i\phi_i(\mathbf{x})\Big)\Big].$$

Since $p_1(x_1, x_2) = \sum_j c_j x_1^{q_j} x_2^{r_j}$ is polynomial,

$$p_1\Big(\sum_{i=1}^n a_i\phi_{i1}(\mathbf{x}), \sum_{i=1}^n a_i\phi_{i2}(\mathbf{x})\Big) =$$

$$\sum_j c_j\Big(\sum_{i=1}^n a_i\phi_{i1}(\mathbf{x})\Big)^{q_j}\Big(\sum_{i=1}^n a_i\phi_{i2}(\mathbf{x})\Big)^{r_j}.$$

$\Big(\sum_{i=1}^n a_i\phi_{i1}(\mathbf{x})\Big)^{q_j}$ can be further expanded by making use of the *Multinomial theorem* [151] as

$$\sum \frac{q_j!}{s_1!\dots s_n!} a_1^{s_1}\dots a_n^{s_n}\phi_{11}(\mathbf{x})^{s_1}\dots\phi_{n1}(\mathbf{x})^{s_n},$$

where $s_1, \dots, s_n \in \mathbb{N}_0$ and $\sum_{i=1}^n s_i = d_1$. For the l^{th} term of the above sum, let us define $g_{l1}(\mathbf{a}) = \frac{q_j!}{s_1!\dots s_n!} a_1^{s_1}\dots a_n^{s_n}$ and $h_{l1}(\phi(\mathbf{x})) = \prod_{i=1}^n \phi_{i1}(\mathbf{x})^{s_i}$, yielding

$$\Big(\sum_{i=1}^n a_i\phi_{i1}(\mathbf{x})\Big)^{q_j} = \sum_l g_{l1}(\mathbf{a}) h_{l1}(\phi(\mathbf{x})).$$

$$\begin{aligned}
\chi_1^{-1}(\mathbf{y}) &= x_1 = \frac{(H_{22} - H_{32}H_{23})y_1 - (H_{12} - H_{32}H_{13})y_2 + H_{23}H_{12} - H_{22}H_{13}}{(H_{32}H_{21} - H_{31}H_{22})y_1 - (H_{32}H_{11} - H_{31}H_{12})y_2 + H_{22}H_{11} - H_{21}H_{12}} \\
\chi_2^{-1}(\mathbf{y}) &= x_2 = \frac{-(H_{21} - H_{31}H_{23})y_1 + (H_{11} - H_{31}H_{13})y_2 - (H_{23}H_{11} - H_{21}H_{13})}{(H_{32}H_{21} - H_{31}H_{22})y_1 - (H_{32}H_{11} - H_{31}H_{12})y_2 + H_{22}H_{11} - H_{21}H_{12}} \\
|J_{\chi^{-1}}(\mathbf{y})| &= \frac{|\mathbf{H}|^2}{((H_{32}H_{21} - H_{31}H_{22})y_1 - (H_{32}H_{11} - H_{31}H_{12})y_2 + H_{22}H_{11} - H_{21}H_{12})^3}
\end{aligned} \tag{A.24}$$

Hence $\left(\sum_{i=1}^{n} a_i \phi_{i1}(\mathbf{x})\right)^{q_j}$ is separable and similarly $\left(\sum_{i=1}^{n} a_i \phi_{i2}(\mathbf{x})\right)^{r_j}$ is also separable. Furthermore, their product is also separable by Lemma A.3.1, thus we proved that p_1 is separable. Similarly, it is easy to see that p_2 is also separable, which completes the proof. **Q.E.D.**

Now the statement of Theorem 6.3.1 is easily seen: $f(\varphi(\mathbf{x}))$ and $|J_\varphi(\mathbf{x})|$ are separable by Lemma A.3.3 and Lemma A.3.2, respectively. Hence their product $f(\varphi(\mathbf{x}))|J_\varphi(\mathbf{x})|$ is also separable by Lemma A.3.1 and using the basic properties of integral calculus, we get Eq. (6.14).

Author's publications

[1] C. Benedek, T. Sziranyi, **Z. Kato**, and J. Zerubia, "A multi-layer MRF model for object-motion detection in uregistered airborne image-pairs," in *Proceedings of International Conference on Image Processing*, (San Antonio, Texas, USA), pp. 141–144, IEEE, IEEE, Sep. 2007.

[2] C. Benedek, T. Sziranyi, **Z. Kato**, and J. Zerubia, "Detection of object motion regions in aerial image pairs with a multilayer Markovian model," *IEEE Transactions on Image Processing*, vol. 18, pp. 2303–2315, Oct. 2009.

[3] M. Berthod, **Z. Kato**, S. Yu, and J. Zerubia, "Bayesian image classification using Markov random fields," *Image and Vision Computing*, vol. 14, pp. 285–295, 1996.

[4] M. Berthod, **Z. Kato**, and J. Zerubia, "DPA: A deterministic approach to the MAP," *IEEE Transactions on Image Processing*, vol. 4, pp. 1312–1314, Sep. 1995.

[5] T. Blaskovics, **Z. Kato**, and I. Jermyn, "A Markov random field model for extracting near-circular shapes," in *Proceedings of International Conference on Image Processing*, (Cairo, Egypt), pp. 1073–1076, IEEE, IEEE, Nov. 2009.

[6] C. Domokos and **Z. Kato**, "Binary image registration using covariant Gaussian densities," in *International Conference on Image Analysis and Recognition*, (A. Campilho and M. Kamel, eds.), (Póvoa de Varzim, Portugal), pp. 455–464, Springer, June 2008.

[7] C. Domokos and **Z. Kato**, "Affine alignment of compound objects: A direct approach," in *Proceedings of International Conference on Image Processing*, (Cairo, Egypt), pp. 169–172, IEEE, IEEE, Nov. 2009.

[8] C. Domokos and **Z. Kato**, "Affine puzzle: Realigning deformed object fragments without correspondences," in *Proceedings of European Conference on Computer Vision*, (K. Daniilidis, P. Maragos, and N. Paragios, eds.), (Crete, Greece), pp. 777–790, Springer, Sep. 2010.

[9] C. Domokos and **Z. Kato**, "Parametric estimation of affine deformations of planar shapes," *Pattern Recognition*, vol. 43, pp. 569–578, March 2010.

[10] C. Domokos and **Z. Kato**, "Simultaneous affine registration of multiple shapes," in *Proceedings of International Conference on Pattern Recognition*, (Tsukuba Science City, Japan), pp. 9–12, IAPR, IEEE, Nov. 2012.

[11] C. Domokos, **Z. Kato**, and J. M. Francos, "Parametric estimation of affine deformations of binary images," in *Proceedings of International Conference on Acoustics, Speech and Signal Processing*, (Las Vegas, Nevada, USA), pp. 889–892, IEEE, IEEE, Apr. 2008.

[12] C. Domokos, J. Nemeth, and **Z. Kato**, "Nonlinear shape registration without correspondences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, pp. 943–958, May 2012.

[13] P. Horvath, I. Jermyn, **Z. Kato**, and J. Zerubia, "A higher-order active contour model of a 'gas of circles' and its application to tree crown extraction," *Pattern Recognition*, vol. 42, pp. 699–709, May 2009.

[14] **Z. Kato**, M. Berthod, and J. Zerubia, "Parallel image classification using multiscale Markov random fields," in *Proceedings of International Conference on Acoustics, Speech and Signal Processing*, (Minneapolis), pp. 137–140, IEEE, IEEE, Apr. 1993.

[15] **Z. Kato**, "Reversible jump Markov chain Monte Carlo for unsupervised MRF color image segmentation," in *Proceedings of British Machine Vision Conference*, (A. Hoppe, S. Barman, and T. Ellis, eds.), (Kingston, UK), pp. 37–46, BMVA, Sep. 2004.

[16] **Z. Kato**, "Segmentation of color images via reversible jump MCMC sampling," *Image and Vision Computing*, vol. 26, pp. 361–371, March 2008.

[17] **Z. Kato**, "A unifying framework for correspondence-less linear shape alignment," in *International Conference on Image Analysis and Recognition*, (A. Campilho and M. Kamel, eds.), (Aveiro, Portugal), pp. 277–284, Springer, June 2012.

[18] **Z. Kato**, "Linear and nonlinear shape alignment without correspondences," in *Proceedings of International Joint Conference on Computer Vision, Imaging and Computer Graphics - Theory and Applications*, (J. Braz and G. Csurka, eds.), pp. 3–17, Rome, Italy: Springer, 2013. Keynote talk.

[19] **Z. Kato**, "A unifying framework for correspondence-less shape alignment and its medical applications," in *Proceedings of International Conference on Intelligent Interactive Technologies and Multimedia*, (A. Agrawal, R. Tripathi, and M. Tiwari, eds.), pp. 40–52, Allahabad, India: Springer, March 2013. Keynote talk.

[20] **Z. Kato**, M. Berthod, and J. Zerubia, "A hierarchical Markov random field model and multi-temperature annealing for parallel image classification," Research Report 1938, INRIA, Sophia Antipolis, France, Aug. 1993.

[21] **Z. Kato**, M. Berthod, and J. Zerubia, "A hierarchical Markov random field model for image classification," in *Proceedings of International Workshop on Image and Multidimensional Digital Signal Processing*, (Cannes, France), IEEE, Sep. 1993.

[22] **Z. Kato**, M. Berthod, and J. Zerubia, "Multiscale Markov random field models for parallel image classification," in *Proceedings of International Conference on Computer Vision*, (Berlin, Germany), pp. 253–257, IEEE, May 1993.

[23] **Z. Kato**, M. Berthod, and J. Zerubia, "Parallel image classification using multiscale Markov random fields," in *Proceedings of International Conference on Acoustics, Speech and Signal Processing*, (Minneapolis, USA), pp. 137–140, IEEE, Apr. 1993.

[24] **Z. Kato**, M. Berthod, and J. Zerubia, "A hierarchical Markov random field model and multi-temperature annealing for parallel image classification," *Computer Vision, Graphics and Image Processing: Graphical Models and Image Processing*, vol. 58, pp. 18–37, Jan. 1996.

[25] **Z. Kato**, M. Berthod, J. Zerubia, and W. Pieczynski, "Unsupervised adaptive image segmentation," in *Proceedings of International Conference on Acoustics, Speech and Signal Processing*, (Detroit, Michigan, USA), pp. 2399–2402, IEEE, May 1995.

[26] **Z. Kato** and T. C. Pong, "A Markov random field image segmentation model using combined color and texture features," in *Proceedings of International Conference on Computer Analysis of Images and Patterns*, (W. Skarbek, ed.), (Warsaw, Poland), pp. 547–554, Springer, Sep. 2001.

[27] **Z. Kato** and T. C. Pong, "Video object segmentation using a multicue Markovian model," in *Joint Hungarian-Austrian Conference on Image Processing and Pattern Recognition*, (D. Chetverikov, L. Czuni, and M. Vincze, eds.), (Veszprem, Hungary), pp. 111–118, KEPAF, OAGM/AAPR, Austrian Computer Society, May 2005.

[28] **Z. Kato** and T. C. Pong, "A Markov random field image segmentation model for color textured images," *Image and Vision Computing*, vol. 24, pp. 1103–1114, Oct. 2006.

[29] **Z. Kato** and T. C. Pong, "A multi-layer MRF model for video object segmentation," in *Proceedings of Asian Conference on Computer Vision*, (P. J. Narayanan, S. K. Nayar, and H.-Y. Shum, eds.), (Hyderabad, India), pp. 953–962, Springer, Jan. 2006.

[30] **Z. Kato**, T. C. Pong, and G. Q. Song, "Multicue MRF image segmentation: Combining texture and color," in *Proceedings of International Conference on Pattern Recognition*, (Quebec, Canada), pp. 660–663, IAPR, IEEE, Aug. 2002.

[31] **Z. Kato**, T. C. Pong, and G. Q. Song, "Unsupervised segmentation of color textured images using a multi-layer MRF model," in *Proceedings of International Conference on Image Processing*, (Barcelona, Spain), pp. 961–964, IEEE, Sep. 2003.

[32] **Z. Kato** and J. Zerubia, *Markov random fields in image segmentation*. *Foundations and Trends in Signal Processing*, Now Publishers, Sep. 2012. 164 pages.

[33] **Z. Kato**, J. Zerubia, and M. Berthod, "Image classification using Markov random fields with two new relaxation methods: Deterministic pseudo annealing and modified Metropolis dynamics," Research Report 1606, INRIA, Sophia Antipolis, France, Feb. 1992.

[34] **Z. Kato**, J. Zerubia, and M. Berthod, "Satellite image classification using a modified Metropolis dynamics," in *Proceedings of International Conference on Acoustics, Speech and Signal Processing*, (San-Francisco, California, USA), pp. 573–576, IEEE, March 1992.

[35] **Z. Kato**, J. Zerubia, and M. Berthod, "Bayesian image classification using Markov random fields," in *Maximum Entropy and Bayesian Methods*, (A. Mohammad-Djafari and G. Demoment, eds.), pp. 375–382, Kluwer Academic Publisher, 1993.

[36] **Z. Kato**, J. Zerubia, and M. Berthod, "Unsupervised parallel image classification using a hierarchical Markovian model," Research Report 2528, INRIA, Sophia Antipolis, France, Apr. 1995.

[37] **Z. Kato**, J. Zerubia, and M. Berthod, "Unsupervised parallel image classification using a hierarchical Markovian model," in *Proceedings of International Conference on Computer Vision*, (Cambridge, MA, USA), pp. 169–174, IEEE, June 1995.

[38] **Z. Kato**, J. Zerubia, and M. Berthod, "Unsupervised parallel image classification using Markovian models," *Pattern Recognition*, vol. 32, pp. 591–604, Apr. 1999.

[39] J. Mitra, **Z. Kato**, S. Ghose, D. Sidibe, R. Mart, X. Llado, A. Oliver, J. C. Vilanova, and F. Meriaudeau, "Spectral clustering to model deformations for fast multimodal prostate registration," in *Proceedings of International Conference on Pattern Recognition*, (Tsukuba Science City, Japan), pp. 2622–2625, IAPR, IEEE, Nov. 2012.

[40] J. Mitra, **Z. Kato**, R. Marti, A. Oliver, X. Llado, S. Ghose, J. C. Vilanova, and F. Meriaudeau, "A non-linear diffeomorphic framework for prostate multimodal registration," in *Proceedings of International Conference on Digital Image Computing: Techniques and Applications*, (Noosa, Queensland, Australia), pp. 31–36, IEEE, Dec. 2011.

[41] J. Mitra, **Z. Kato**, R. Marti, A. Oliver, X. Llado, D. Sidibe, S. Ghose, J. C. Vilanova, J. Comet, and F. Meriaudeau, "A spline-based non-linear diffeomorphism for multimodal prostate registration," *Medical Image Analysis*, vol. 16, pp. 1259–1279, Aug. 2012.

[42] C. Molnar, **Z. Kato**, and I. Jermyn, "A multi-layer phase field model for extracting multiple near-circular objects," in *Proceedings of International Conference on Pattern Recognition*, (Tsukuba Science City, Japan), pp. 1427–1430, IAPR, IEEE, Nov. 2012.

[43] J. Nemeth, C. Domokos, and **Z. Kato**, "Nonlinear registration of binary shapes," in *Proceedings of International Conference on Image Processing*, (Cairo, Egypt), pp. 1001–1004, IEEE, November 2009.

[44] J. Nemeth, C. Domokos, and **Z. Kato**, "Recovering planar homographies between 2D shapes," in *Proceedings of International Conference on Computer Vision*, (Kyoto, Japan), pp. 2170–2176, IEEE, September 2009.

[45] J. Nemeth, **Z. Kato**, and I. Jermyn, "A multi-layer 'gas of circles' Markov random field model for the extraction of overlapping near-circular objects," in *Proceedings of Advanced Concepts for Intelligent Vision Systems*, (J. Blanc-Talon, W. Philips, D. Popescu, P. Scheunders, and R. Kleihorst, eds.), (Ghent, Belgium), pp. 171–182, Springer, Aug. 2011.

[46] Z. Santa and **Z. Kato**, "Elastic registration of 3D deformable objects," in *Proceedings of International Conference on Digital Image Computing: Techniques and Applications*, (Fremantle, Australia), pp. 1–7, IEEE, Dec. 2012.

[47] Z. Santa and **Z. Kato**, "A unifying framework for non-linear registration of 3D objects," in *Proceedings of International Conference on Cognitive Infocommunications*, (Kassa, Slovakia), pp. 547–552, IEEE, IEEE, Dec. 2012.

[48] Z. Santa and **Z. Kato**, "Correspondence-less non-rigid registration of triangular surface meshes," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, (Portland, Oregon, USA), IEEE, IEEE, June 2013. Accepted.

[49] A. Tanács, C. Domokos, N. Sladoje, J. Lindblad, and **Z. Kato**, "Recovering affine deformations of fuzzy shapes," in *Proceedings of Scandinavian Conferences on Image Analysis*, (A.-B. Salberg, J. Y. Hardeberg, and R. Jenssen, eds.), (Oslo, Norway), pp. 735–744, Springer, June 2009.

[50] A. Tanacs and **Z. Kato**, "Fast linear registration of 3D objects segmented from medical images," in *Proceedings of International Conference on BioMedical Engineering and Informatics*, (Shanghai, China), pp. 299–303, IEEE, Oct. 2011.

[51] A. Tanács, N. Sladoje, J. Lindblad, and **Z. Kato**, "Estimation of linear deformations of 3D objects," in *Proceedings of International Conference on Image Processing*, (Hong Kong, China), pp. 153–156, IEEE, IEEE, Sep. 2010.

Bibliography

[52] G. L. Andriole, E. D. Crawford, R. L. Grubb, S. S. Buys, D. Chia, T. R. Church, M. N. Fouad, E. P. Gelmann, D. J. Reding, J. L. Weissfeld, L. A. Yokochi, B. O'Brien, J. D. Clapp, J. M. Rathmell, T. L. Riley, R. B. Hayes, B. S. Kramer, G. Izmirlian, A. B. Miller, P. F. Pinsky, P. C. Prorok, J. K. Gohagan, and C. D. Berg, "Mortality results from a randomized prostate-cancer screening trial," *The New England Journal of Medicine*, vol. 360, no. 13, pp. 1310–1319, 2009.

[53] V. Arsigny, O. Commowick, N. Ayache, and X. Pennec, "A fast and log-euclidean polyaffine framework for locally linear registration," *Journal of Mathematical Imaging and Vision*, vol. 33, pp. 222–238, February 2009.

[54] R. Azencott, "Markov fields and image analysis," *Proc. AFCET, Antibes*, 1987.

[55] C. Baillard and H. Maître, "3-D reconstruction of urban scenes from aerial stereo imagery: a focusing strategy," *Comput. Vis. Image Underst.*, vol. 76, no. 3, pp. 244–258, 1999.

[56] Y. Bard, *Nonlinear Parameter Estimation*. Academic Press, 1974.

[57] S. A. Barker and P. J. W. Rayner, "Unsupervised image segmentation using Markov random field models," *Pattern Recognition*, vol. 33, pp. 587–602, Apr. 2000.

[58] R. J. Baxter, *Exactly Solved Models in Statistical Mechanics*. Academic Press, 1990.

[59] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding*, vol. 110, pp. 346–359, August 2008.

[60] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape context," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 509–522, April 2002.

[61] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape context," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 509–522, April 2002.

[62] C. Benedek and T. Sziranyi, "Change detection in optical aerial images by a multilayer conditional mixed markov model," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 47, pp. 3416–3430, Oct. 2009.

[63] J. Besag, "Spatial interaction and the statistical analysis of lattice systems (with discussion)," *Journal of the Royal Statistical Society, ser. B*, vol. 36, no. 2, pp. 192–236, 1974.

[64] J. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society, ser. B*, vol. 48, no. 3, pp. 259–302, 1986.

[65] P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, pp. 239–256, Feb. 1992.

[66] A. Blake and A. Zisserman, *Visual reconstruction*. MIT Press, 1987.

[67] F. L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 567–585, June 1989.

[68] B. Braathen, W. Pieczynski, and P. Masson, "Global and Local Methods of Unsupervised Bayesian Segmentation of Images," *Machine Graphics and Vision*, vol. 2, no. 1, pp. 39–52, 1993.

[69] A. M. Bronstein and M. M. Bronstein, "Regularized partial matching of rigid shapes," in *Proceedings of the Europian Conference on Computer Vision*, (D. Forsyth, P. Torr, and A. Zisserman, eds.), (Marseille, France), pp. 143–154, Springer, October 2008.

[70] A. M. Bronstein, M. M. Bronstein, A. M. Bruckstein, and R. Kimmel, "Analysis of two-dimensional non-rigid shapes," *International Journal of Computer Vision*, vol. 78, pp. 67–88, June 2008.

[71] L. G. Brown, "A survey of image registration techniques," *ACM Computing Surveys*, vol. 24, pp. 325–376, December 1992.

[72] A. S. Bryant and R. J. Cerfolio, "The maximum standardized uptake values on integrated FDG-PET/CT is useful in differentiating benign from malignant pulmonary nodules," *The Annals of Thoracic Surgery*, vol. 82, pp. 1016–1020, 2006.

[73] P. Carroll and K. Shinohara, "Transrectal ultrasound guided prostate biopsy," Tech. Rep., Department of Urology, University of California, San Francisco, 2010. http://urology.ucsf.edu/patientGuides.html, accessed [30th Dec, 2010].

[74] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *International Journal of Computer Vision*, vol. 22, no. 1, pp. 61–79, 1997.

[75] G. Celeux and J. Diebolt, "The SEM algorithm: a probabilisitic teacher algorithm derived from the EM algorithm for the mixture problem," *Computational Statistics Quaterly*, vol. 2, pp. 73–82, 1985.

[76] B. Chalmond, "An iterative Gibbsian technique for reconstruction of M-ary images," *Pattern Recognition*, vol. 22, no. 6, pp. 747–762, 1989.

[77] B. Chalmond, *Modeling and Inverse Problems in Image Analysis*. Springer, 2003.

[78] Y. Chen, H. Tagare, S. Thiruvenkadam, F. Huang, D. Wilson, K. Gopinath, R. Briggs, and E. 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.

[79] L. Cohen, "On active contour models and balloons," *Computer Vision, Graphics and Image Processing: Image Understanding*, vol. 53, pp. 211–218, March 1991.

[80] D. Cremers, F. Tischhauser, J. Weickert, and C. Schnorr, "Diffusion snakes: Introducing statistical shape knowledge into the Mumford-Shah functional," *International Journal of Computer Vision*, vol. 50, no. 3, pp. 295–313, 2002.

[81] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, ser. B*, vol. 39, no. 1, pp. 1–38, 1977.

[82] Y. Deng, , and B. S. Manjunath, "Unsupervised segmentation of color-texture regions in images and video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 800–810, Aug. 2001.

[83] X. Descombes and J. Zerubia, "Marked point process in image analysis," *IEEE Signal Processing Magazine*, pp. 77–84, Sep. 2002.

[84] M. Downing, P. Undrill, P. Ashcroft, D. Hukins, and J. Hutchison, "Automated femoral measurement in total hip replacement radiographs," in *Proceedings of International Conference on Image Processing and Its Applications*, (Dublin, Ireland), pp. 843–847, IEEE, July 1997.

[85] B. Erdőhelyi and E. Varga, "Semi-automatic bone fracture reduction in surgical planning," in *Proceedings of the International Conference on Computer Assisted Radiology and Surgery*, (Berlin, Germany), pp. S98–S99, Springer, June 2009.

[86] B. Flach and D. Schlesinger, "Combining shape priors and MRF-segmentation," in *Proceedings of the 2008 Joint IAPR International Workshop on Structural, Syntactic, and Statistical Pattern Recognition*, (Orlando, Florida), pp. 177–186, Lecture Notes In Computer Science, 2008.

[87] C. Florea, C. Vertan, and L. Florea, "Logarithmic model-based dynamic range enhancement of hip x-ray images," in *Proceedings of International Conference on Advanced Concepts for Intelligent Vision Systems*, (Delft, Netherlands), pp. 587–596, Springer, August 2007.

[88] J. Flusser, J. Kautsky, and F. Šroubek, "Implicit moment invariants," *International Journal of Computer Vision*, vol. 86, pp. 72–86, January 2010.

[89] J. Flusser and T. Suk, "Pattern recognition by affine moment invariants," *Pattern Recognition*, vol. 1, pp. 167–174, January 1993.

[90] J. Flusser and T. Suk, "Projective moment invariants," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 1364–1367, October 2004.

[91] J. Flusser, T. Suk, and B. Zitová, *Moments and Moment Invariants in Pattern Recognition*. Wiley & Sons, Oct. 2009.

[92] A. Foulonneau, P. Charbonnier, and F. Heitz, "Multi-reference shape priors for active contours," *Int. Journal of Computer Vision*, vol. 81, pp. 68–81, January 2009.

[93] J. M. Francos, R. Hagege, and B. Friedlander, "Estimation of multidimensional homeomorphisms for object recognition in noisy environments," in *Proceedings of Conference on Signals, Systems and Computers*, (Pacific Grove, CA, USA), pp. 1615–1619, November 2003.

[94] T. Funkhouser and P. Shilane, "Partial matching of 3D shapes with priority-driven search," in *Proceedings of the Eurographics Symposium on Geometry Processing*, (Sardinia, Italy), pp. 1–12, Eurographics, ACM SIGGRAPH, June 2006.

[95] R. Gan and A. C. S. Chung, "Distance-intensity for image registration," in *Proceedings of Computer Vision for Biomedical Image Applications*, (Beijing, China), pp. 281–290, Springer, October 2005.

[96] D. Geman, "Bayesian image analysis by adaptive annealing," in *Proceedings of International Geoscience and Remote Sensing Symposium*, (Amherst, USA), pp. 269–277, IEEE, IEEE, Oct. 1985.

[97] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721–741, 1984.

[98] S. Ghose, A. Oliver, R. Marti, X. Llado, J. Freixenet, J. C.Vilanova, and F. Meriaudeau, "Prostate segmentation with texture enhanced active appearance model," in *Proceedings of International Conference on Signal-Image Technology and Internet Based Systems*, (Kuala Lumpur, Malaysia), pp. 18–22, IEEE, Dec. 2010.

[99] N. Giordana and W. Pieczynski, "Estimation of generalized multisensor hidden Markov chains and unsupervised image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 465–475, May 1997.

[100] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Prentice Hall, 2008.

[101] A. Goshtasby, "Registration of images with geometric distortions," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 26, pp. 60–64, January 1988.

[102] V. Govindu and C. Shekhar, "Alignment using distributions of local geometric properties," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, pp. 1031–1043, October 1999.

[103] U. Grenander, *General Pattern Theory*. Oxford University Press, 1993.

[104] A. Guimond, J. Meunier, and J. P. Thirion, "Average brain models: a convergence study," *Comput. Vision Image Understand.*, vol. 77, pp. 192–210, 2000.

[105] H. Guo, A. Rangarajan, S. Joshi, and L. Younes, "Non-rigid registration of shapes via diffeomorphic point matching," in *Proceedings of International Symposium on Biomedical Imaging: From Nano to Macro*, (Arlington, VA, USA), pp. 924–927, IEEE, April 2004.

[106] L. Gupta and T. Sortrakul, "A Gaussian-mixture-based image segmentation algorithm," *Pattern Recognition*, vol. 31, no. 3, pp. 315–325, 1998.

[107] R. Hagege and J. M. Francos, "Parametric estimation of multi-dimensional affine transformations:an exact linear solution," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*, (Philadelphia, USA), pp. 861–864, IEEE, March 2005.

[108] R. Hagege and J. M. Francos, "Linear estimation of sequences of multi-dimensional affine transformations," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*, (Toulouse, France), pp. 785–788, IEEE, May 2006.

[109] R. Hagege and J. M. Francos, "Parametric estimation of affine transformations: An exact linear solution," *Journal of Mathematical Imaging and Vision*, vol. 37, pp. 1–16, Jan. 2010.

[110] M. S. Hansen, M. F. Hansen, and R. Larsen, "Diffeomorphic statistical deformation models," in *Proceedings of International Conference on Computer Vision*, (Rio de Janeiro, Brazil), pp. 1–8, IEEE, October 2007.

[111] K. Hardinge, M. L. Porter, P. R. Jones, D. W. L. Hukins, and C. J. Taylor, "Measurement of hip prostheses using image analysis. the maxima hip technique," *Journal of Bone and Joint Surgery*, vol. 73-B, pp. 724–728, September 1991.

[112] R. Hartley, "In defense of the eight-point algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 580–593, June 1997.

[113] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their application," *Biometrika*, vol. 57, pp. 97–109, 1970.

[114] D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes, "Medical image registration," *Physics in Medicine and Biology*, vol. 46, pp. R1–R45, March 2001.

[115] D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes, "Medical image registration," *Physics in Medicine and Biology*, vol. 46, pp. R1–R45, March 2001.

[116] F. Hjouj and D. W. Kammler, "Identification of reflected, scaled, translated, and rotated objects from their Radon projections," *IEEE Transactions on Image Processing*, vol. 17, pp. 301–310, March 2008.

[117] J. Ho, M.-H. Yang, A. Rangarajan, and B. Vemuri, "A new affine registration algorithm for matching 2D point sets," in *Proceedings of IEEE Workshop on Applications of Computer Vision*, (Austin, Texas, USA), p. 25, IEEE, February 2007.

[118] M. Holden, "A review of geometric transformations for nonrigid body registration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 111–128, Jan. 2008.

[119] P. Horvath, *The 'Gas of Circles' Model and its Application to Tree Crown Extraction*. PhD thesis, University of Nice - Sophia Antipolis and University of Szeged, Dec. 2007.

[120] P. Horvath and I. Jermyn, "A new phase field model of a 'gas of circles' for tree crown extraction from aerial images," in *Proceedings of International Conference on Computer Analysis of Images and Patterns*, (Vienna, Austria), pp. 702–709, August 2007.

[121] M.-K. Hu, "Visual pattern recognition by moment invariants," *IRE Transactions on Information Theory*, vol. 8, pp. 179–187, February 1962.

[122] Y. Hu, H. U. Ahmed, Z. Taylor, C. Allem, M. Emberton, D. Hawkes, and D. Barratt, "MR to ultrasound registration for image-guided prostate interventions," *Medical Image Analysis, In Press*, 2011. doi:10.1016/j.media.2010.11.003.

[123] R. Huang, V. Pavlovic, and D. N. Metaxas, "A graphical model framework for coupling MRFs and deformable models," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, (Washington, DC, USA), pp. 739–746, IEEE, IEEE, June 2004.

[124] P. Jain and C. V. Jawahar, "Homography estimation from planar contours," in *Proceedings of International Symposium on 3D Data Processing, Visualization, and Transmission*, (Chapel Hill, NC, USA), pp. 877–884, June 2006.

[125] A. Kadyrov and M. Petrou, "Affine parameter estimation from the trace transform," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 1631–1645, October 2006.

[126] S. Kaneko, Y. Satohb, and S. Igarashi, "Using selective correlation coefficient for robust image registration," *Pattern Recognition*, vol. 36, pp. 1165–1173, May 2003.

[127] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.

[128] D. Kersten, P. Mamassian, and A. Yuille, "Object perception as Bayesian inference," *Annual Review of Psychology*, vol. 55, pp. 271–304, 2004.

[129] C. Kervrann and F. Heitz, "Statistical deformable model-based segmentation of image motion," *IEEE Transactions on Image Processing*, vol. 8, pp. 583–588, 1999.

[130] S. Khan and M. Shah, "Object based segmentation of video using color, motion and spatial information," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, (Kauai, Hawaii), pp. 746–751, IEEE, Dec. 2001.

[131] S. Klein, J. P. Pluim, M. Staring, and M. A. Viergever, "Adaptive stochastic gradient descent optimisation for image registration," *International Journal of Computer Vision*, vol. 81, pp. 227–239, March 2009.

[132] W. Kong and B. B. Kimia, "On solving 2D and 3D puzzles using curve matching," in *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, (Kauai, Hawaii), pp. 1–8, IEEE, Dec. 2001.

[133] M. P. Kumar, S. Kuthirummal, C. V. Jawahar, and P. J. Narayanan, "Planar homography from fourier domain representation," in *Proceedings of International Conference on Signal Processing and Communications*, (Bangalore, India), pp. 560–564, IEEE, December 2004.

[134] M. P. Kumar, S. Kuthirummal, C. V. Jawahar, and P. J. Narayanan, "Planar homography from Fourier domain representation," in *Proceedings of International Conference on Signal Processing and Communications*, (Bangalore, India), pp. 560–564, IEEE, December 2004.

[135] R. Kumar, R. H. Sawhney, S. Samarasekera, S. Hsu, H. Tao, Y. Guo, K. Hanna, A. Pope, R. Wildes, D. Hirvonen, M. Hansen, and P. Burt, "Aerial video surveillance and exploitation," in *Proceeding of the IEEE*, pp. 1518–1539, 2001.

[136] P. V. Laarhoven and E. Aarts, *Simulated annealing: Theory and applications*. Kluwer Academic Publisher, 1987.

[137] S. Lakshmanan and H. Derin, "Simultaneous parameter estimation and segmentation of Gibbs random fields using simulated annealing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 799–813, Aug. 1989.

[138] Y. LeCun and C. Cortes, "The MNIST database of handwritten digits," database, `http://yann.lecun.com/exdb/mnist/`.

[139] V. Lepetit and P. Fua, "Keypoint recognition using randomized trees," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 1465–1479, September 2006.

[140] E. Littmann and H. Ritter, "Adaptive color segmentation–a comparison of neural and statistical methods," *IEEE Transactions on Neural Networks*, vol. 8, pp. 175–185, Jan. 1997.

[141] M. Lourakis, "homest: A C/C++ library for robust, non-linear homography estimation," software, `http://www.ics.forth.gr/~lourakis/homest/`, 2008.

[142] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–110, November 2004.

[143] J. B. A. Maintz and M. A. Viergever, "A survey of medical image registration," *Medical Image Analysis*, vol. 2, pp. 1–36, March 1998.

[144] J. B. A. Maintz and M. A. Viergever, "A survey of medical image registration," *Medical Image Analysis*, vol. 2, pp. 1–36, March 1998.

[145] S. Manay, D. Cremers, B.-W. Hong, A. Y. Jr., and S. Soatto, "Integral invariants for shape matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 1602–1618, October 2006.

[146] S. Mann and R. W. Picard, "Video orbits of the projective group a simple approach to featureless estimation of parameters," *IEEE Transactions on Image Processing*, vol. 6, pp. 1281–1295, September 1997.

[147] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *SIAM Journal on Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963.

[148] S. Marsland and C. Twining, "Constructing diffeomorphic representations for the groupwise analysis of nonrigid registrations of medical images," *IEEE Transactions on Medical Imaging*, vol. 23, pp. 1006–1020, August 2004.

[149] P. Masson and W. Pieczynski, "SEM Algorithm and Unsupervised Statistical Segmentation of Satellite Images," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 31, pp. 618–633, May 1993.

[150] J. C. McBride and B. B. Kimia, "Archaeological fragment reconstruction using curve-matching," in *Proceedings of the International Conference on Computer Vision and Pattern Recognition Workshop*, (Madison, Wisconsin, USA), pp. 1–8, IEEE, June 2003.

[151] R. Merris, *Combinatorics*. *Wiley-Interscience Series in Discrete Mathematics and Optimization*, John Wiley & Sons, 2 ed., August 2003.

[152] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087–1092, 1953.

[153] M. Miller and L. Younes, "Group actions, homeomorphisms, and matching: A general framework," *International Journal of Computer Vision*, vol. 41, pp. 61–84, Feb. 2001.

[154] M. I. Miller, U. Grenander, O. J. A., and D. L. Snyder, "Automatic target recognition organized via jump-diffusion algorithms," *IEEE Transactions on Image Processing*, vol. 6, pp. 157–174, Jan. 1997.

[155] J. Mitra, A. Srikantha, D. Sidibé, R. Martí, A. Oliver, X. Lladó, J. C. Vilanova, and F. Meriaudeau, "A shape-based statistical method to retrieve 2D TRUS-MR slice correspondence for prostate biopsy," in *Proc. of SPIE Medical Imaging*, (San Diego, Calfornia, USA), pp. 83143M–1–9, February 2012.

[156] D. Mumford, "The Bayesian rationale for energy functionals," in *Geometry-Driven Diffusion in Computer Vision*, (B. Romeny, ed.), pp. 141–153, Kluwer Academic Publisher, 1994.

[157] D. Mumford, "Pattern theory: a unifying perspective," in *Perception as Bayesian Inference*, (D. Knill and W. Richards, eds.), pp. 25–62, Cambridge University Press, 1996.

[158] D. Mumford and J. 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.

[159] O. Musse, F. Heitz, and J.-P. Armspach, "Topology preserving deformable image matching using constrained hierarchical parametric models," *IEEE Transactions on Image Processing*, vol. 10, pp. 1081–1093, July 2001.

[160] M. Nielsen, P. Johansen, A. D. Jackson, B. Lautrup, and S. Hauberg, "Brownian warps for non-rigid registration," *Journal of Mathematical Imaging and Vision*, vol. 31, pp. 221–231, July 2008.

[161] J. C. Noordam, G. W. Otten, A. J. M. Timmermans, and B. v. Zwol, "High-speed potato grading and quality inspection based on a color vision system," in *Proceedings of SPIE Machine Vision Applications in Industrial Inspection*, (K. W. T. Jr., ed.), pp. 206–220, 2000.

[162] A. Oprea and C. Vertan, "A quantitative evaluation of the hip prosthesis segmentation quality in x-ray images," in *Proceedings of International Symposium on Signals, Circuits and Systems*, (Iasi, Romania), pp. 1–4, IEEE, July 2007.

[163] G. Papaioannou, E.-A. Karabassi, and T. Theoharis, "Reconstruction of three-dimensional objects through matching of their parts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 114–124, Jan. 2002.

[164] C. F. Paulo and P. L. Correia, "Automatic detection and classification of traffic signs," in *Proceedings of Workshop on Image Analysis for Multimedia Interactive Services*, (L. O'Conner, ed.), (Santorini, Greece), pp. 11–14, IEEE, June 2007.

[165] G. Perrin, X. Descombes, and J. Zerubia, "A marked point process model for tree crown extraction in plantations," in *Proc. IEEE International Conference on Image Processing (ICIP)*, (Genova, Italy), Sep. 2005.

[166] J. Pettersson, H. Knutsson, and M. Borga, "Non-rigid registration for automatic fracture segmentation," in *Proceedings of International Conference on Image Processing*, (Atlanta, GA, USA), pp. 1185–1188, IEEE, October 2006.

[167] W. Pieczynski, "Statistical image segmentation," in *Proceedings of Machine Graphics and Vision*, (Naleczow, Poland), pp. 261–268, May 1992.

[168] R. Pless, T. Brodsky, and Y. Aloimonos, "Detecting independent motion: The statistics of temporal continuity," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 68–73, 2000.

[169] E. Rahtu, M. Salo, J. Heikkilä, and J. Flusser, "Generalized affine moment invariants for object recognition," in *Proceedings of International Conference on Pattern Recognition*, (Hong-kong), pp. 634–637, IEEE, August 2006.

[170] M. Reuter, "Hierarchical shape segmentation and registration via topological features of Laplace-Beltrami eigenfunctions," *International Journal of Computer Vision*, vol. 89, no. 2, pp. 287–308, 2010.

[171] E. Rignot and R. Chellappa, "Maximum a posteriori classification of multifrequency, multilook, synthetic aperture radar intensity data," *Journal of the Optical Society of America a-Optics Image Science and Vision*, vol. 10, no. 4, pp. 573–582, 1993.

[172] M. Rochery, I. H. Jermyn, and J. Zerubia, "Higher order active contours and their application to the detection of line networks in satellite imagery," in *Proceedings of Workshop on Variational, Geometric and Level Set Methods in Computer Vision*, (at ICCV, Nice, France), Oct. 2003.

[173] M. Rochery, I. H. Jermyn, and J. Zerubia, "Higher order active contours," *International Journal of Computer Vision*, vol. 69, pp. 27–42, August 2006.

[174] M. Rochery, I. H. Jermyn, and J. Zerubia, "Phase field models and higher-order active contours," in *Proc. IEEE International Conference on Computer Vision (ICCV)*, (Beijing, China), October 2005.

[175] K. Rohr, H. S. Stiehl, R. Sprengel, T. M. Buzug, J. Weese, and M. H. Kuhn, "Landmark-based elastic registration using approximating thin-plate splines," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 526–534, June 2001.

[176] M. Rousson and N. Paragios, "Shape priors for level set representations," in *Proceedings of European Conference on Computer Vision*, (Copenhagen, Denmark), pp. 78–92, Springer, 2002.

[177] C. Russell, D. Metaxas, C. Restif, and P. Torr, "Using the P^n Potts model with learning methods to segment live cell images," in *International Conference on Computer Vision*, (Rio de Janeiro, Brazil), pp. 1–8, IEEE, IEEE, Oct. 2007.

[178] R. M. Rustamov, "Laplace-Beltrami eigenfunctions for deformation invariant shape representation," in *Proceedings of the Eurographics Symposium on Geometry Processing*, (Barcelona, Spain), pp. 1–9, Eurographics, ACM SIGGRAPH, July 2007.

[179] J. Salvi, C. Matabosch, D. Fofi, and J. Forest, "A review of recent range image registration methods with accuracy evaluation," *Image and Vision Computing*, vol. 25, pp. 578–596, May 2007.

[180] A. Shekhovtsov, I. Kovtun, and V. Hlaváč, "Efficient MRF deformation model for non-rigid image matching," *Computer Vision and Image Understanding*, vol. 112, pp. 91–99, October 2008.

[181] K. M. Simonson, S. M. Drescher, and F. R. Tanner, "A statistics-based approach to binary image registration with uncertainty analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 112–125, January 2007.

[182] K. M. Simonson, S. M. Drescher, and F. R. Tanner, "A statistics-based approach to binary image registration with uncertainty analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 112–125, January 2007.

[183] A. Srivastava, S. Joshi, W. Mio, and X. Liu, "Statistical shape analysis: Clustering, learning, and testing," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 4, pp. 590–602, 2005.

[184] T. Suk and J. Flusser, "Affine normalization of symmetric objects," in *Proceedings of International Conference on Advanced Concepts for Intelligent Vision Systems*, (J. Blanc-Talon, W. Philips, D. Popescu, and P. Scheunders, eds.), (Antwerp, Belgium), pp. 100–107, Springer, September 2005.

[185] H. D. Tagare, D. Groisser, and O. Skrinjar, "Symmetric non-rigid registration: A geometric theory and some numerical techniques," *Journal of Mathematical Imaging and Vision*, vol. 34, pp. 61–88, May 2009.

[186] A. Tanacs, E. Mate, and A. Kuba, "Application of automatic image registration in a segmentation framework for pelvic CT images," in *Proceedings of CAIP*, pp. 628–635, 2005.

[187] M. R. Teague, "Image analysis via the general theory of moments," *Journal of the Optical Society of America*, vol. 70, pp. 920–930, August 1980.

[188] Z. Tu, X. Chen, A. Yuille, and S.-C. Zhu, "Image parsing: Unifying segmentation, detection, and recognition," *International Journal of Computer Vision*, vol. 63, pp. 113–140, July 2005.

[189] Z. Tu, S. Zheng, and A. Yuille, "Shape matching and registration by data-driven EM," *Computer Vision and Image Understanding*, vol. 109, pp. 290–304, March 2008.

[190] L. van Gool, T. Moons, E. Pauwels, and A. Oosterlinck, "Vision and Lie's approach to invariance," *Image and Vision Computing*, vol. 13, pp. 259–277, May 1995.

[191] N. Vandenbroucke, L. Macaire, and J. Postaire, "Color image segmentation by supervised pixel classification in a color texture feature space. Application to soccer image segmentation," in *Proceedings of the International Conference on Pattern Recognition*, (Barcelona, Spain), pp. 621–624, 2000.

[192] J. A. Ventura and W. Wan, "Accurate matching of two-dimensional shapes using the minimal tolerance zone error," *Image and Vision Computing*, vol. 15, pp. 889–899, December 1997.

[193] J. Verschelde, "Algorithm 795: PHCpack: a general-purpose solver for polynomial systems by homotopy continuation," *ACM Transactions on Mathematical Software*, vol. 25, pp. 251–276, June 1999.

[194] J. Verschelde, "PHCpack: a general-purpose solver for polynomial systems by homotopy continuation," software, July 1999.

[195] J. Verschelde, P. Verlinden, and R. Cools, "Homotopies exploiting newton polytopes for solving sparse polynomial systems," *SIAM Journal on Numerical Analysis*, vol. 31, pp. 915–930, June 1994.

[196] J. C. Vilanova, C. Barceló-Vidal, J. Comet, M. Boada, J. Barceló, J. Ferrer, and J. Albanell, "Usefulness of prebiopsy multi-functional and morphologic MRI combined with the free-to-total PSA ratio in the detection of prostate cancer," *American Journal of Roentgenology*, vol. 196, no. 6, pp. W715–W722, 2011.

[197] J. Wang and Y. Liu, "Characteristic line of planar homography matrix and its applications in camera calibration," in *Proceedings of International Conference on Pattern Recognition*, (Hong-Kong, China), pp. 147–150, IEEE, August 2006.

[198] S. Winkelbach, R. Westphal, and T. Goesling, "Pose estimation of cylindrical fragments for semi-automatic bone fracture reduction," in *Proceedings of Annual Symposium of the German Association for Pattern Recognition*, (B. Michaelis and G. Krell, eds.), (Magdeburg, Germany), pp. 556–573, Springer, September 2003.

[199] G. Winkler, *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods*. Springer, 2003.

[200] S. Wörz and K. Rohr, "Physics-based elastic registration using non-radial basis functions and including landmark localization uncertainties," *Computer Vision Image Understanding*, vol. 111, no. 3, pp. 263–274, 2008.

[201] A. Yezzi, L. Zöllei, and T. Kapurz, "A variational framework for joint segmentation and registration," in *Proceedings of IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, (Kauai, HI, USA), pp. 44–51, IEEE, Dec. 2001.

[202] L. Zagorchev and A. Goshtasby, "A comparative study of transformation functions for nonrigid image registration," *IEEE Transactions on Image Processing*, vol. 15, pp. 529–538, March 2006.

[203] J. Zhang, Y. Ge, S. H. Ong, C. K. Chui, S. H. Teoh, and C. H. Yan, "Rapid surface registration of 3D volumes using a neural network approach," *Image and Vision Computing*, vol. 26, no. 2, pp. 201–210, 2008.

[204] Y. Zhang, C. Wen, and Y. Zhang, "Recognition of symmetrical images using affine moment invariants in both frequency and spatial domains," *Pattern Analysis & Applications*, vol. 5, pp. 316–325, August 2002.

[205] Z. Zhang, R. Deriche, O. D. Faugeras, and Q. T. Luong, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry," *Artificial Intelligence*, vol. 78, pp. 87–119, October 1995.

[206] S. C. Zhu and A. 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.

[207] B. Zitová and J. Flusser, "Image registration methods: A survey," *Image and Vision Computing*, vol. 21, pp. 977–1000, October 2003.