5-2005

Multimodal Three Dimensional Scene Reconstruction, The Gaussian Fields Framework

Fayçal Boughorbel
University of Tennessee - Knoxville

To the Graduate Council:

I am submitting herewith a dissertation written by Fayçal Boughorbel entitled "Multimodal Three Dimensional Scene Reconstruction, The Gaussian Fields Framework." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Electrical Engineering.

Mongi A. Abidi, Major Professor

We have read this dissertation and recommend its acceptance:

Daniel B. Koch, Seong G. Kong, Hairong Qi, Jian Huang, Andreas Koschan, Besma Abidi

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a dissertation written by Fayçal Boughorbel entitled "Multimodal Three Dimensional Scene Reconstruction, The Gaussian Fields Framework." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Electrical Engineering.

Mongi A. Abidi

Major Professor

We have read this dissertation
and recommend its acceptance:

Daniel B. Koch

Seong G. Kong

Hairong Qi

Jian Huang

Andreas Koschan

Besma Abidi

Accepted for the Council:

Anne Mayhew

Vice Chancellor and Dean of
Graduate Studies

(Original signatures are on file with official student records.)

Multimodal Three Dimensional Scene Reconstruction, The Gaussian Fields Framework

A Dissertation Presented for the Doctor of Philosophy Degree

The University of Tennessee, Knoxville

Fayçal Boughorbel

May 2005

To my parents Bouaziz Boughorbel and Khira Ouni

Acknowledgments

Coming to the completion of this work, and to a special moment in the quest of learning that spanned most of my life, I owe many thanks to countless people, some of which I will mention, but many others I hope will forgive my omission. First and foremost I would like to thank my father Mr Bouaziz Boughorbel who spared no effort to help me reach the best education. From him and from my caring mother Mrs Khira Ouni I derive the motivation to go far. I need not say more they know what I feel for them.

To Dr Mongi Abidi I owe much more than the generous financial support and academic guidance. He tempered my impatience with wisdom, helped me go through moments of uncertainty, and provided me with advice about life and work that I will not forget. Without his constant help I cannot imagine coming to this point. Special thanks to Dr Andreas Koschan for his technical advice and for the wonderful conversations we had and that spanned all horizons. I am also indebted Dr Hairong Qi for teaching me the basics of image processing and for her kind support. I would like to thank my committee members: Dr Daniel Koch, Dr Seong Kong, Dr Jian Huang, and Dr Besma Abidi for their time and fruitful comments. To Dr Andrei Gribok I owe several illuminating remarks that greatly strengthened my work, as well as some good passes on the soccer field!

Then I cannot forget my IRIS family: Brad Grinstead who survived next to me in the Lab for all these years. Yohan Fougerolle was always great. Special thanks to David Page, Sun Yiyong, Rangan, Vivek, and Michael Roy; and to all the other IRIS students and former students that I worked with. Vicki Courtney-Smith and Kim Kate deserve many thanks for their professionalism in handling administrative matters and for their kindness. I must acknowledge my aunt Amina for her boundless care during my undergraduate engineering years in Tunis. To my friends and relatives back home Mohamed Abidi, Mohamed Rezgui, Mourad Nasri, Mohamed Saadellaoui, and Laaj, who where always there when I needed them: thank you. Last and certainly not least I owe a lot to the patience and care of my fiancé Boutheina and to the unwavering support of my little brother Sabri.

Abstract

The focus of this research is on building 3D representations of real world scenes and objects using different imaging sensors. Primarily range acquisition devices (such as laser scanners and stereo systems) that allow the recovery of 3D geometry, and multi-spectral image sequences including visual and thermal IR images that provide additional scene characteristics. The crucial technical challenge that we addressed is the automatic point-sets registration task. In this context our main contribution is the development of an optimization-based method at the core of which lies a unified criterion that solves simultaneously for the dense point correspondence and transformation recovery problems. The new criterion has a straightforward expression in terms of the datasets and the alignment parameters and was used primarily for 3D rigid registration of point-sets. However it proved also useful for feature-based multimodal image alignment. We derived our method from simple Boolean matching principles by approximation and relaxation. One of the main advantages of the proposed approach, as compared to the widely used class of Iterative Closest Point (ICP) algorithms, is convexity in the neighborhood of the registration parameters and continuous differentiability, allowing for the use of standard gradient-based optimization techniques. Physically the criterion is interpreted in terms of a Gaussian Force Field exerted by one point-set on the other. Such formulation proved useful for controlling and increasing the region of convergence, and hence allowing for more autonomy in correspondence tasks. Furthermore, the criterion can be computed with linear complexity using recently developed Fast Gauss Transform numerical techniques. In addition, we also introduced a new local feature descriptor that was derived from visual saliency principles and which enhanced significantly the performance of the registration algorithm. The resulting technique was subjected to a thorough experimental analysis that highlighted its strength and showed its limitations. Our current applications are in the field of 3D modeling for inspection, surveillance, and biometrics. However, since this matching framework can be applied to any type of data, that can be represented as N-dimensional point-sets, the scope of the method is shown to reach many more pattern analysis applications.

Contents

Tables

Figures

1 INTRODUCTION

Stating that among all the senses that humans possess vision is the most ubiquitous and the most useful in our lives may seem obvious and redundant, but the workings of such an important pillar of our cognition is so much misunderstood that this may be worth recalling at the beginning of this dissertation. In fact scientists showed that the majority of brain power is devoted to visual processing. We are so dependant on vision that we have to use the visual medium as a way of communicating and recording information through graphic symbols that we call letters and numbers, and even as a way of reasoning. Our understanding of abstract and convoluted concepts is always enhanced when we describe them by images and graphs. Since we like images so much and as our technology evolved we have equipped many of our machines with eyes of their own.

The proliferation of digital imaging systems is rapidly transforming many areas of our lives. Current technologies, however, remain overwhelmingly devoted to acquiring and transmitting the images to its human end user. In most cases, such as when we look at the images of friends or family members transmitted to us from their wireless phones, that is all what we want the systems to do. The image in this case is the end product and is consumed for the subjective psychological satisfaction that it generates for us. Now imagine a scenario where a military pilot or a tank commander are operating in the dark of night, their night vision and other imaging systems show a moving vehicle on the ground. They are not sure if the vehicle is hostile or friendly. Clear images can save innocent lives in such contexts. Other examples of this abound including the medical imagery on which doctors are so dependant today for diagnosis and for intervention.

The problem that we have with the flood of images acquired by our cameras is that their value is mostly wasted if a human is not in the loop. Reducing the role of humans as the only interprets of visual information is the most important goal of the discipline of computer vision within which our research lies. Given that biological vision has proved very difficult to understand let alone to mimic, our only hope for accomplishing automated image understanding lies in the use of the mathematical tools with which we are more comfortable. One of advantages that we have with our technology is that we can extend our perception of the world beyond the visual spectrum. This offers a good opportunity to compensate for our limited success in understanding the world from color alone. At the heart of computer vision lies the task of reconstructing three dimensional computer models that describe objects and scenes. In addition to the geometry of objects it is also very useful to have the color and other multi-spectral information associated with each point of a scene. The spatial alignment, or registration, of different imaging modalities is at the heart of this research. The registration problem emerges because of the limited window through which imaging systems capture the world; limited in both the spectral and spatial domains. The narrow field of view of cameras or mostly the complex topologies of real world objects require us to view them from different positions in order to have a better and more useful description.

For the acquisition of scene geometry different systems were developed. Stereovision systems use the same principals of parallax and triangulation employed by human vision to reconstruct the geometry of objects. Years of intensive research led to some success in several applications such as robotic navigation and manipulation in hazardous environments [Maimone98], or planetary exploration [Goldberg02]. Nevertheless this approach is not yet useful for applications that require precise measurements of the scene. The main problems with stereo are: (1) the matching or correspondence task, which is also at the core of the registration problem, and (2) the rapid decrease in depth accuracy with the increase in distance. Another popular class of 3D scene digitization systems employs laser range scanners which are using either time

of flight or triangulation principles. Laser scanners offer the advantage of accuracy and usually provide a high resolution sampling of static and rigid object surfaces. We will mostly use this last method for our acquisition of scene geometry. In addition to range maps we will employ in our reconstruction framework color and Infra-Red images. The ultimate goal is twofold: to automatically build 3D models with multimodal texture overlay for human interpretation, and more importantly to integrate useful information in the different sensory inputs for fusion and automatic machine decision. Our means to achieve this end is a unified approach to 3D and 2D image alignment based on the optimization of a novel criterion that will be presented, argued for, and analyzed in detail as this dissertation progresses. Applications that we targeted are virtualized reality for simulation and reverse engineering, remote inspection in hazardous environments (mainly for DoE's radioactive waste cleaning), and Biometrics.

1.1 Our Framework

The primary emphasis of this dissertation will be on the registration of multiple free-from shapes for object modeling. The second technical goal is the alignment of multimodal 2D imagery. Both tasks are active research areas and we will present an up-to-date overview of the state of the art in the following chapter. Our work will employ the most basic representation of shapes assuming their description as a cloud of points. Therefore our techniques will belong to the same class as that of the ubiquitous Iterative Closest Point (ICP) algorithm. It is in fact the limitations of ICP that we are attempting to overcome. The latter method while increasingly popular since its discovery (or invention) by Besl and McKay in 1992 [Besl92], has several shortcomings. Our research will concentrate on extending the range of convergence of point-based registration, which is one of the major problems with ICP. While these techniques can ensure an accurate registration when closely initialized they limit the autonomy of current 3D modeling systems by requiring the human to be in the loop.

To achieve our goals we started by devising a robust feature descriptor for general unorganized and noisy point-sets. Several local and global descriptors were proposed for surface-based registration, using mostly differential properties. Given that we work with the most general case of point-sets we will employ a recently developed and powerful tool for feature inference namely tensor voting. The philosophical reason behind this choice, as well as behind the choice of the point-sets representation, is that current 3D digitization devices actually provide point-sets sampled from the surfaces. Surface topology and differential properties are currently just inferred from these samples. In the registration task the goal is precisely to reconstruct these surfaces from the combination of several datasets. The redundancy of information will allow for the accurate recovery of shapes. Hence it is more suitable, when possible, to register the raw point-sets without any processing so that we don't loose irrecoverable shape information. The efficient Tensor Voting framework [Medioni00] will help us in designing a local feature descriptor which will measure the visual importance of points, also known as saliency. This descriptor will prove robust to noise and information rich. It will allow the implicit embedding of salient feature information and confidence in a convenient format suitable for our registration algorithm. It will also have the advantage of being computationally efficient.

Having a good local descriptor such as our point saliency measure will help significantly with the correspondence task, which is the main challenge for registration. In the work that will be described in this dissertation we depart from the ICP criterion and designing a new energy function that quantifies registration. In this we are guided by the limitations of ICP methods. Our development starts with a simple combinatorial matching criterion that is consistent with a rigorous definition of the registration task. A continuously differentiable energy function is obtained from this criterion by the method known as mollification, which is simply a smoothing by convolution with a Gaussian kernel. We will interpret this criterion physically in terms of Gaussian force fields that are exerted by one of the point-sets on the other. The strength of this field

will depend on the distance between the points and on the similarity of their shape and other attributes.

Our formulation proved straightforward to implement, especially given the possibility of using standard gradient-based optimization techniques instead of the specialized heuristic employed by ICP. Even better is the recent availability of a fast computation method for our criterion in the form the Fast Gauss Transform [Elgammal03][Yang03]. This technique allows for the reduction of computational complexity from square to linear. By combining these elements and by studying the properties of the Gaussian criterion we end up with an elegant and practical framework for 3D registration. Quantitative and qualitative analysis will show that our algorithm allows for a significant increase in the range of convergence as compared with ICP, and hence offers a serious alternative to this standard technique. We will also show that the Gaussian Fields criterion can be used for multimodal image registration as well. Practical examples of 3D modeling, mainly photo-realistic object reconstruction and multi-sensor face modeling, will stress our ultimate objective of automating multimodal scene description as well as our application areas. In the process of developing our methods we also addressed in a novel way the important computer vision problem of Shape from Motion. In this effort lie the seeds of both the saliency measure and of the registration approach. The overall multimodal integration pipeline that focused our work is summarized in Fig. 1.1.

1.2 Contributions

In summary, the most important contributions of this work are the following:

- **The Gaussian Fields Criterion for Point-Sets Registration**

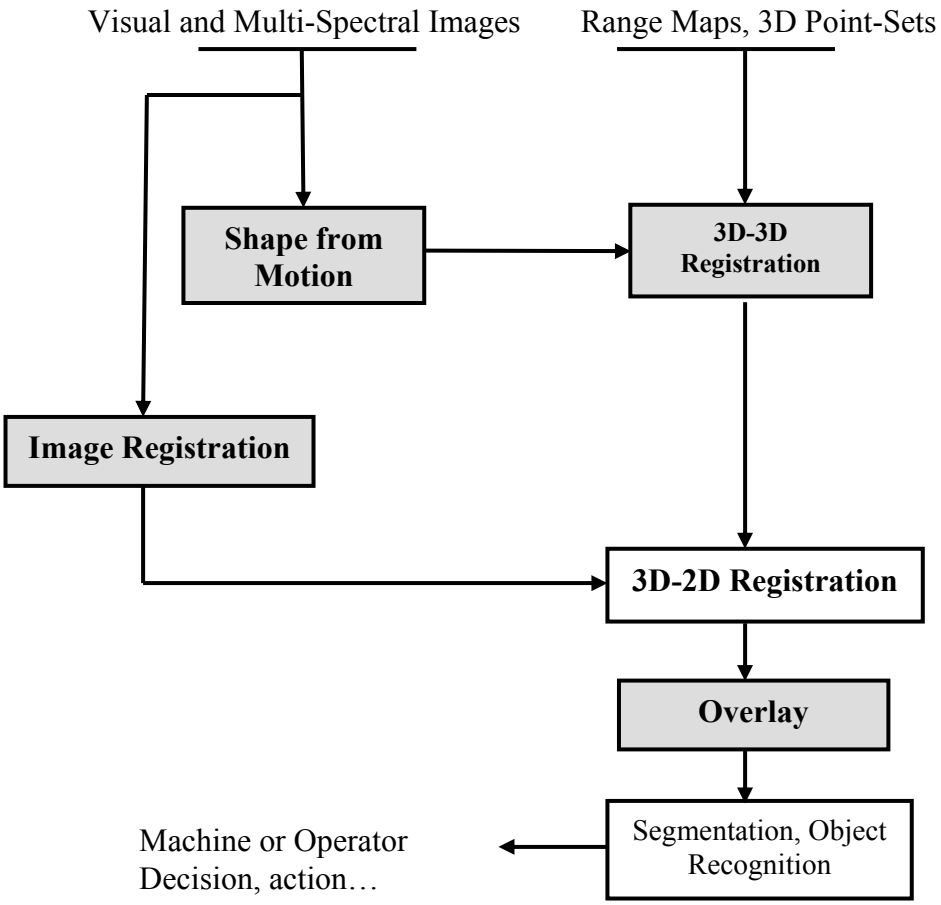


Fig. 1.1. General flowchart of our 3D multimodal modeling system. The shaded boxes are the tasks that were addressed in this dissertation.

In this research we introduce a new optimization-based approach to 3D and 2D point-sets registration allowing the simultaneous recovery of point correspondences and aligning transformations. Our overriding goal is to obtain a differentiable smooth criterion with local convexity in the neighborhood of the aligned position. The new criterion is based on Gaussian Force fields as an efficient device for attracting the two datasets into the registered position. By using a standard quasi-Newton optimization strategy we are able to extend the region of convergence of current automatic methods significantly. Furthermore the Fast Gauss Transform numerical technique offered us a powerful and well developed tool for the efficient implementation of our method. While we emphasized the 3D registration task our framework generalizes easily to N-D point-sets matching and applies directly to feature-based multimodal and single sensor image alignment. We believe that this new approach is an important contribution to the state of the art in both 3D and 2D free-form registration.

- **A New Local Feature Descriptor for Visual Saliency Measure**

In conjunction with the Gaussian criterion we developed a new local feature descriptor based on Tensor Voting. This descriptor was specifically designed with the consideration of differentiability in mind. It is explicitly expressed in terms of the point-sets and is robust to noise and clutter. The local saliency descriptor embeds surface and curve information that is very useful for matching. It is also computationally efficient unlike several other moment invariants. Experimental analysis showed the robustness of this criterion to high levels of noise and its discriminatory power which enhanced the performance of our registration algorithm.

In addition to these two key contributions we also worked on the simultaneous recovery of camera pose and scene structure without search for explicit correspondences [Boughorbel03]. While this work initiated the other ideas on structure saliency and

ultimately contributed to our alignment framework, it was not as thoroughly investigated in this dissertation as the others due to our different focus.

1.3 Document Organization

The remainder of this dissertation is organized as follows:

- Chapter 2 gives a literature review of the topics most relevant to our research. Namely: 3D Free-from registration, image registration, tensor voting, and shape from motion.
- Chapter 3 describes our work on designing the feature saliency descriptor starting from our initial efforts at camera motion recovery and presenting a first attempt at designing a point-set registration criterion.
- Chapter 4 is the core theory chapter that develops the Gaussian Fields framework using mollification and relaxation approaches. It argues theoretically for its advantages, and gives the details of the optimization strategy used.
- Chapter 5 presents Fast Gauss Transform methods and shows its usefulness for our algorithm.
- Chapter 6 contains a thorough experimental analysis on synthetic and real datasets of the 3D registration method. Included are studies of robustness to noise, overlap, feature choice, as well as convergence properties. This chapter shows the results of IR to color registration and examples of multimodal 3D reconstruction.
- Chapter 7 will present a short summary of the dissertation's seminal points, a discussion with concluding remarks, and opportunities for future research.

2 RELATED WORK

While the main thrust of this research is the registration of 3D point-sets under rigid transformations, several other topics were also addressed in the process of developing our new method and in the context of multi-modal reconstruction. These include the tensor voting framework used for robust inference of features, single and multi-sensor image registration, and shape recovery from camera motion. In this chapter we present the relevant literature situating our work within the state of the art.

2.1 3D Free-Form Registration

Due to their limited field of view and to the occlusion problem most 3D imaging systems will provide partial scans of a scene. In order to build complete description of scene geometry, we need to merge together several of these partial views. Since these datasets are originally represented in the local sensor coordinates frame, registration is a fundamental step in most 3D modeling pipelines. In the applications most relevant to our work the views are related by rigid transformations (R,t) : 3D rotations and translations. In some other fields such as medical imaging non-rigid transformations may be needed [Maintz98]. In this case the focus is mostly on fusing different modalities, or on comparing structures with important shape variability [Toga99], for a brief overview of image registration techniques see section 2.2. The free-form shape registration problem was the focus of significant research efforts during the last several years. At the core of many registration approaches lies the classic problem of absolute orientation: recovering the rigid transformations using a set given of 3D correspondences [Faugeras86][Horn87].

In the literature, a common distinction is found between fine and coarse registration methods [Campbell01], which are often used in a two-stage fashion. Coarse or preliminary registration can be obtained by several techniques, including hardware solutions, such as rotating tables, which were used to scan small objects. But more commonly, a set of matches is used for the alignment of the views. These correspondence points can be obtained interactively as in most commercial 3D modeling packages [RapidForm02]. Fully automatic 3D matching using exhaustive constrained search techniques were attempted [Chen99]. However, these methods are computationally expensive and sensitive to noise. Invariant features are more frequently employed to reduce the search space, playing an important role in systems that aim for the automation of the reconstruction process. To achieve fine registration, by far the most widely used algorithm is the Iterative Closest Point (ICP) algorithm and its very numerous variants and extensions. This accurate method is commonly used as a refinement step after an initial coarse registration was obtained.

2.1.1 Registration with invariant features

Given a set of corresponding points between two 3D data sets, Horn [Horn87] derived a closed from solution to the absolute orientation problem. Similar results were also obtained in [Faugeras86]. Automatically establishing the set of correspondences to be used in such algorithm is a common interest to both registration and object recognition tasks. Several feature descriptors were used to represent free-form surfaces and point sets. In the class of global descriptors spherical representations such as the Spherical Attribute Image (SAI), which mapped surface curvature values into a tessellated sphere, were employed for 3D registration [Higuchi95][Hebert95]. Also to this category belongs the work of Lucchese et al [Lucchese02] extending frequency-domain methods to range data registration. Park and Subbarao [Park04] employed the invariant Stable Tangent Plan (STP) for crude registration.

Many local representations were also proposed to encode surface shape information. Stein and Medioni [Stein92] used the variation of the surface normals at local patches for matching, defining a local feature map called Splash. Thirion [Thirion96] extracted critical points and contours using Gaussian curvature extrema. Chua and Jarvis [Chua96] introduced a new local representation called the Point Signature. In this descriptor, a sphere centered at the surface point intersected with the surface resulting in a contour. The distance of these contour points to a plane approximating the tangent plane at the point is parametrized and stored. This approach results in a pose invariant description of the local shape information. Spin Images are another popular invariant local representation, proposed by Johnson et al [Johnson97]. These are basically histograms of distances and angles of neighboring surface points, which can be used efficiently for registration and recognition. Belonging to this same class of local descriptors, is the Monge patches used by Wyngaerd and Van Gool for rough alignment [Wyaengard02]. Most of these features rely on local normal or curvature information. Since we are mostly focused on the registration unorganized noisy point-sets, where surfaces were not defined yet, differential surface attributes will not be used from the start. Nevertheless, using a voting process we are able to infer local surface information in noisy and sparse datasets.

2.1.2 Iterative Closest Point algorithms

The ICP algorithm was first introduced by Besl and MacKay in [Besl92]. Its basic version aligns a set $S = \{s_1,...,s_{N_s}\}$ of 3D scene points with a geometric model $M = \{m_1,...,m_{N_m}\}$, by minimizing the sum of the squared distances between the scene points and the model. For every point $s_i \in S$, the distance to M is defined as: $d(s_i, M) = \min_{m \in M} \|s_i - m\|$.

The algorithm is summarized in the following:

1. Start with an initial transformation (R_0, t_0).
2. For $k = 1,...,k_{\max}$ or until stopping criteria met do:

   2.1. Compute $s_i^{k-1} = R_{k-1} s_i + t_{k-1}$.

   2.2. Build the correspondence set $C^{k-1} = \bigcup_{s_i \in S} \{(s_i^{k-1}, \arg\min_{m \in M} \| s_i^{k-1} - m \|)\}$.

   2.3. Using the pairs C compute the transformation that minimizes the sum of squared distances [Horn87].

The ICP algorithm was shown to converge monotonically to a local minimum. Therefore, the initial estimate of the transformation should be sufficiently close to the correct registration. Another limitation of the original version is that it requires large overlap between the datasets to be aligned. Step 2.3 in the algorithm is commonly solved by one of the feature matching techniques presented in 2.1.1. Independently, Chen and Medioni [Chen92] developed an algorithm similar to ICP. Using orientation information, they devised a least squares matching metric based on the distance, in the direction of the normal to a scene point, to the tangent plane at the model. This approach allowed the incorporation of local shape information, as well as the handling of partially overlapping datasets.

Active research on the free-form registration problem gave a rise to a proliferation of improved ICP algorithms. Zhang [Zhang94] proposed a method based on robust statistics, allowing for better handling of outliers, occlusions, and partial overlap. The algorithm used heuristics to remove inconsistent matches. While the ICP algorithm was first used in the context of registration of clouds of points, Turk and Levoy [Turk94] devised a modified registration metric that dealt with polygon meshes. Their technique is able to register partially overlapping views, by imposing mesh-based constraints on the selection of nearest points. In other improvements of ICP, Masuda

and Yokoya [Masuda95] used a Least Mean Squares (LMS) error measure that is robust to partial overlap, and Dorai et al [Dorai97] proposed a Minimum Variance Estimate (MVE) of the registration error, that produced less error than the least squares error measure. In the same effort at robustness Trucco et al [Trucco99] employed Least Median Squares (LMedS), in the so-called Robust ICP version (RICP), to register noisy point-sets with missing data. Zinsser et al [Zinsser03] designed an algorithm coined the Picky ICP which combined the strengths of different currents variants to achieve both robustness and increased speed. Also for reducing the speed of ICP other approaches were also investigated, such as the use of k-D trees to handle the datasets [Zhang94], or the use of spatial subdivisions to partition mesh vertices [Turk94]. These data structures helped speed the search for the nearest point and reduced the computational cost significantly. Volumetric constraints and Voronoi diagrams were at the core of the fast Morphological ICP proposed by Kapoutsis et al [Kapoutsis99]. For points sets-to-surface registration a recent work by Pottmann et al [Pottmann04] devised a technique based on instantaneous kinematics, achieving substantial gains in speed over ICP methods.

Going for the two 3D views to the case of multi-views registration, it was clear that using the original ICP in a sequential fashion may lead to error propagation. Hence, the development of several techniques that attempted to minimize, and also to balance, the error distribution in the registration of multiple views. Turk and Levoy [Turk94] used a reference cylindrical scan of the object and registered all the other views to it. Blais and Levine [Blais95] used the camera calibration to project pixels from one view onto the other range images. The resulting optimization problem was approached through a simulated annealing technique starting from the initial transformations. In yet another extension of the ICP algorithm, Bargevin et al [Bargevin96] showed that the transformations could not be decoupled and proposed a technique that handled multiple range images simultaneously. The method balances the registration errors between the views resulting in an overall error less than the sensor noise. Several

heuristics were also used in this work to achieve robustness to missing and unreliable data. Eggert et al [Eggert96] used both point position and normal information in the search for correspondences. All the views were represented in a single coordinate frame, and the search was performed on the combined point sets.

Using analogy to physics Eggert et al [Eggert96], and stoddart et al [Stoddart96] devised a force-based optimization method to register the data sets. The closest points were connected by springs assuring better convergence to local minima. Formulating the multiple-views registration problem in graph theoretic framework, Neugebauer et al [Neugebauer97] represented the different datasets as nodes and the transformations as links. The registration task in this approach consisted of finding a network with balanced registration errors. For improved computational efficiency, a multi-resolution hierarchical approach was adopted, where the number of points used at the start is low and increases as the registration proceeds. A similar graph theoretical algorithm was proposed by Huber and Hebert [Huber03]. Recently a hybrid algorithm employing invariant features was used to enhance the performance of the standard ICP [Sharp02]. The search for closest matches was performed in the extended positional and feature space. For recent surveys and comparisons of several common variants of the ICP algorithm the following references [Dalley02][Rodrigues02][Rusinkiewicz01] can be consulted.

Several researchers studied the relationship between surface shape complexity and the registration accuracy. In this context, Pennec and Thirion [Pennec97] characterized the uncertainty of point set registration. Stoddart et al [Stoddart96] defined a registration index measuring shape information, and Brujic and Ristic [Brujic96] used Monte Carlo simulation to study the dependence of accuracy and complexity. To enhance the alignment in the case of scenes with low shape information content, Pito [Pito97] designed a registration aid that can be placed with the scanned objects ensuring precise matching.

Several of the limitations of the ICP framework stem from the non-differentiability of its cost function which imposed the use of a specialized heuristic for optimization. Addressing the registration in the context of gradient-based optimization has attracted some interest recently. We mainly refer here to the work of Fitzgibbon [Fitzgibbon03] who showed that a Levenberg-Marquardt approach to the point-set registration problem offered several advantages over current ICP algorithms. The proposed method used Chamfer distance transforms [Borgefors88] to compute derivatives for the ICP criterion. In addition robust estimation via a Huber kernel [Huber81] was employed with the effect of significantly widening the basins of convergence of existing techniques. This work is one of the most closely related to our efforts in 3D registration, since we also aim at designing a criterion that can be optimized through general purpose non-linear techniques. The main disadvantage of Fitzgibbon's technique is that it is limited to datasets on a grid, where the Chamfer distance transforms and discrete derivatives are easily evaluated. For sparse unorganized point-sets this method cannot be directly applied. Our work is also in the same class of techniques as the one by Charpiat et al [Charpiat03] which approximates the Hausdorff distance with a differentiable metric on shape space. The resulting Hausdorff warping can be used for PDE-driven fitting and recognition of shapes.

By introducing a new registration method that uses a straightforward differentiable cost function, directly and explicitly expressed in terms of the point coordinates and the registration parameters, we are able to overcome several problems with the ICP-based methods described above. The smooth behavior of our registration criterion, as well as other characteristics which will be discussed in this dissertation, combined with the use of a standard optimization scheme extends the range of convergence. The application of the technique to the multiple views case is straightforward, and the registration metric can be extended to cases of non-rigid registration. Given the generality of our framework we can apply the Gaussian Fields method to single and

multi-sensor image registration tasks In the following section we briefly present previous work done in this field.

2.2 Image Registration

Intensively researched, image registration has a wide range of applications in areas such as pattern recognition, medical imaging, and remote sensing. In the case of single sensor registration the purpose is to combine several images in order to overcome the limited view of the camera. Multimodal registration is mainly the step preceding fusion, where the fused information is exploited for recognition and decision. The two main components that define registration methods are the measure of similarity between the images or between smaller areas of the images (mostly square windows), and the transformations modeling the mappings aligning them.

In the case of single sensor registration similarity measures belonging to the class of correlation measures such as normalized cross-correlation with its different variants, the sum of squared differences (SSD), and the sum of absolute differences (SAD), have been used for a long time. A method belonging to this class, but that was used for multimodal imagery, is the Correlation Ratio (CR). Roche et al [Roche98] give a good comparison of the CR with other multi-modal similarity measures. Some more sophisticated criteria can be built from correlation measures, such as in the work of Irani et al [Irani98] on Infra-red and electro-optical image registration, where a global criterion was obtained by summing the local cross-correlation measures of small patches in extracted energy images. This approach does not require the global statistical correlation of the images, which violated in most cases of multimodal imagery, but just the local one. Beside the assumptions of the statistical relations between the images correlation techniques may suffer from a flat similarity measure, requiring some sharpening through the use of edge of other feature maps.

Another important class of similarity measures is based on Fourier analysis and phase correlation. It is mostly used when the images are acquired under significantly varying conditions, or when they are corrupted with correlated and frequency dependant noise. The main issue in this class of methods is the type of transformations used. De Castro and Morandi [De Castro87] showed the case of alignment with translations and rotations. The use of Fourier-Mellin Transform [Chen94] allowed the recovery of scale as well. The third important family of similarity measures is the one based on Mutual Information; a widely popular technique in medical imaging. Introduced by Viola and Wells [Viola97] where it was applied to the registration of MRI images and to the alignment of 3D models with images and employed a gradient descent optimization approach. The method was almost simultaneously discovered by Maes et al [Maes97]. This similarity measure is derived from information theory and uses the concept of entropy to measure the statistical dependence between the images. For a comparison of mutual information with other image registration criteria see [Roche00][Penny98]. A related information theory method called cross entropy was recently used for volumetric image registration by Zhu [Zhu02].

Most of these measures were used for area-based registration. These methods are suited for when we don't have prominent local features, another drawback mentioned above is the assumption of global statistical dependence. Area-based similarity measures could be also applied to continuous feature maps, computed in a pre-processing step, in order to reduce the differences between multimodal images or images with wide illumination differences. An example is the directional derivative images employed by Irani et al [Irani98] for multi-sensor registration, and the local frequency maps used by Liu et al [Liu02]. Another approach, relevant to our work presented in chapter 6 on multimodal registration, is to use binary images obtained by thresholding the feature maps. Huttenlocher employed the Hausdorff measure and showed that it performs better than cross correlation for this task [Huttelocher93].

To model the transformations aligning the images several approaches were used. Global parametric transformations, such as affine and projective warps, assume that a single set of parameters will be valid for the whole image. More accurate local parametric mappings were employed for registration and image warping [Zitova03]. Other methods while global in nature are capable of modeling local deformations, these include radial basis functions techniques, chief among which are Thin-plate Splines methods [Bookstein89]. While these above methods were mostly used in the case feature correspondences are available, other techniques recover both correspondences and transformations at the same time. For example another approach used in the case of images with complex and/or local deformation is elastic registration, introduced by Bajcsy [Bajcsy89]. These techniques modeled the images as pieces of an elastic rubber sheet that are subject to external stretching forces and internal smoothness and stiffness constraints that bring them into alignment with minimal amount of bending and stretching. The external forces are derived by optimization of a similarity function defined on the intensity values or on the boundary structures.

In the case of very localized deformations registration can be addressed by the so-called fluid registration, where a viscous fluid model was used to model the flow of one image in the process of aligning with the reference image [Woolny02]. In addition other non-rigid methods are commonly used including diffusion registration [Thirion98][Andersen01], and PDE-driven level sets [Hermosillo02]. For a thorough recent survey on image registration techniques we refer to Zitova and Flusser [Zitova03]. In this research the application of our new matching criterion to image registration was somewhat limited when compared to our work on 3D alignment; nonetheless we are convinced that our framework will be very useful for both multimodal and single-sensor image alignment under different non-rigid transformation models.

2.3 Tensor Voting

The framework of tensor voting was introduced in [Guy97][Medioni00] as an efficient tool for robust feature extraction in noisy datasets. At its core lies the salient feature inference engine that encodes both feature information and feature confidence in a unified and intuitive representation. Since the usual first-order description (vector representation) of features is not able to handle both aspects, in the tensor voting approach, a second order representation based on tensors was adopted, offering a convenient way to encode the visual importance of features called saliency. We will later explain in more details the formalism used for this representation (chapter 3). Additionally, tensor representation allows for an easy propagation of feature information from one site to neighboring sites through the process of Tensor Voting, a process that relies on the collection of votes at a given site using tensor summation. The basic features that were considered in 2D and 3D are point (junction and isolated point), curve, and surface elements.

Tensor Voting will be used extensively in this work for inferring robust features. Its implementation can be thought of as a convolution operation. At the end of the voting step, dense saliency maps are created for the different features: junctions, curves, and surfaces. Building the saliency maps is achieved by decomposing the collected votes at each site into a collection of basic tensors the combination of which can generate any general tensor. The next step will be the extraction of features as local extrema in the saliency maps. The recovery of continuous curves and surfaces is done through marching techniques in the dense domain.In the case of non-oriented point sets, which will be our primary data in this research, a preliminary voting step will be performed to associate orientation to the points. In this step, the voting will be performed along straight lines joining the sites and using the basic so-called stick field (Fig. 2.1).

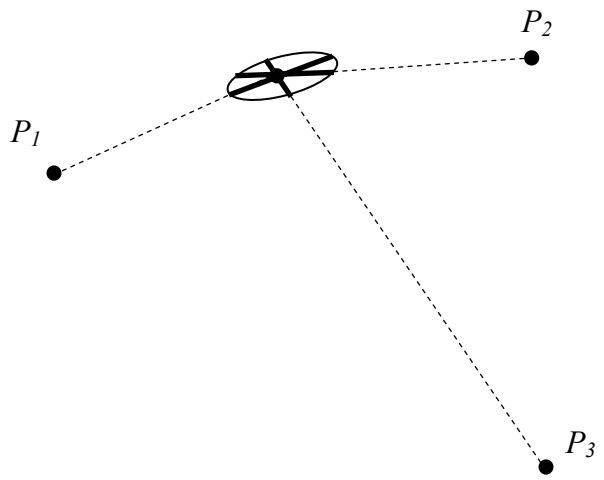


Fig. 2.1. Tensor voting in the case of non-oriented points. The first pass: the votes are represented by the sticks directed along the lines joining the sites (and encoding tangents to curve elements), and the lengths of the sticks decrease for farther voting sites. The ellipsoid reflects the orientation uncertainty of the tangent at the site. It is also the geometric expression of the resulting tensor sum.

Since we are mainly working with non-oriented points the technique that will be described later in this dissertation is based on this first pass algorithm. Tensor voting for the extraction of features in noisy datasets was shown to be a robust, fast, and intuitive method. It is also one of the few techniques that can handle sparse point-sets which contain large gaps of missing data. The method has a remarkable generality that initiated applications to many areas of computer vision. Applications include solving for epipolar geometry estimation using 8-Dimensional Tensor Voting [Tang01], the recovery of motion layers [Tong04][Nicolescu03], the tracking of segmented objects in image sequences [Kornprobst00], image repairing [Jia03], and color and texture segmentation [Jia04]. The tensor-voting framework was also extended to handle polarity information in addition to direction [Tong01].

2.4 Shape from Motion

The roots of the Shape from Motion task can be traced back to the early work of photogrammetrists at inferring measurements in a scene from photographs [Slama80]. With the emergence of the field of computer vision, and increased interest in the three dimensional reconstruction of objects from images, a significant research effort was devoted to SFM [Jebara99]. The main objective of SFM is to use feature correspondences between images, mainly point-features, in order to infer the motion of the imaging camera and then to reconstruct the matches through triangulation. Shape from Motion algorithms usually lead to sparse structure estimation, but a dense stereo matching step is added in many systems [Pollefeys99].

The first step in any automatic motion and shape recovery from image sequences is the tracking of features, a delicate task in itself, which is still an open research area [Schmid00]. Once the correspondences recovered, several approaches were adopted to tackle SFM but we can generally divide the methods into linear and nonlinear classes. Although the SFM problem is commonly considered as a non-linear problem, projective geometry was used for the modeling of the multiple views formation in the framework of linear algebra. In the linear methods the relationship between features across images is encoded in algebraic structures such as the Fundamental Matrix [Faugeras92], which relates corresponding points in two views, and Trifocal tensor encoding correspondence between three views, as well as other higher order tensors that relate N views [Hartley00]. Linear methods are elegant mathematically and convenient computationally, unfortunately, they are very sensitive to noise and are in practice followed by a nonlinear refinement step. Non-linear methods define and minimize a geometric cost function either in the image space in a bundle adjustment fashion, or in the 3D Euclidean space [Faugeras93]. To handle image sequences effectively these methods were also implemented in the framework of recursive estimation [Azarbayejani95][Soatto98].

A recent research trend focused on un-calibrated cameras and several interesting results were obtained [Hartley00][Faugeras01]. They show that in several cases and given the right set of correspondences the internal parameters of cameras can be recovered in an auto-calibration module, the work is based on projective geometry stratification and invariants. In our implementation of Shape from Motion we assume calibrated cameras in order to have the highest possible accuracy.

Most of the work mentioned previously considered the tasks of motion recovery and structure reconstruction separately and sequentially. An important exception is the class of algorithms known as the Factorization methods [Kanade97]. In this approach, a closed form solution of the SFM problem is sought, where shape and motion are obtained simultaneously. It was obtained at the cost of simplified camera models (resulting in approximate reconstructions) such as the orthographic [Tomasi92], and paraperspective models [Poelman94]. The Factorization algorithm was extended to handle features other than points such as lines and planes [Morris98][Quan96], and to the case of multiple moving objects [Costeira95]. There is also continuing work on adapting it to perspective camera models, although in these cases iterative techniques are usually required [Han99].

3 A FEATURE SALIENCY DESCRIPTOR FOR REGISTRATION AND POSE RECOVERY

In this chapter we present our derivation of a new and effective local measure of visual saliency that we will use as point-feature descriptor for registration. The organization of the chapter reflects the gradual way and the thought process through which we came to this new measure. Our first goal was the challenging of recovering camera motion without search for explicit correspondences. Work on this problem led us to the idea of quantifying the visual saliency of a given 3D reconstruction. Using the framework of Tensor voting we were able to derive a straightforward local measure of point saliency. The next step was to employ this measure for 3D point-sets registration, which we approached from the principle structure saliency maximization. This first formulation showed several advantages, but also important limitations, which we overcame by further simplification in the formalism leading to the Gaussian Fields registration technique that will be presented in the next chapter.

3.1 Tensor Encoding of Features

The definition of saliency tensors as presented in Medioni et al [Medioni00] follows from the common representation of orientation uncertainty by an ellipsoid in 3D and an ellipse in 2D (Fig. 3.1). The uncertainty ellipsoid is a geometric description of the covariance matrix T associated with vectors such as the tangents to curves and normals to surfaces.

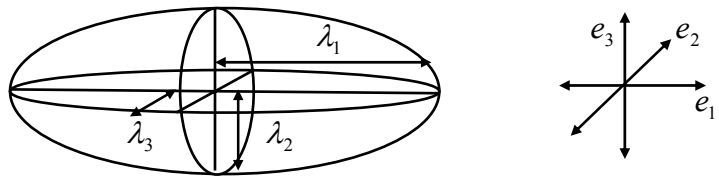


Fig. 3.1. The ellipsoid describing orientation uncertainty and the associated eigen-decomposition.

In 3D the matrix T can be decomposed into its eigenvalues $(\lambda_1, \lambda_2, \lambda_3)$ and eigenvectors (e_1, e_2, e_3) as follows:

$$T = \begin{bmatrix} e_1 & e_2 & e_3 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \begin{bmatrix} e_1^T \\ e_2^T \\ e_3^T \end{bmatrix} \tag{3.1}$$

So that $T = \lambda_1 e_1 e_1^T + \lambda_2 e_2 e_2^T + \lambda_3 e_3 e_3^T$, where $\lambda_1 \geq \lambda_2 \geq \lambda_3$ are the principal axes of the orientation ellipsoid. The saliency of a feature is determined by the size and shape of the uncertainty ellipsoid and depends directly on these eigenvalues. As expressed in (3.1) T has the characteristic representation of a tensor.

Point features are encoded using the so-called *ball* tensor, geometrically described by a circle in 2D and a sphere in 3D. Curve elements (*Curvels*) in 2D, consisting of the pair point + tangent vector, are encoded using the covariance matrix of the tangent

vector in a tensor known as the *stick* tensor. In 3D a curve element is represented by a *plate* tensor encoding the uncertainty of normal orientation. Surface elements (*Surfels*) are represented by the covariance of the normals to the surface. Table 3.2 summarizes the encoding of the different features which will be employed later. The voting process will usually introduce tensors that are not as singular as the ones shown in Table 3.2.

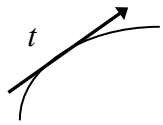
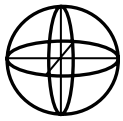
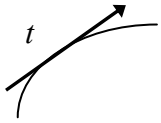
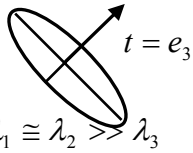
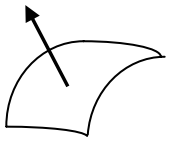
In general any saliency tensor S can be decomposed as combination of these basic saliency tensors as follows:

$$S = (\lambda_1 - \lambda_2)e_1 e_1^T + (\lambda_2 - \lambda_3)(e_1 e_1^T + e_2 e_2^T) + \lambda_3 (e_1 e_1^T + e_2 e_2^T + e_3 e_3^T) \qquad (3.2)$$

where $e_1 e_1^T$ represents a *stick*, $e_1 e_1^T + e_2 e_2^T$ a *plate,* and $e_1 e_1^T + e_2 e_2^T + e_3 e_3^T$ a *ball*. It follows from this decomposition that at each point $\lambda_1 - \lambda_2$ is the component measure of *surface-ness*, $\lambda_2 - \lambda_3$ a measure of *curve-ness* and λ_3 a measure of *point-ness* [Medioni00].

The input for the feature inference algorithms described in [Medioni00] consists of sparse tokens, usually points with or without associated orientation. These tokens are encoded in terms of tensors as shown above. Tensor voting is conducted by propagating the saliency information using *voting fields*. Each site in the data will cast a vote at the other sites using an associated voting function. Different voting fields correspond to the *ball*, *plate* and *stick* components of the saliency tensor. The design of the fields was done by taking in consideration perceptual organization principles and using analogies to physical models of potential fields. The votes will decay exponentially with distance to account for higher influence of neighboring sites. A detailed description of the derivation of the voting functions is presented in [Guy97][Medioni00].

Table. 3.2. Encoding of basic features in 2D and 3D in the tensor framework [Medioni00].

Encoding of 2D tokens		
Point	**Disk**  $\lambda_1 \cong \lambda_2 \cong \lambda$	$\begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$
 ***Curvel***	***Stick***  $\lambda_1 >> \lambda_2$	$\begin{bmatrix} t_x^2 & t_x t_y \\ t_y t_x & t_y^2 \end{bmatrix}$
Encoding of 3D tokens		
Point	***Ball***  $\lambda_1 \cong \lambda_2 \cong \lambda_3 \cong \lambda$	$\begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix}$
 ***Curvel***	***Plate***  $\lambda_1 \cong \lambda_2 >> \lambda_3$	$\begin{bmatrix} 1-t_x^2 & -t_x t_y & -t_x t_z \\ -t_x t_y & 1-t_y^2 & -t_y t_z \\ -t_x t_z & -t_y t_z & 1-t_z^2 \end{bmatrix}$
 ***Surfel***	***Stick***  $\lambda_1 \cong \lambda_2 >> \lambda_3$	$\begin{bmatrix} n_x^2 & n_x n_y & n_x n_z \\ n_x n_y & n_y^2 & n_y n_z \\ n_x n_z & n_y n_z & n_z^2 \end{bmatrix}$

3.2 Camera Motion Estimation by Search of the Most Salient Structure

3.2.1 Recovering motion parameters from point matches

The basis of most perspective camera motion recovery algorithms is the *coplanarity* constraint (Fig. 3.3), relating a 3D world point M to its projections in two images. Because the second camera's reference frame is related to the first one by the rigid transformations (R,t) : the 3D rotation and translation, the coplanarity constraint imposes that the rays m_1 and m_2, pointing from the cameras optical centers to the world point, be located in the same plane. Algebraically this can is expressed as follows:

$$m_1.(t \times Rm_2) = 0 \tag{3.3}$$

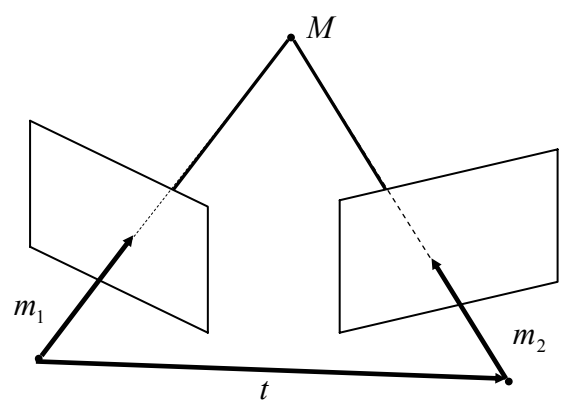


Fig. 3.3. The co-planarity constraint implies that the two rays m_1 and m_2 pointing from two cameras optical centers toward the world point M lie on the same plane with the baseline vector $t = (t_x, t_y, t_z)^T$.

The common approach to computing the motion parameters (R,t) uses a two-stage technique: starting with a linear method, followed by a non-linear refinement step [Faugeras93][Higgins81]. In the linear method the coplanarity constraint (3.3) is rewritten as:

$$m_1^T E m_2 = 0 \tag{3.4}$$

where E is a 3x3 matrix, known as the *Essential Matrix*, that embeds the motion parameters: $E = [t]_\times R$, with $[t]_\times = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}$ the anti-symmetric matrix representing the cross product in a linear fashion. A system of homogeneous equations in the entries of E is then set from the point matches. Commonly more than 8 matches are used resulting in an over-constrained system, which is solved using singular value decomposition (SVD). The actual motion parameters are easily computed from the essential matrix. The scale of the translation, and hence of the reconstruction, is inherently ambiguous in the SFM problem, therefore we have only five independent parameters to recover: $(\varphi_x, \varphi_y, \varphi_z, t_y, t_z)$, the three rotation angles and two translations. Unless some knowledge about the actual dimensions of the scene is available [Morris01], the scale will be arbitrary.

For the non-linear refinement step several formulations were proposed, where the rotation matrix was parametrized as an orthonormal matrix or using unit quaternions [Horn90]. A classic nonlinear method is based on the least squares minimization of the Longuet-Higgins cost function [Faugeras93]:

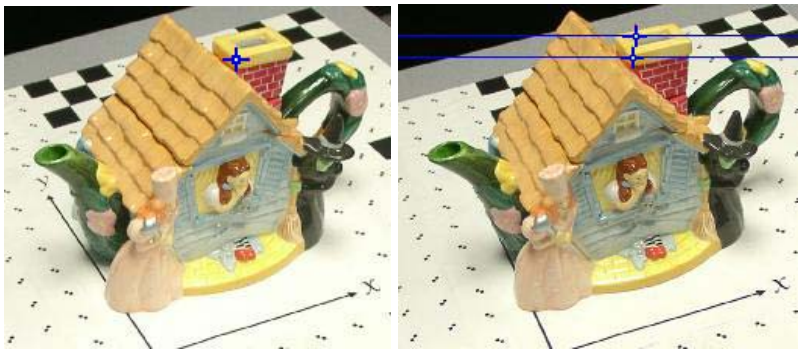
$$\min_{R,t} \sum_{i=1\ldots N_{matches}} \left\| m_1^{\,i} . (t \times R m_2^{\,i}) \right\|^2 \tag{3.5}$$

After the recovery of the motion parameters, 3D points are reconstructed through robust triangulation from matches in two or more images. In addition to the two views case, extensive work was done recently on computing the relationship between multiple views within the framework of projective geometry [Hartley00][Faugeras01]. In these cases relations between the views are described by multi-linear algebraic structures such as the 4trilinear tensor used for three images.

3.2.2 SFM without Correspondences

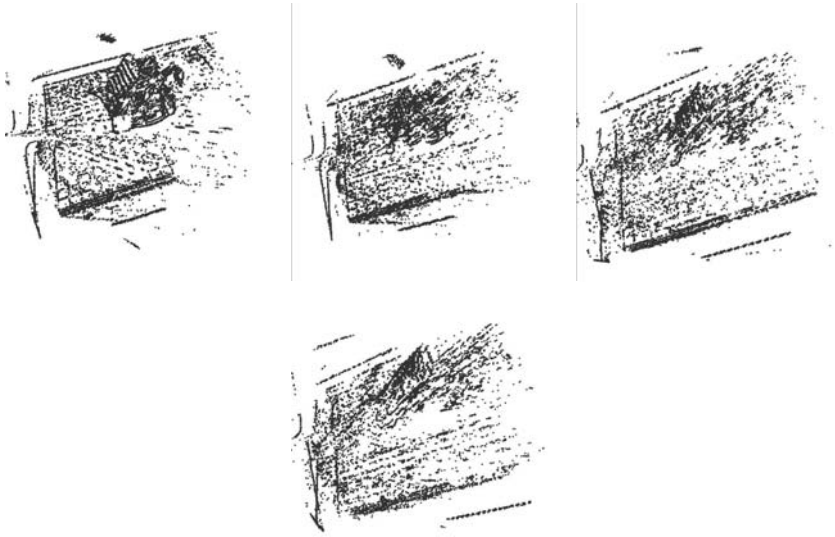
We outlined earlier (literature review section) Dellaert's approach to the recovery of camera motion without prior knowledge of point correspondences [Deallert00]. We will describe here a different approach that we developed to address this problem [Boughorbel03]. The idea behind our method is based on the following facts: (1) If the 3D pose parameters, relating two perspective views of a scene are at hand, it is known that we can use the *epipolar* geometry (Fig. 3.3) to constrain the search for correspondences to a line search. At the correct camera pose the structure reconstructed from these matches (using for example window based matching) will be close to the actual scene geometry, although it can be noisy. (2) If we are far from the correct pose parameters, relying on the epipolar geometry will lead to false matches since the search for correspondences will be done in the wrong area (Fig. 3.4). In this case, the reconstruction will be close to a random cloud of points.

Fig. 3.5 shows the reconstructed structure, resulting from edge-based matching, as the relative rotation parameter φ_x is moved away from its correct value. The set of triangulated 3D points is exhibiting the expected increasing disorder. The key to exploiting this fact for pose estimation is to quantify how structured the set of points is. Or to use another term how *salient* a set of points is.



(a) (b)

Fig. 3.4. The search for a match for the point in the left (a) image is limited by the epipolar constraints to a line search. For the correct pose parameters we are more likely to find the correct match (lower point in (b)). But in the case of wrong pose parameters the matching process leads to wrong correspondences (upper point in (b)). (Images from CVPR01 calibrated test dataset, http://vision.cse.psu.edu/cvpr2001/main1.html).



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

Fig. 3.5. The degradation of the structure reconstructed from the images of Fig. 3.4 (from. (a) to (d)) as we move away from the correct pose parameters.

Compared to the work done in [Dellaert00], this approach reduces the size of the parameter space to be searched. In fact, in our case the focus is mostly on recovering the pose parameters rather than on estimating accurately the structure, which can be done later in a refinement step. So, in our method instead of searching for the structure and pose at the same time we only try to find the pose parameters, greatly reducing the complexity of the algorithm. Furthermore, once we devise a simple structure-saliency metric we can avoid the complex probability distributions arising in [Dellaert00].

3.2.3 Tensor Voting for Structure Saliency Estimation

The voting process that we presented earlier (sections 2.3 and 3.1) can be thought of as a convolution operation ($*$) with a digital mask T_v. If Σ is a set of 3D points, in our case reconstructed from two images at a given relative pose of the cameras, and Σ_v is defined by:

$$\Sigma_v = \{(P, S(P)) : P \in \Sigma\} = T_v * \Sigma \tag{3.6}$$

representing the points with their associated saliency S after the voting. Then the number of points in the reconstruction which belong to a surface, and are not isolated points, can be computed as follows:

$$N_{Surf} = \sum_{P \in \Sigma_v} \delta_{Surf}(P, S(P)) \tag{3.7}$$

where $\delta_{Surf}(P, S(P))$ is defined by:

$\delta_{Surf}(P) = 1$ if P is voted as a surface point,

$\delta_{Surf}(P) = 0$ otherwise.

The function $\delta_{Surf}(P)$ depends on the eigenvalues of $S(P)$ and employs the appropriate thresholds to determine points that belong to two types of salient structures: surfaces and curves as explained in [Medioni00]. Fig. 3.6 shows the 3D reconstruction for different pose parameters, starting at the correct parameters at the left and gradually changing one of the rotation parameters. In the lower sequence are shown the resulting salient points after the tensor voting and feature inference process. Isolated points were removed. The actual variation of N_S for this sequence is shown in Table 3.7. We see that N_S can be used as a point-set saliency measure and employed in a pose recovery framework.

This criterion is not expressed analytically in terms of the datasets and the transformation parameters and optimization will have to be heuristic. But the notion of evaluating the saliency of a set of points was the important idea that will be exploited for our core problem of general point-sets registration. In the next section we will shift our focus and present our derivation of an analytically expressed saliency criterion for the purpose of point-sets and free-form shape registration. While the main emphasis was first on the global saliency measure of a set of points, the local descriptor turned out to be more useful in and of its self.

3.3 The Local Saliency Descriptor

Our purpose at this point is to design a simple local feature descriptor which will be based on the principles of tensor voting. Being based on a robust feature inference framework the new descriptor is expected to have a good performance in the presence of noise and in the case of sparse and non-uniformly sampled datasets. This performance will be confirmed later in through experimental analysis. In this work we are mainly interested in the basic case of non-oriented clouds of points. To derive the expression of our descriptor we employ the first pass of Tensor Voting.

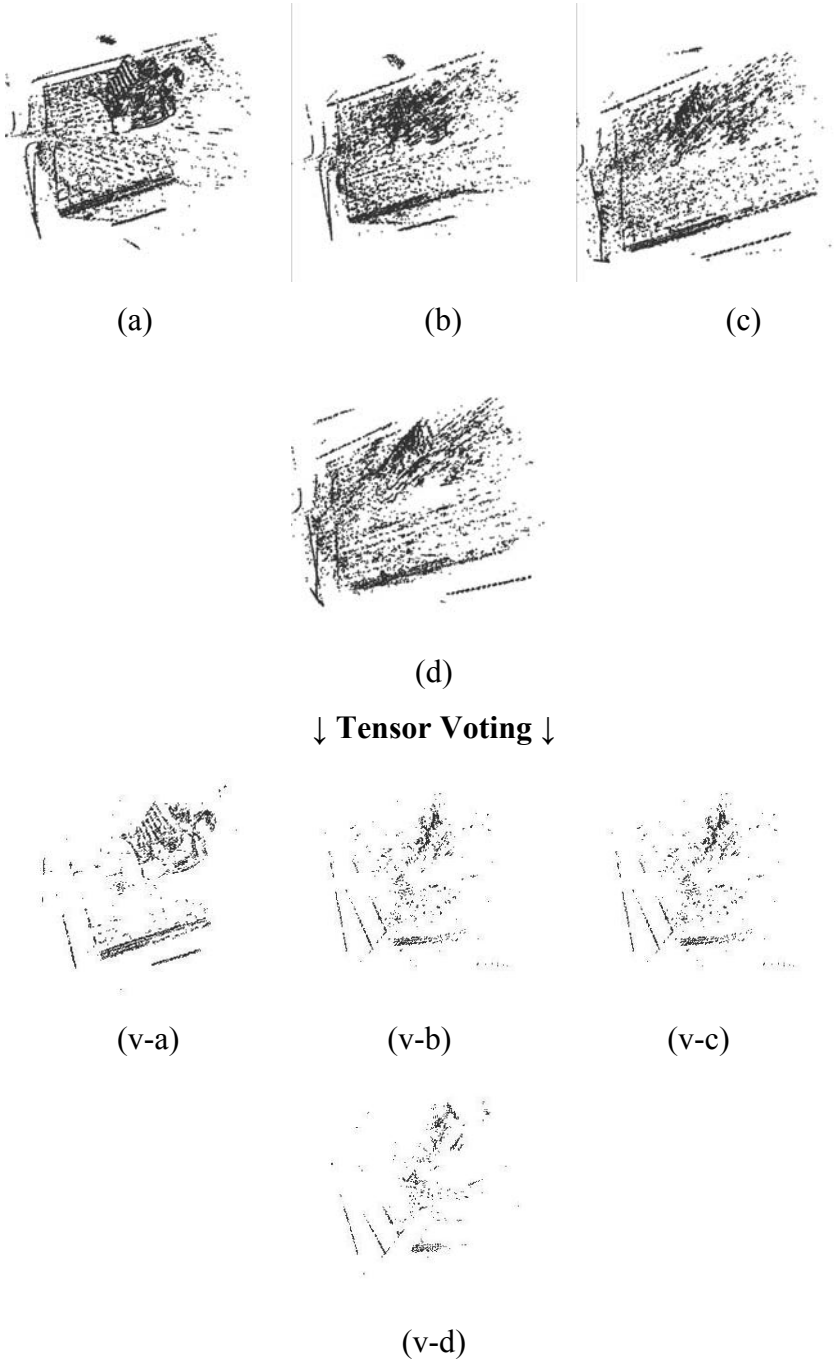


Fig. 3.6. Applying tensor voting and removing non-salient feature-points (v-a to v-d) from the reconstructed point-sets (a to b) for different pose parameters. Results for the correct pose are shown in (a) and (v-a).

Table. 3.7. Variation of N_S with φ_x: the number of salient points will decrease significantly as the quality of the reconstruction degrades (φ_x^0: correct rotation).

φ_x	φ_x^0	$\varphi_x^0 + 2^0$	$\varphi_x^0 + 4^0$	$\varphi_x^0 + 6^0$
N_S	1930	858	425	224

We recall that in the 2D case the voting process for a point-set P is performed using a stick field, which encodes unit vectors pointing from one site $X_i = (x_i, y_i)^T \in P$ to another site $X_j = (x_j, y_j)^T \in P$ where the vote is being cast. Following the formalism described in [Medioni00] the unit vector pointing from the first site to the second (indexed i and j) $t^{ij} = \dfrac{X_i - X_j}{\|X_i - X_j\|} = (t_x^{ij}, t_y^{ij})^T$ is encoded in tensor representation using the covariance matrix of equation (3.8).

$$T_{ij} = \begin{bmatrix} (t_x^{ij})^2 & t_x^{ij} t_y^{ij} \\ t_x^{ij} t_y^{ij} & (t_y^{ij})^2 \end{bmatrix} \tag{3.8}$$

$$= \frac{1}{(x_i - x_j)^2 + (y_i - y_j)^2} \begin{bmatrix} (x_i - x_j)^2 & (x_i - x_j)(y_i - y_j) \\ (x_i - x_j)(y_i - y_j) & (y_i - y_j)^2 \end{bmatrix}$$

As we mentioned above the strength of the votes is made to decay exponentially with distance. This decay is controlled by a parameter σ, which is related to the scale of the

dataset and to the desired smoothness in the variation of the descriptor from one site to neighboring ones. At one site *i* where local votes are collected the resulting tensor sum is:

$$T_i = \sum_{X_j \in P} \exp(-\frac{\left\| X_i - X_j \right\|^2}{\sigma^2}) T_{ij} \tag{3.9}$$

$$= \begin{bmatrix} \sum_{X_j \in P} \exp(-\frac{\left\| X_i - X_j \right\|^2}{\sigma^2}) \frac{(x_i - x_j)^2}{(x_i - x_j)^2 + (y_i - y_j)^2} & \sum_{X_j \in P} \exp(-\frac{\left\| X_i - X_j \right\|^2}{\sigma^2}) \frac{(x_i - x_j)(y_i - y_j)}{(x_i - x_j)^2 + (y_i - y_j)^2} \\ \sum_{X_j \in P} \exp(-\frac{\left\| X_i - X_j \right\|^2}{\sigma^2}) \frac{(x_i - x_j)(y_i - y_j)}{(x_i - x_j)^2 + (y_i - y_j)^2} & \sum_{X_j \in P} \exp(-\frac{\left\| X_i - X_j \right\|^2}{\sigma^2}) \frac{(y_i - y_j)^2}{(x_i - x_j)^2 + (y_i - y_j)^2} \end{bmatrix}$$

Geometrically the accumulated votes will be represented by an eigen-system ellipsoid with principle axes λ_1 and λ_2 (in 2D). Our idea for quantifying the local saliency is to use a measure corresponding to the area of the bounding box A_i of the eigen-system ellipse. In 2D it will be sufficient to compute the determinant of the tensor sum at site *i*:

$$D_i = \det T_i = (\lambda_1 \lambda_2)^2 = A_i^2 \tag{3.10}$$

Using D_i as a local saliency measure offers many advantages. First, we obtain a straightforward analytic expression in terms of the point coordinates. This expression can be computed quickly for every point in the dataset, with linear computational complexity on a grid: $O(N)$, and in $O(N \log N)$ for the case of general point-sets. Furthermore, for feature-points with degenerate stick tensors the measure is zero. Given that using our one pass voting scheme this latter case will occur only in flat areas (zero curvature), such as lines in 2D and planar areas in 3D, the local saliency descriptor can thus be related to the basic differential feature which is curvature. It will have the added advantage of being robustly computed from noisy and sparse datasets. The expression of the descriptor in 2D is then given by:

$$D_i = \left(\sum_{X_j \in P} \frac{\exp(-\frac{\|X_i - X_j\|^2}{\sigma^2})(x_i - x_j)^2}{(x_i - x_j)^2 + (y_i - y_j)^2} \right) \left(\sum_{X_j \in P} \frac{\exp(-\frac{\|X_i - X_j\|^2}{\sigma^2})(y_i - y_j)^2}{(x_i - x_j)^2 + (y_i - y_j)^2} \right) - \left(\sum_{X_j \in P} \frac{\exp(-\frac{\|X_i - X_j\|^2}{\sigma^2})(x_i - x_j)(y_i - y_j)}{(x_i - x_j)^2 + (y_i - y_j)^2} \right)^2 \quad (3.11)$$

For the 3D case, we derive the expression of the descriptor using a similar reasoning. Voting will be done using plate tensors encoding the uncertainty of normals (Fig. 3.8). Site $X_i = (x_i, y_i, z_i)^T \in P$ will vote at neighboring site $X_j = (x_j, y_j, z_j)^T \in P$ in the direction of the unit vector $t^{ij} = \frac{X_i - X_j}{\|X_i - X_j\|} = (t_x^{ij}, t_y^{ij}, t_z^{ij})^T$ with the plate tensor expressed as:

$$T_{ij} = \begin{bmatrix} 1-(t_x^{ij})^2 & -t_x^{ij}t_y^{ij} & -t_x^{ij}t_z^{ij} \\ -t_x^{ij}t_y^{ij} & 1-(t_y^{ij})^2 & -t_y^{ij}t_z^{ij} \\ -t_x^{ij}t_z^{ij} & -t_y^{ij}t_z^{ij} & 1-(t_z^{ij})^2 \end{bmatrix} \quad (3.12)$$

$$= \frac{1}{\|X_i - X_j\|^2} \begin{bmatrix} (y_i - y_j)^2 + (z_i - z_j)^2 & -(x_i - x_j)(y_i - y_j) & -(x_i - x_j)(z_i - z_j) \\ -(x_i - x_j)(y_i - y_j) & (x_i - x_j)^2 + (z_i - z_j)^2 & -(y_i - y_j)(z_i - z_j) \\ -(x_i - x_j)(z_i - z_j) & -(y_i - y_j)(z_i - z_j) & (x_i - x_j)^2 + (y_i - y_j)^2 \end{bmatrix}$$

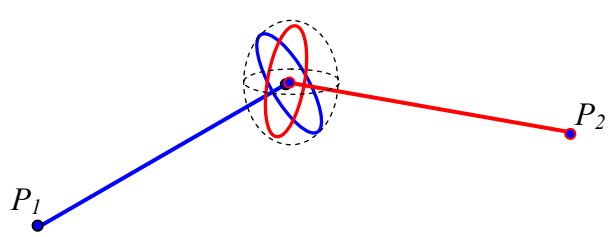


Fig. 3.8. Voting for feature saliency in 3D using plate tensors, the accumulated votes are described by the ellipsoid bounding the two plates.

The expression of the accumulated votes at a given site is similar to (3.9), while the local saliency measure is given by the determinant of the tensor T_i (square of the volume V_i of the bounding box of the uncertainty ellipsoid):

$$D_i = \det T_i = (\lambda_1 \lambda_2 \lambda_3)^2 = V_i^{\,2} \tag{3.13}$$

3.4 A 3D Registration Criterion Derived from Joint Voting

The development of the feature saliency descriptor presented in the previous section was primarily geared toward the task of 3D point-sets registration. The emphasis on the analytic expression of the local saliency descriptor was motivated by our goal to address the registration task in the framework of standard gradient-based optimization techniques. Our first alignment criterion is based on the computation of global saliency for the joint datasets to be registered. The method relies on the fact that at the unregistered position the point-sets will normally have little interaction, due to the local nature of saliency inference. On the other hand, at the aligned position structures common to the registered point-sets will overlap resulting in a local increase in the number of votes, and hence of the measure of feature saliency as defined above. This is illustrated in Fig. 3.9. If we are to register the datasets $P_1, P_2, ..., P_N$, and if $P(\Theta) = P_1 \cup P_2 ... \cup P_N$, with Θ being the transformation parameter vector, we can define, for a given pose Θ , the criterion that measures the total saliency of the resulting structure as:

$$C(\Theta) = \sum_{X \in P(\Theta)} S(X) \tag{3.14}$$

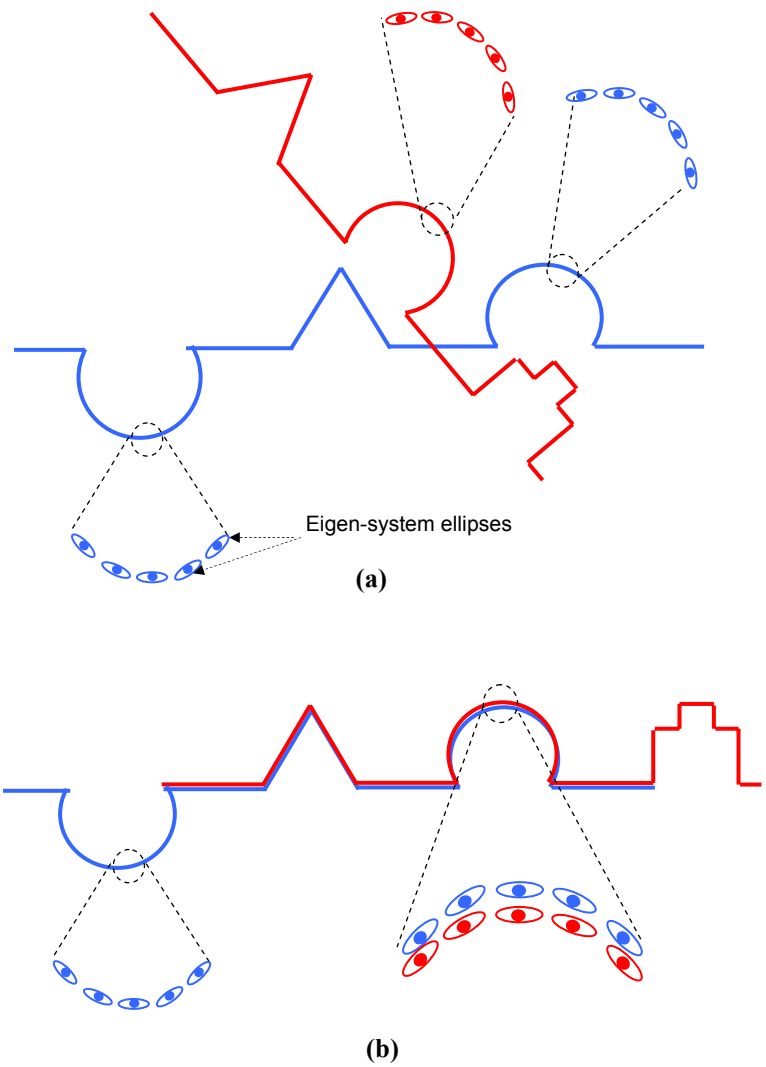


Fig. 3.9. Given the local nature of Tensor Voting the two datasets will have little interaction at the unregistered position (a). When the two patterns overlap (b) local saliency (proportional to uncertainty ellipse size) in the intersection will increase while remaining the same in the other areas.

the summation of all local descriptors. This criterion derives from the same ideas that we employed for camera motion recovery. Its analytic expression is straightforward to compute, allowing for the use of a range of well known optimization techniques [Teukolsky92].

The method was tested on real datasets and was able to register several free-form shapes reconstructed from real-world data. Nonetheless, some fundamental limitations led as to the simplification of the expression (3.14) resulting in a different paradigm altogether. The first problem with the above criterion is presence of poles in the expression of the saliency descriptor (3.11), due to the use of unit vectors in the voting. This means that when points become too close the criterion and in particular its derivatives become instable. The other, more serious, problem is due to the nature of the voting process itself. We noticed that in the case of datasets containing a significant amount of smooth surfaces (and curves) the criterion will tend to have multiple modes none of which will be at exactly the registered position. This is illustrated in Fig. 3.10. The reason for this behavior is due to the fact that for smooth surfaces the saliency tensors are nearly degenerate and the local descriptor's value is small. Therefore, in the case where these surfaces are registered the overall saliency will be small. It will be also small in the case where the two smooth surfaces are far apart, beyond voting range. It is in the intermediate case of displacements by about one σ in the direction normal to the surfaces that we have an increase in the saliency. The increase is due to the effects of one surface on the other in the form of vertical tensor components. Therefore if we employ our method we risk obtaining datasets mis-registered by up to one σ. These drawbacks are due to the idea performing both voting and registration at the same time. We found that we can overcome these problems by decoupling the two tasks.

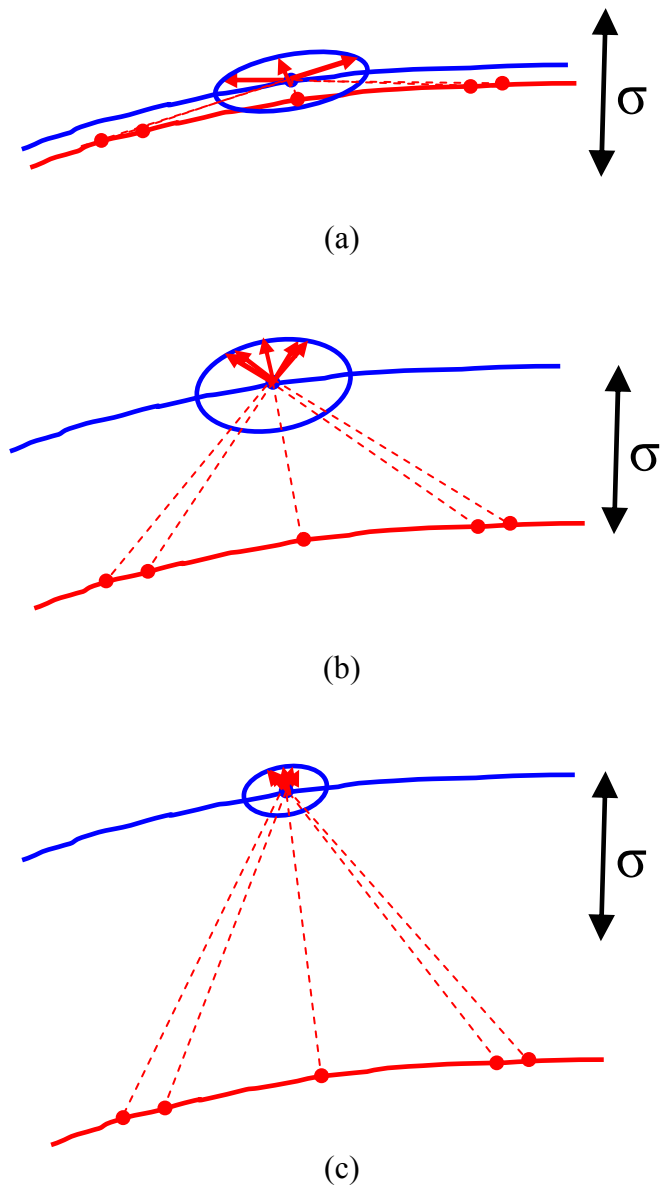


Fig. 3.10. Tensor voting in the case of smooth curves. In the case of nearly registered curves we will have almost degenerate tensors with the eigen-system ellipses having a small area (a). The ellipse will be also small for curves that are separated vertically by a distance larger than σ due to the rapid decay of the votes (c). But in the intermediate case of a vertical displacement around one σ, we can have a larger saliency due to the vertical tensor components (b). Hence we have a maximum saliency at the wrong position.

The local saliency features, will be computed for each dataset as descriptors invariant to rigid transformations, proving to be very robust to noise. A new criterion, which can be seen as a simplification of the previous one will retain the ideas of exponentially decaying vote from one dataset on the other, although in this case we will have a scalar field instead of a tensor one. This will represent the seminal contribution of our work and will be introduced in the next chapter.

4 GAUSSIAN FIELDS FOR FREE-FORM SHAPE REGISTRATION

In this core theory chapter we focus on designing a new energy maximization algorithm for registering free-from shapes represented as point-sets. To overcome several of the shortcomings of standard registration techniques, in particular the ICP algorithm, we aim at accomplishing the following objectives:

1. The criterion to be maximized should be differentiable and preferably convex in the neighborhood of the registered position, allowing for the use of standard optimization techniques.
2. The method should not need any explicit set of point correspondences.
3. The method will incorporate as much available information as possible, in addition to point coordinates, including local shape descriptors or associated intensity values.
4. The method will allow for as large of a region of convergence as possible and reduce dependence on close initialization.
5. The resulting algorithm must be computationally efficient.

We will also show that the formulation of our new criterion addresses the important issues of information content and shape complexity, and discuss the applicability of the method to various real-world tasks.

4.1 A Discrete Combinatorial Criterion

In developing our algorithm we start with a rigorous and consistent definition of the registration task. Registration is in fact a special sub-problem of pattern matching, where the purpose is to locate a 'model' in the 'data'. In the registration task the goal is the recovery of both the correspondence and spatial transformations that ensure the best match. For most applications where registration is employed the assumption is that a significant overlap between the model and the data exists. While it is difficult to have a general definition that will encompass all special cases of this typical ill-posed problem, a good intuitive definition can consist of stating that the registered position is the one resulting in the maximum point-to-point overlap of model and data (considering the noiseless case). Such definition uses a minimum amount of information about the datasets, just the position of points, and could be augmented by requiring local shape similarity between the points. We will show the enhancing effect of using this local information on the quality of the registration.

We start by introducing a very simple combinatorial criterion satisfying the maximum (point-to-point) overlap of two point-sets $M = \{P_i\}_{i=1\ldots N_M}$ and $D = \{Q_j\}_{j=1\ldots N_D}$, that are registered by a transformation Tr^*. We assume at this point the noiseless case. For the problem to be well–posed we need also to assume that M and D have a maximum point-to-point overlap at the aligned position. Then the following measure (4.1) will have a global maximum at Tr^*:

$$E(Tr) = \sum_{\substack{i=1\ldots N_M \\ j=1\ldots N_D}} \delta(d(Tr(P_i), Q_j)) \tag{4.1}$$

with $\delta(t) = 1$ for $t = 0$

and $\delta(t) = 0$ otherwise

Where $d(P,Q)$ is the distance (in our case Euclidean) between points. Incorporating local shape similarity in this criterion is straightforward and requires just using a higher dimensional representation of the datasets where points are defined by both position and a vector of shape attributes: $M = \{(P_i, S(P_i))\}_{i=1...N_M}$ and $D = \{(Q_j, S(Q_j))\}_{j=1...N_D}$.

4.2 Mollification and the Gaussian Criterion

Obviously the resulting discrete criterion is not continuous with respect to the alignment transformations and can be visualized by a collection of "spikes" in parameter space. The resulting optimization problem will not be practical since it is difficult to find the global maximum of discrete combinatorial functions. One of the core ideas of upon which our approach is built is to find a smooth approximation of the combinatorial criterion using an analytical method known as *Mollification.* This approach was used as a tool to regularize ill-posed problem with respect to differentiability [Murio93].

Given the Gaussian kernel $\rho_\sigma(t) = \exp(\frac{-t^2}{\sigma^2})$, and an arbitrary non-differentiable function $f(t)$ defined on $\Omega \subset \Re^d$, a 'mollified' function $f_\sigma(t)$ can be obtained by convolution:

$$f_\sigma(t) = (\rho_\sigma * f)(t) = \int_\Omega \exp(\frac{-(t-s)^2}{\sigma^2}) f(s) ds \tag{4.2}$$

The resulting function will be an approximation of the original one such that: $\lim_{\sigma \to 0} f_\sigma(t) = f(t)$. Furthermore we will have $f_\sigma \in C^\infty(\Omega)$. This operation is also known as the Gauss Transform and is encountered in many applications. Now if we apply discrete mollification to our combinatorial registration criterion (4.1) we have:

$$E_\sigma(Tr) = \int \exp(-\frac{(d(Tr(P_i),Q_j)-s)^2}{\sigma^2})\{\sum_{\substack{i=1\ldots N_M \\ j=1\ldots N_D}} \delta(d(Tr(P_i),Q_j))\}ds$$

$$= \sum_{\substack{i=1\ldots N_M \\ j=1\ldots N_D}} \int \exp(-\frac{(d(Tr(P_i),Q_j)-s)^2}{\sigma^2})\delta(d(Tr(P_i),Q_j))ds$$

$$= \sum_{\substack{i=1\ldots N_M \\ j=1\ldots N_D}} \int \exp(-\frac{(d(Tr(P_i),Q_j)-s)^2}{\sigma^2})\delta(s)ds = \sum_{\substack{i=1\ldots N_M \\ j=1\ldots N_D}} \exp(-\frac{d^2(Tr(P_i),Q_j)}{\sigma^2}) \quad (4.3)$$

The mollified criterion is a straightforward sum of Gaussians of distances between all pairs of model and data points. Expression (4.3) can be re-interpreted physically as the integration of a potential field whose sources are located at points in one of the datasets dataset and targets in the other one. In the noisy case the Gaussian criterion can account for the noise affecting the position of points by relaxing the parameter σ to values near that of noise variance. Fig 4.1 and Fig. 4.2 illustrate the working of the discrete combinatorial criterion and the mollified version.

Having met the first of our objectives, which is differentiability, we now examine the possibility of extending the basin of convergence of our criterion. (We are focusing here on the case of rigid registration where $Tr(Q_j) = RQ_j + t$). Being the sum of closely packed Gaussian functions, the profile of the criterion with respect to the transformation parameters will generally have the appearance of a Gaussian, with local convexity in the neighborhood of the registered position.

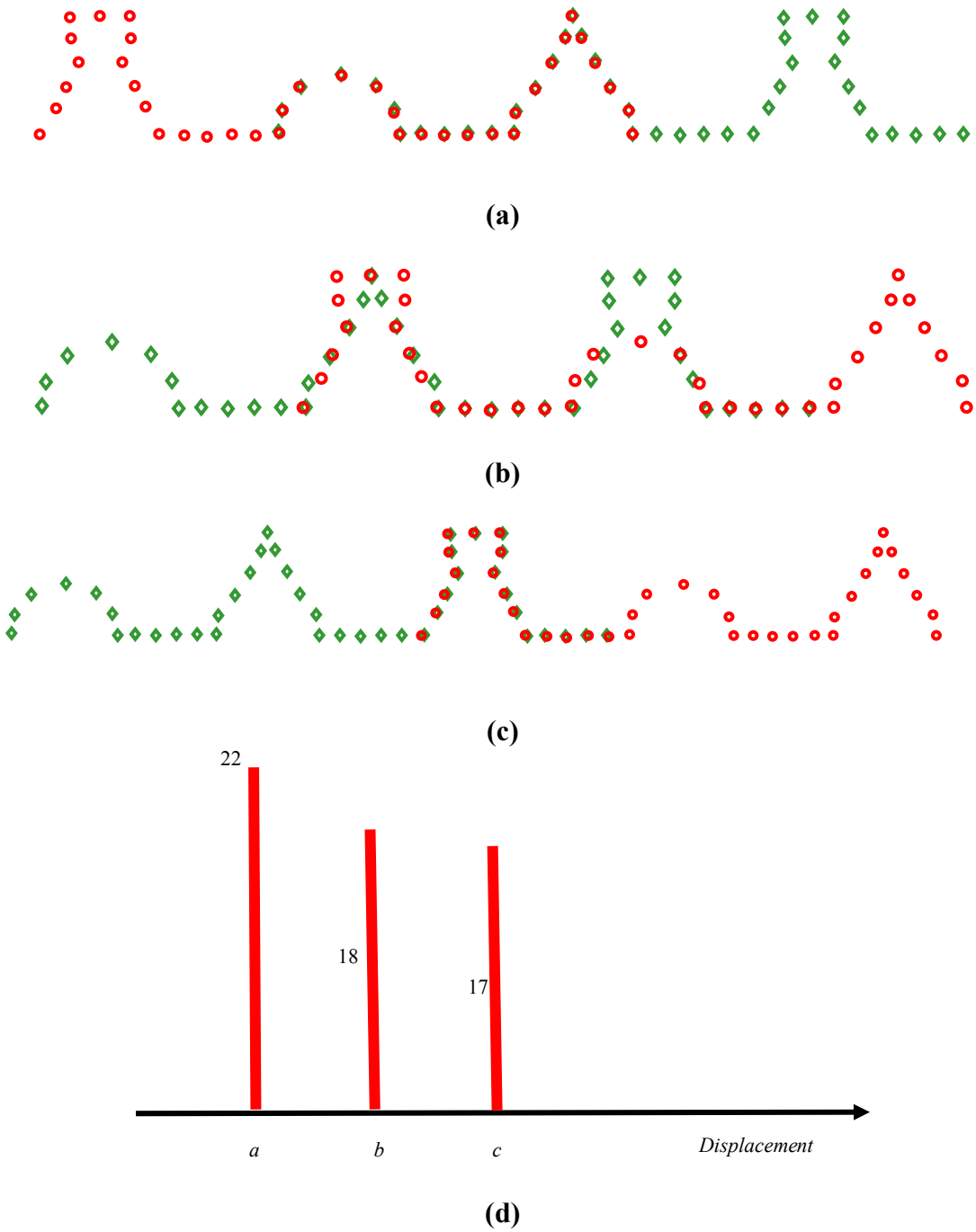
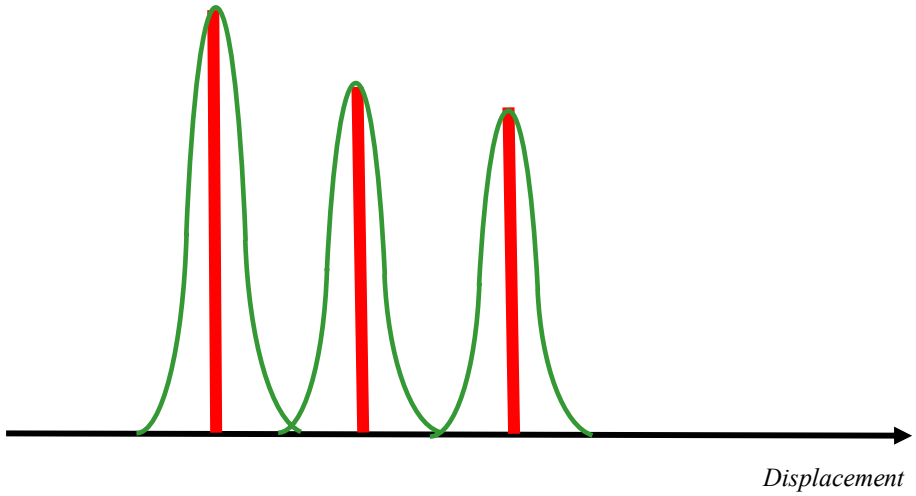
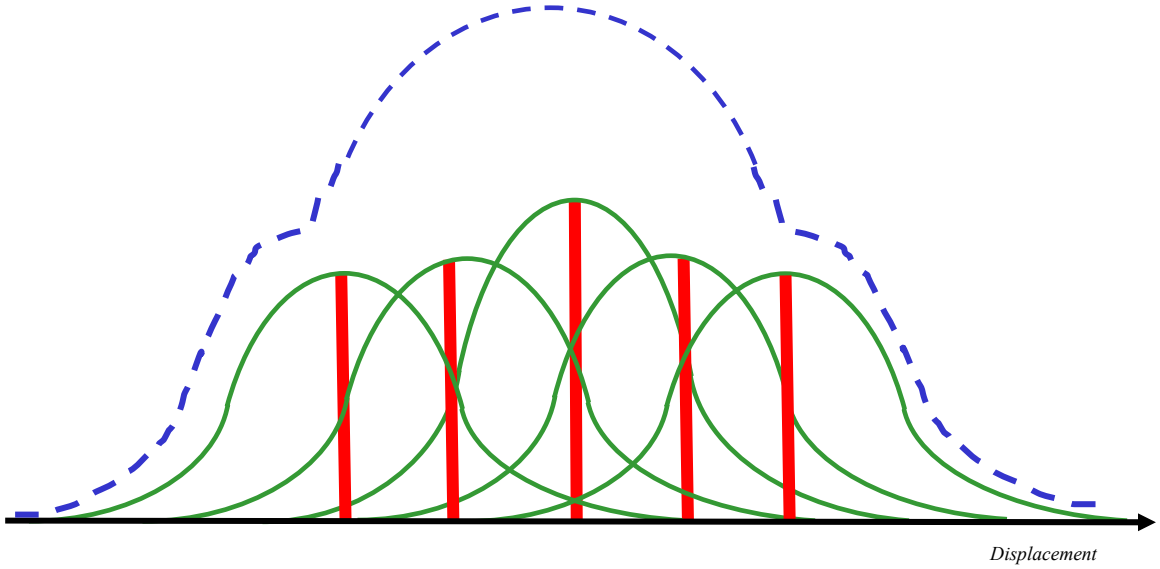


Fig. 4.1. Illustration of the discrete combinatorial criterion. Two point-sets are shown in their registered position (a), and for two relative displacements in the horizontal direction. The criterion (d) will count the number of points that overlap at each position.



(a)



(b)

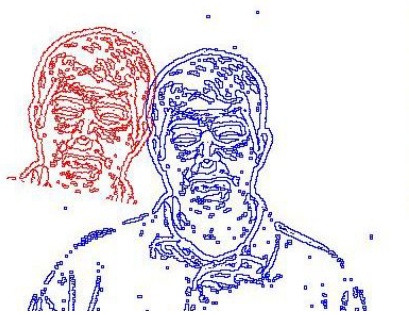
Fig. 4.2. Mollification converts the discrete combinatorial criterion into a smooth sum of Gaussians (a). For σ relaxed we will have an overlap between the different Gaussians. The mixture of these will be our registration criterion, having a dominant peak around the registered position (b).

These are important properties that allow for the use of standard and well-proven gradient-based optimization techniques. Extending the width of the basin of convergence is easily done by increasing the parameter σ. However this relaxation will come at the price of decreasing the localization accuracy of the criterion. The tradeoff between registration accuracy and size of the region of convergence (ROC) is mainly due to the effect of outliers (i.e. the areas that are outside the intersection of model and data). This tradeoff can be illustrated with the behavior of the matching criteria with and without attributes as shown in Fig. 4.3. The profile of the criterion was plotted for a relative displacement of the two point sets of Fig. 4.3(a).

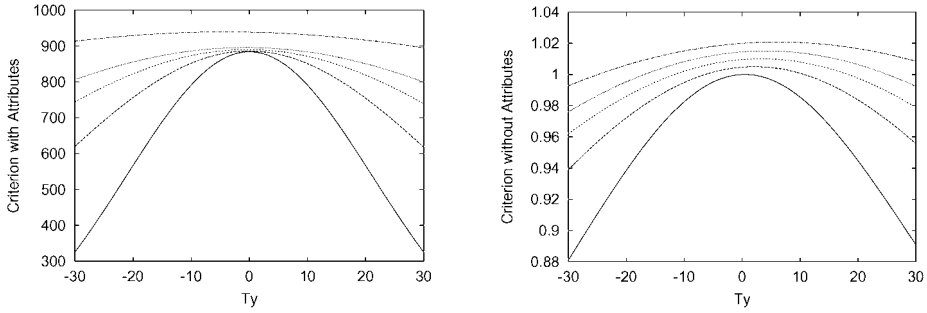
Several plots are shown with increasing σ. For the non-attributed case, where Euclidean distance between point locations is employed, (Fig. 4.3(b)) we notice that as σ increases the width of the Gaussian bell increases too. However the maximum will also drift away from the correct location. When we use the Gaussian criterion augmented with moment invariants, as attributes associated with the points, the maximum is more stable (Fig. 4.3(c)), with nearly no drift for the range of values of σ shown. In the analysis section we will use real datasets to study the localization error as a function of the force range parameter σ. Assuming that at the registered position the model point-set $M = \{P_i\}_{i=1\ldots N_M}$ is completely included in the data point-set $D = \{Q_j\}_{j=1\ldots N_D}$, and that the points of D with corresponding matches in M are labeled from 1 to N_M, the criterion can be broken into two components representing both inliers-inliers and outliers-inliers interaction:

$$E_\sigma(Tr) = \sum_{\substack{i=1\ldots N_M \\ j=1\ldots N_M}} \exp\left(-\frac{d^2(Tr(P_i),Q_j)}{\sigma^2}\right) + \sum_{\substack{i=1\ldots N_M \\ j=N_M+1\ldots N_D}} \exp\left(-\frac{d^2(Tr(P_i),Q_j)}{\sigma^2}\right)$$

$$E_\sigma(Tr) = E_\sigma^{in-in}(Tr) + E_\sigma^{out-in}(Tr) \tag{4.4}$$



(a)



(b) (c)

Fig. 4.3. Profiles of the Gaussian energy function for a displacement around the registered position of the datasets shown in (a). In (b) the profiles are plotted in the case without attributes for σ = 30,50,70,90,150 (from narrowest to widest). Plots with moment invariants as attributes for the same values of σ are shown in (c). The scale of the datasets is about 200×200 pixels. (For (b) magnitudes were rescaled for comparison).

It can be shown that for rigid transformations $E_{\sigma}^{in-in}(R,t)$ will have the same maximum for any σ. For small values of the decay parameter the second term will have no effect on the global maximum of the function at the registered position since $\lim_{\sigma \to 0} E_{\sigma}^{out-in}(R^*,t^*) = 0$ (again assuming the uniqueness of the aligned position). To reduce the effect of the outliers and ensure good localization error for the Gaussian registration criterion, while at the same time increasing the area of convergence, it is suitable to associate as much information as possible to the points. For example in the case of range data registration 3D scanners can provide additional intensity or color information for each sample acquired. Even in the case where only geometry is available we can compute for each point a vector of local shape descriptors. The inclusion of this additional information is achieved by extending the distance measure between points in the criterion as follows:

$$E_{\sigma,\Sigma}(Tr) = \sum_{\substack{i=1\ldots N_M \\ j=1\ldots N_D}} \exp\left(-\frac{\left\|Tr(P_i)-Q_j\right\|^2}{\sigma^2} - (S(Tr(P_i))-S(Q_j))^T \Sigma^{-1} (S(Tr(P_i))-S(Q_j))\right)) \quad (4.5)$$

Where the $\|..\|$ is the Euclidean distance in 3D and the Σ associated with the attribute vector is just a diagonal matrix with positive components generalizing the mollification to higher dimensions, this matrix will also allow for the proper scaling of the different attributes before combination. If we let:

$$\omega_{\Sigma}^{ij}(Tr) = \exp(-(S(Tr(P_i))-S(Q_j))^T \Sigma^{-1} (S(Tr(P_i))-S(Q_j))))$$

Then the Gaussian criterion will be:

$$E_{\sigma,\Sigma}(Tr) = \sum_{\substack{i=1\ldots N_M \\ j=1\ldots N_M}} \omega_{\Sigma}^{ij}(Tr)\exp\left(-\frac{\left\|Tr(P_i)-Q_j\right\|^2}{\sigma^2}\right) + \sum_{\substack{i=1\ldots N_M \\ j=N_M+1\ldots N_D}} \omega_{\Sigma}^{ij}(Tr)\exp\left(-\frac{\left\|Tr(P_i)-Q_j\right\|^2}{\sigma^2}\right)$$

$$= E_{\sigma,\Sigma}^{in-in}(Tr) + E_{\sigma,\Sigma}^{out-in}(Tr) \tag{4.6}$$

In the case where the attributes are invariant to the aligning transformations ω_{Σ}^{ij} will not depend on Tr. If Σ is chosen with as small components as possible (just the noise level), the term ω_{Σ}^{ij} can be expected to reduce the value of $E_{\sigma,\Sigma}^{out-in}(Tr)$ more than that of $E_{\sigma,\Sigma}^{in-in}(Tr)$. This will be helpful when the spatial decay parameter σ is relaxed to extend the region of convergence by allowing good localization of the registered position and reducing the need for close initialization.

Our framework accounts also for two important aspect in free-form registration, namely the quantity and quality of information available in the datasets, and the amount of overlap between the two datasets. Both factors strongly influence the accuracy of the alignment. In the context of Gaussian Fields the effect of the outliers will be generally important if the overlap region is small relative to the size of the two datasets. The information contained in the datasets, such as the complexity of the shapes to be registered, is directly encoded in the point attributes. The increase in information content results in the decrease of the weights ω_{Σ}^{ij} and reduces $E_{\sigma,\Sigma}^{out-in}(Tr)$ further than $E_{\sigma,\Sigma}^{in-in}(Tr)$, since the latter term accounts for the corresponding sets. In fact the quantity $\sum_{i,j}\omega_{\Sigma}^{ij}$ can be a good measure for shape complexity, having a high value for simple shapes and a low value for complex ones.

To optimize the resulting continuously differentiable registration criterion we can use a large variety of gradient-based strategies such as conjugate gradient and quasi-Newton methods. In addition we also need to devise an efficient scheme for increasing the region of convergence. From our earlier observations about the Gaussian Fields approach two main approaches can be adopted either separately or together. The first will rely on computing a large number of local descriptors from the point-sets and

include any additional independent information (such as associated color or intensity) to increase the ROC without increasing localization error too much. This approach can lead to a higher accuracy at the cost of additional computations. A second method based on tuning the parameter σ can be devised. In such strategy we start initially with a large sigma and proceed to find the transformation parameters maximizing the Gaussian criterion then we will decrease the value of the decay parameter and maximize again starting from the previous parameters. This last scheme will be chosen because of its computational efficiency and is described in more details in section 4.6.

4.3 Attributes in the Case of 3D Rigid Registration

In the case of 3D free-from registration, which is our main application of the Gaussian Fields method, we have the choice of several local feature descriptors as point attributes. When surfaces are extracted from the point-sets several descriptors based on differential properties can be employed, starting with curvature. Given that we represent our shapes as point-sets, a natural idea is to use 3D moment invariants. Additionally, we employ the local descriptor presented in Chapter 3 that we called point saliency. The three moment invariants [Sedjadi87] are commonly used for object recognition tasks and were also employed in registration algorithms such as in the extension of ICP by Sharp et al [Sharp02]. The moments J_1, J_2, and J_3 are defined for a local neighborhood N around a point $P(X_P, Y_P, Z_P)$ by:

$$
\begin{aligned}
J_1 &= \mu_{200} + \mu_{020} + \mu_{002} \\
J_2 &= \mu_{200}\mu_{020} + \mu_{200}\mu_{002} + \mu_{020}\mu_{002} - \mu_{110}^2 - \mu_{101}^2 - \mu_{011}^2 \\
J_3 &= \mu_{200}\mu_{020}\mu_{002} + 2\mu_{110}\mu_{101}\mu_{011} - \mu_{002}\mu_{110}^2 - \mu_{020}\mu_{101}^2 - \mu_{200}\mu_{011}^2
\end{aligned}
\tag{4.7}
$$

with

$$\mu_{pqr} = \sum_{(X,Y,Z)\in N} (X - X_P)^p (Y - Y_P)^q (Z - Z_P)^r \tag{4.8}$$

For computational efficiency and for robustness to noise we will employ in our later experimental analysis only two descriptors J_1 and the point saliency measure S.

4.4 Application to Multimodal Image Registration and Tracking

The Gaussian Fields approach can be readily employed for image and other 2D registration tasks. The method can be applied to discrete feature maps, obtained from gradient or frequency maps for example. This method allows for the registration of multimodal as well as single modality images and provides a potentially good approach to the task of edge-based registration and fusion. In this case we have a 2D point-sets registration task under appropriately chosen motion models such as the affine or the more common projective 8-paramters model. Radial-basis functions and other non-rigid warps can also be employed for the alignment when necessary. The main issues will be (1) the choice of shape attributes preserving the differentiability of the criterion, and (2) imposing regularizing constraints on the transformations so that spurious solutions are avoided. Although affine attributes exist, invariance in the case of other warps is not guaranteed. But given that invariance is not necessary and is just sought for simplifying gradient computation, it is sufficient for our optimization framework to use descriptors that are differentiable with respect to the transformation parameters.

For single modality registration and for tracking applications it is also highly desirable to use color or intensity attributes in addition to local shape descriptors. In addition to parametric warps the Gaussian Fields criterion can be also employed to recover a dense point-flow that registers images and shapes, for this purpose regularized variational and Partial Differential Equations (PDE) methods can be readily applied.

While we emphasized here a general point-sets representation that is suitable for integrating edge maps by efficiently using shape information, the method can be directly applied to the original 'continuous' grid representation, with the pixels (and voxels in 3D) considered as points having associated color and intensity attributes.

4.5 Relaxing Absolute Orientation, Closed-Form Approximations

If the point-sets have sufficient complexity σ can be chosen large (compared to the size of the datasets) for a bounded localization error (a fact that will be verified experimentally). Therefore we can attempt the approximation of the criterion (4.5) using a first order development of the exponential:

$$E_{\sigma,\Sigma}(R,t) = \sum_{\substack{i=1\ldots N_M \\ j=1\ldots N_D}} \omega_{\sigma,\Sigma}^{ij} \left(1 - \frac{\left\| RP_i + t - Q_j \right\|^2}{\sigma^2}\right) \tag{4.10}$$

Maximizing (4.10) is equivalent to minimizing:

$$\widetilde{E}_{\sigma,\Sigma}(R,t) = \sum_{\substack{i=1\ldots N_M \\ j=1\ldots N_D}} \omega_{\sigma,\Sigma}^{ij} \left\| RP_i + t - Q_j \right\|^2 \tag{4.11}$$

The resulting problem can be seen as a relaxation of the well known absolute orientation task where a list of point correspondences between the datasets is given [Horn87]. Proceeding in a similar way as for absolute orientation we need to add the orthonormality constraints on the rotation matrix $R = \begin{pmatrix} r_1^T \\ r_2^T \\ r_3^T \end{pmatrix}$ using Lagrange multipliers. We obtain the modified criterion:

$$\widetilde{E}_{\sigma,\Sigma}(R,t) = \sum_{\substack{i=1\ldots N_M \\ j=1\ldots N_D}} \omega_{\sigma,\Sigma}^{ij} \left\| RP_i + t - Q_j \right\|^2 + \sum_{k=1\ldots 3} 2\lambda_k (r_k^T r_k - 1) + 2\lambda_4 r_1^T r_2 + 2\lambda_5 r_1^T r_3 + 2\lambda_6 r_2^T r_3 \tag{4.12}$$

We first express the translation as a function of the rotation by computing the partial derivative with respect to the rotation and setting it to zero which gives:

$$t = \frac{\sum_{\substack{i=1\ldots N_M \\ j=1\ldots N_D}} \omega_{\sigma,\Sigma}^{ij} Q_j - R \sum_{\substack{i=1\ldots N_M \\ j=1\ldots N_D}} \omega_{\sigma,\Sigma}^{ij} P_i}{\sum_{\substack{i=1\ldots N_M \\ j=1\ldots N_D}} \omega_{\sigma,\Sigma}^{ij}} = \overline{Q} - R\overline{P} \tag{4.13}$$

Computing the partials with respect to the rotation parameters will result in the following system of equations:

$$\begin{aligned}
&\left[\sum_{\substack{i=1\ldots N_M \\ j=1\ldots N_D}} \omega_{\sigma,\Sigma}^{ij} (P_i - \overline{P})(P_i - \overline{P})^T\right] r_1 + \lambda_1 r_1 + \lambda_4 r_2 + \lambda_5 r_3 = \sum_{\substack{i=1\ldots N_M \\ j=1\ldots N_D}} \omega_{\sigma,\Sigma}^{ij} (Q_j{}^x - \overline{Q}^x)(P_i - \overline{P}) \\
&\left[\sum_{\substack{i=1\ldots N_M \\ j=1\ldots N_D}} \omega_{\sigma,\Sigma}^{ij} (P_i - \overline{P})(P_i - \overline{P})^T\right] r_2 + \lambda_4 r_1 + \lambda_2 r_2 + \lambda_6 r_3 = \sum_{\substack{i=1\ldots N_M \\ j=1\ldots N_D}} \omega_{\sigma,\Sigma}^{ij} (Q_j{}^y - \overline{Q}^y)(P_i - \overline{P}) \\
&\left[\sum_{\substack{i=1\ldots N_M \\ j=1\ldots N_D}} \omega_{\sigma,\Sigma}^{ij} (P_i - \overline{P})(P_i - \overline{P})^T\right] r_3 + \lambda_5 r_1 + \lambda_6 r_2 + \lambda_3 r_3 = \sum_{\substack{i=1\ldots N_M \\ j=1\ldots N_D}} \omega_{\sigma,\Sigma}^{ij} (Q_j{}^z - \overline{Q}^z)(P_i - \overline{P})
\end{aligned} \tag{4.14}$$

We define the matrices

$$A = \sum_{\substack{i=1\ldots N \\ j=1\ldots N'}} \omega_{ij} (Q_j - \overline{Q})(Q_j - \overline{Q})^T$$

$$\Lambda = \begin{bmatrix} \lambda_1 & \lambda_4 & \lambda_5 \\ \lambda_4 & \lambda_2 & \lambda_6 \\ \lambda_5 & \lambda_6 & \lambda_3 \end{bmatrix} \text{ and } B = (b^x b^y b^z) \text{ with}$$

$$b^x = \sum_{\substack{i=1...N \\ j=1...N'}} \omega_{ij} (P_i^x - \overline{P}^x)(Q_j - \overline{Q})$$

$$b^y = \sum_{\substack{i=1...N \\ j=1...N'}} \omega_{ij} (P_i^y - \overline{P}^y)(Q_j - \overline{Q})$$

$$b^z = \sum_{\substack{i=1...N \\ j=1...N'}} \omega_{ij} (P_i^z - \overline{P}^z)(Q_j - \overline{Q})$$

Then our system of equations (4.10) can be rewritten as:

$$AR^T + R^T \Lambda = B \tag{4.15}$$

Observing that A and Λ are symmetric and using the SVD decomposition $B = UDV$ the rotation matrix is recovered:

$$R = V^T U^T \tag{4.16}$$

Hence, the rigid registration parameters are computed in closed form. This approach will not lead to accurate registration by itself; however it is very useful for finding a reasonable initialization to the more accurate iterative alignment step.

4.6 Optimization Strategy

For rigid transformation our criterion is always continuously differentiable. Also we notice a very dominant mode around the registered position. For sufficiently dense point-sets this mode will itself have a shape close to a Gaussian given that it is a mixture of Gaussians closely located in parameter space. We can safely assume a smooth convex behavior around the registered position. This can be demonstrated in the limit by considering, for the sake of simplicity, the two dimensional case. For a small value of σ and small rigid displacements near the registered position (i.e. a ball

of radius ε around the rotation angle and translation vector (φ, t)) the Gaussian criterion (4.3) can be approximated as follows

$$\sum_{\substack{i=1\ldots N_M \\ j=1\ldots N_D}} \exp(-\frac{d^2(Tr(P_i),Q_j)}{\sigma^2}) = \sum_{\substack{i=1\ldots N_M \\ j=1\ldots N_D}} \exp(-\frac{(\cos\varphi P_i^x - \sin\varphi P_i^y + t_x - Q_j^x)^2 + (\sin\varphi P_i^x + \cos\varphi P_i^y + t_y - Q_j^y)^2}{\sigma^2})$$

$$\approx \sum_{\substack{i=1\ldots N_M \\ j=1\ldots N_D}} 1 - \frac{(P_i^x - \varphi P_i^y + t_x - Q_j^x)^2 + (\varphi P_i^x + P_i^y + t_y - Q_j^y)^2}{\sigma^2} \tag{4.17}$$

Using the two approximations for a small rotation $\cos\varphi \approx 1$ and $\sin\varphi \approx \varphi$, in addition to the first order approximation resulting from the small displacement compared with σ:

$$\exp(-\frac{d^2(Tr(P_i),Q_j)}{\sigma^2}) \approx 1 - \frac{d^2(Tr(P_i),Q_j)}{\sigma^2}$$

Clearly in this limit case we obtain an expression (4.17) which is quadratic in the rigid parameters demonstrating the convexity of the criterion. In practice even in the relaxed case we can assume safely convexity in the neighborhood of the registered position.

To optimize the Gaussian Fields criterion we employed one of the most standard gradient-based schemes, namely the quasi-Newton algorithm [Teukolsky92]. The gradient of the criterion (4.5) with respect to a transformation parameter α is given by:

$$\frac{\partial E_{\sigma,\Sigma}(Tr)}{\partial\alpha} = \sum_{\substack{i=1\ldots N_M \\ j=1\ldots N_D}} \frac{-2\omega_\Sigma^{ij}}{\sigma^2} \frac{\partial Tr(P_i)}{\partial\alpha}.(Tr(P_i) - Q_j)\exp(-\frac{\left\|Tr(P_i) - Q_j\right\|^2}{\sigma^2}) \tag{4.18}$$

The quasi-Newton scheme uses the analytic expression for the gradient along with an approximation of the Hessian to update descent directions which are extensions of the

basic Newton step. In each descent direction a line search routine is used to find the optimum. Also objective constraints are imposed to avoid excessively small and large steps. As a stopping criterion the relative change in the variables between successive steps, as well as a check of the magnitude of the gradient, are used.

One of the most important questions that emerged in our earlier discussion of the Gaussian criterion is the issue of the drift associated with the relaxation of the force range parameter σ. The compromise was between an accurate localization with a small value of σ and larger region of convergence for a larger σ at the expense of registration accuracy. To strike a balance between these two constraints we devised a simple scheme based on adapting the values of σ during the optimization process. The scheme consists of two or more runs of the quasi-Newton routine with values of sigma decreasing according to suitable schedule (Fig. 4.4). Using such approach which recalls annealing algorithms [Kirkpatrick83][Teukolsky92] we can start far from the registration parameter while having a good chance of ending with an accurate registration. The main issue here will be a choice of a good reduction schedule for σ. The constraints to be considered are: (1) the lower bound on σ is the noise level, and (2) avoiding being trapped at local minima. We can try to avoid a local maximum by studying the rate at which the global maximum is drifting with change of the force range parameter. We need to ensure that this drift is not resulting in the next run starting from outside the dominant (usually convex) mode (Fig. 4.5). By studying for several datasets the rate at which the maximum of the criterion drifts with respect to σ and the width of the dominant mode at half the maximum for different values of the force range parameter we can determine the value $\eta \in (0,1)$ which multiplies σ at each run of the global scheme such that we avoid the local maxima. Experimental results show that in practice and with no prior knowledge about initialization we will only need two steps. The first of which can exploit the closed form approximation shown previously.

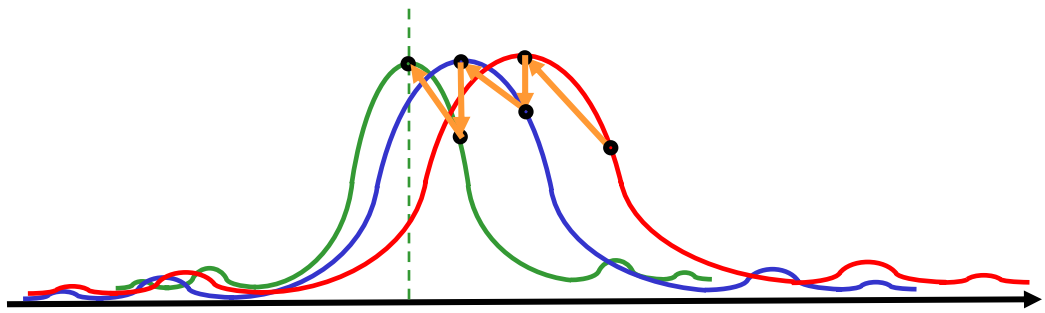


Fig. 4.4. The global optimization strategy that strikes a balance between the width of the region of convergence and the accuracy of registration is based on adapting the parameter σ, starting form a large value and reducing σ until convergence. We will show experimentally that we just need two steps.

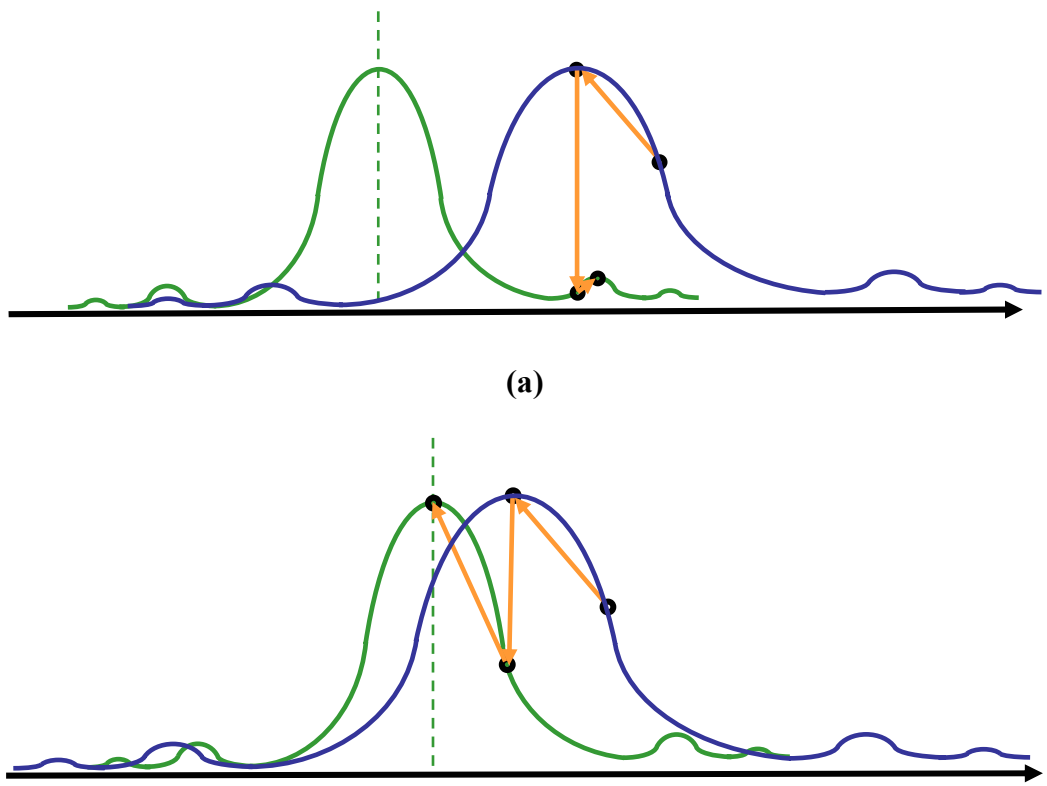


Fig. 4.5. In the global strategy we should be careful to avoid being trapped in a local maximum as seen in (a). This can happen if we decrease the value σ too much. We must guarantee that we start the next iteration of the optimization routine within the dominant mode (b). Experimental analysis will study the best schedule.

4.7 Bias and Variance Dilemma and Evaluation of the Algorithm

So far we stressed that one of the main advantages of the Gaussian Fields framework is the possibility of controlling the region of convergence using one single parameter: σ. In addition to allowing for adaptive optimization schemes this parameter offers the possibility of physical understanding of the uncertainty associated with the registration algorithm. An important topic in many fields of science is the quantitative evaluation of uncertainty. Some of the most striking examples are found in quantum physics where the complimentarity principle of Niels Bohr showed that particles can have different natures according to the type of experiments used to observe them. From this principle Heisenberg derived his well known uncertainty inequalities stating that the precise measurement of a particle's velocity will result in the imprecise measurement of its position. In the same vein, but more relevant to our work, is the so-called Bias and Variance dilemma, which states that for many systems there is a tradeoff between the precision and variance of the state variables characterizing the system. Usually high precision is compounded by a large variance of the variables, and stable systems showing a low variance in the variables will be mostly biased (i.e. less precise). In learning theory, in particular, it was found that most estimators will be subject to the Bias-Variance dilemma [Hastie01][Geman92]. In section 6.3.4 we will use an objective measurement of uncertainty based on the Mean Squared Error (MSE), computed with respect to a distribution of initial transformations, to compare the performance of the Gaussian Fields algorithm with that of ICP. This criterion conveniently combines both Bias and Variance and represents a natural way of evaluating our method.

5 FAST GAUSS TRANSFORM METHODS

The Gaussian Fields registration algorithm, applied to the alignment of two point-sets M and D containing respectively N_M and N_D points, requires several evaluations of the criterion within the optimization scheme. The computational cost of direct evaluation of the mixture of Gaussians is $O(N_M N_D)$, growing quickly with the size of the datasets processed. The same limitation is encountered in other computer vision applications where similar Gaussian mixtures need to be computed, especially for the task of Gaussian kernel density estimation. A solution to this problem was first proposed in the context of potential Fields estimation for particle physics, where an algorithm with asymptotic linear complexity known as the Fast Gauss Transform (FGT) was devised by Greengard and Strain [Greengard91] for computing Gaussian potentials. More recently FGT was applied to the task of kernel density estimation for color modeling and tracking by Elgammal et al [Elgammal03].

5.1 Fast Multipole Methods

Several methods were employed to reduce the computational cost of kernel density estimation including the use of k-nearest neighbor search with special data structures and branch and bound methods [Postaire82][Devroye85][Fukunaga89][Jeon94]. For data with grid structure the Fast Fourier Transform was also employed for evaluating density estimates [Silverman82]. The FGT algorithm is derived from a more general and very efficient class of numerical techniques known as the Fast Multi-pole Methods (FMM) [Greengard87] which was primarily employed for the fast

summation of potential fields generated by a large number of sources such as those encountered in electrostatic and gravitational potential problems. The FMM method was also extended to other applications including the solving the Helmholtz and Maxwell equations and the interpolation scattered data [Gimerov03]. The task of FMM methods is stated as the evaluation of sums of the form:

$$v(y_j) = \sum_{i=1}^{N} u_i \varphi_i(y_j), \; j = 1,...,M \tag{5.1}$$

For which using direct evaluation will cost $O(MN)$. In the FMM approach the functions φ_i are expanded in multiple (singular) series and local (regular) series that are centered at the locations x_0 and y_0 as follows:

$$\varphi(y) = \sum_{n=0}^{p-1} b_n(x_0) S_n(y - x_0) + \varepsilon(p) \tag{5.2}$$

$$\varphi(y) = \sum_{n=0}^{p-1} a_n(x_0) R_n(y - y_0) + \varepsilon(p) \tag{5.3}$$

where S_n and R_n are respectively the multi-pole (singular) and local (regular) basis functions, a_n and b_n the expansion coefficients and ε the error introduced by truncating the series after p terms. The key to reducing the number of operations involved in estimating the sum (5.1) is to express it using the series (5.2) and (5.3). For instance substituting (5.3) in (5.1) we obtain:

$$v(y_j) = \sum_{i=1}^{N} u_i \varphi_i(y_j) = \sum_{i=1}^{N} u_i \sum_{n=0}^{p-1} c_{ni} R_n(y_j - y_0) \tag{5.4}$$

By rearranging the order of the summations we get:

$$v(y_j) = \sum_{n=0}^{p-1} [\sum_{i=1}^{N} u_i c_{ni}] R_n(y_j - y_0) = \sum_{n=0}^{p-1} C_n R_n(y_j - y_0) \tag{5.5}$$

The resulting p-term series (5.5) can be evaluated at the M evaluation points with a cost of $O(Mp)$ operations. Constructing the coefficients C_n requires $O(Np)$ operations, hence the total computational expense is $O((M+N)p)$. In the original FMM framework the potential functions φ_i are not valid over the entire domain, therefore so-called translation operators are used to convert singular expansions around clusters of points into regular expansions that are evaluated at the evaluation points.

5.2 The Fast Gauss Transform Method

The Fast Gauss Transform is a direct application of the FMM approach to the problem of evaluating sums of the form:

$$G(t_i) = \sum_{j=1}^{N} f_j \exp(-(\frac{s_j - t_i}{\sigma})^2), \; i = 1,...,M \tag{5.6}$$

which are slightly more general than those used in our registration algorithm. In the expression (5.6) $\{s_j\}_{j=1,...,N}$ are the centers of the Gaussians known as 'sources' and $\{t_i\}_{i=1,...,M}$ the 'targets'.

The basis of the FGT is the expansion of (5.6) in terms of Hermite and Taylor series. Hermite expansion centered at s_0 results in (5.7):

$$\exp(\frac{-(t-s)^2}{\sigma^2}) = \exp(\frac{-(t-s_0-(s-s_0))^2}{\sigma^2}) \tag{5.7}$$

$$= \exp(\frac{-(t-s_0)^2}{\sigma^2})\sum_{n=0}^{\infty}\frac{1}{n!}(\frac{s-s_0}{\sigma})^n H_n(\frac{t-s_0}{\sigma}) = \sum_{n=0}^{p-1}\frac{1}{n!}(\frac{s-s_0}{\sigma})^n h_n(\frac{t-s_0}{\sigma}) + \varepsilon(p)$$

Where H_n are the *Hermite polynomials* defined by the Rodrigues formula: $H_n(t) = (-1)^n e^{t^2} D^n e^{-t^2}$, $t \in \Re$ with $D = d/dt$, and where the *Hermite functions* $h_n(t)$ are defined by $h_n(t) = e^{-t^2} H_n(t)$.

Similarly using a Taylor expansion centered at t_0 we obtain:

$$\exp(-(\frac{t-s}{\sigma})^2) = \exp(\frac{-(t-t_0-(s-t_0))^2}{\sigma^2}) \tag{5.8}$$

$$\approx \sum_{n=0}^{p-1}\frac{1}{n!}h_n(\frac{s-t_0}{\sigma})(\frac{t-t_0}{\sigma})^n + \varepsilon(p)$$

The first expansion (5.7) is used as counterpart to the multi-pole (far-field) expansion in FMM, while the second is used as the local (near-field) expansion, and the same approach for reducing the computational cost is used in the FGT. In the one dimensional case we can compute the Hermite functions using the following recurrence:

$$h_{n+1}(t) = 2th_n(t) - 2nh_{n-1}(t)\,,\ t \in \Re \tag{5.9}$$

5.2.1 Basic results

The extension to the multivariate case is straightforward. Using the same notations as the original FGT papers the Gaussian in $\Re^d$ is simply the product of univariate Gaussians:

$$\exp(|t-s)|^2) = \exp(-(t_1 - s_1)^2 - \ldots - (t_d - s_d)^2) \tag{5.10}$$

For convenience multi-index notations were adopted. A multi-index $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_d)$ is a d-tuple of nonnegative integers, playing the role of multi-dimensional index. For any multi-index α and any $t \in \Re^d$ we have the following basic definitions:

$$|\alpha| = \alpha_1 + \alpha_2 + \ldots + \alpha_d$$

$$\alpha! = \alpha_1!\alpha_2!\ldots\alpha_d! \tag{5.11}$$

$$t^\alpha = t_1^{\alpha_1} t_2^{\alpha_2} \ldots t_d^{\alpha_d}$$

And the Hermite functions are defined by:

$$h_\alpha(t) = h_{\alpha_1}(t_1)\ldots h_{\alpha_d}(t_d) \tag{5.12}$$

The Hermite expansion of a Gaussian in $\Re^d$ is then simply:

$$\exp(-|t-s|^2) = \sum_{\alpha \geq 0} \frac{(t-s_0)^\alpha}{\alpha!} h_\alpha(s - s_0) \tag{5.13}$$

Greengard and Strain [Greengard91] use three basic lemmas in the FGT algorithm, the first describes how to transform the field due to all sources in a box into a single rapidly converging Hermite expansion about the center of the box:

Lemma 1

Let N_B sources s_j lie in box B with center s_B and side length $r\sqrt{2}\sigma$ with $r<1$. Then the Gaussian Field due to the source in B, $G(t)=\sum_{j=1}^{N} f_j \exp(-(\frac{|s_j - t_i|}{\sigma})^2)$ is equal to a single Hermite expansion about s_B:

$G(t)=\sum_{\alpha\geq 0} A_\alpha h_\alpha(\frac{t-s_B}{\sigma})$, the coefficients A_α are given by:

$$A_\alpha = \frac{1}{\alpha!}\sum_{j=1}^{N_B} f_j(\frac{s_j - s_B}{\sigma})^\alpha$$

The second Lemma shows how to convert an Hermite expansion about s_B into a Taylor expansion about t_C. The Taylor series converges rapidly in a box of side $r\sqrt{2}\sigma$ about t_C, with $r<1$.

Lemma2

The Hermite expansion $G(t)=\sum_{\alpha\geq 0} A_\alpha h_\alpha(\frac{t-s_B}{\sigma})$ has the following Taylor expansion about an arbitrary point t_C:

$G(t)=\sum_{\beta\geq 0} B_\beta h_\alpha(\frac{t-t_C}{\sigma})$, the coefficients B_β are given by

$$B_\beta = \frac{(-1)^{|\beta|}}{\beta!} \sum_{\alpha \geq 0} A_\alpha h_{\alpha+\beta}(\frac{t_C - s_B}{\sigma})$$

Greengard and Strain [Greengard91] use also a third lemma which is a variant of Lemma 2 in which the Hermite series is truncated before converting it to a Taylor series. Truncation errors for all three cases were given and a corrected error bound was presented recently in [Baxter02].

5.2.2 Algorithm and workloads

In summary the essential feature of the FGT method (illustrated in Fig. 5.1) is the clustering of multiple sources using Hermite series and of multiple targets using Taylor series. In the original algorithm the space is subdivided in to boxes of side $r\sqrt{2}\sigma$ with $r \leq 1/2$. Each source is then assigned to the box B in which it lies and each target to the box C where it lies. The original work details the effect on the precision of including only the n nearest boxes. After subdividing the parameter space there are four basic ways in which the FGT algorithm accounts for the influence of N_B sources in a box *B* on M_C targets in a box *C*.

The different combinations are primarily based on the number of sources and targets in the boxes, the overall objective being the reduction of computations while keeping the accuracy of the evaluation as high as possible. The first case is the basic situation when the number of sources and targets is sufficiently low to allow for direct evaluation. The second one arises when the number of sources is low but the number of targets is high, in this case the Taylor expansion 5.8 is used to cluster the targets and accumulate the N_B in the truncated series. When the number of sources in a box is high Hermite series allows for the clustering of sources which, the resulting series are evaluated at each target.

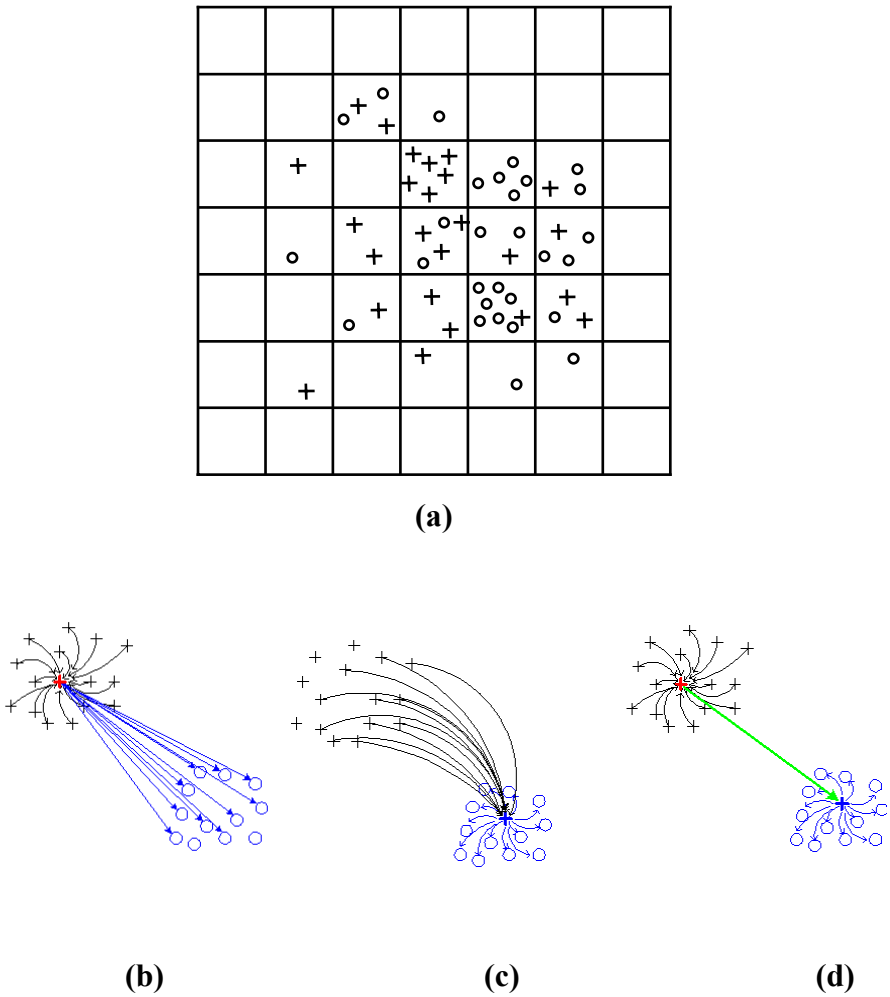


Fig. 5.1. Fast Gauss Transform and the different clustering schemes, sources are represented with crosses (+) and targets with circles (○). The space is subdivided into a regular grid (a). Hermite expansions are used to cluster several sources, which are evaluated at targets (b).Clustering of targets using Taylor expansion (c). Sources are clustered using Hermite expansion then transformed into a Taylor series near a target cluster. Figure redrawn from [Elgammal03].

Finally if both N_B and M_C are large, the Hermite series expansion clusters the sources is shifted into the Taylor series grouping the targets. The original paper [Greengard03] gives the workload for each case:

1- The cost of evaluating N Gaussians at M points is of the order $O(MN)$.
2- N_B Gaussians accumulated in Taylor series: the total cost of both computing the coefficients of the Taylor series and of evaluating the series at the sources is $O((2n+1)^d p^d N) + O(p^d M)$. Considering that influence is limited to the $(2n+1)^d$ boxes within range.
3- Hermite series directly evaluated: $O(p^d N) + O((2n+1)^d p^d M)$.
4- Hermite series accumulated in Taylor series: $O((2n+1)^d dp^{d+1} N_{box}) + O(p^d N) + O(p^d M)$ where N_{Box} accounts for box-box interaction and is bounded by $\min((r\sqrt{2}\sigma)^{-\frac{d}{2}}, M)$.

The algorithm defines a threshold and switches between the different cases in order to reduce the overall CPU load at end asymptotic linear behavior is achieved.

5.2.3 Limitations of the FGT method

While the original FGT was successful in many applications Yang et al [Yang03] pointed to two major drawbacks. The first is due to the exponential growth of complexity with dimensionality seen in the factor $O(p^d)$ showing up in the above mentioned workloads. This effectively limits the FGT to problems involving at most three dimensions. The second defect is due to the use of box data structure. The original algorithm subdivides the space into boxes using a uniform mesh. In higher dimensions such a simple space subdivision scheme is not appropriate since in most real applications, including our registration task, the data is clustered in lower dimensional manifolds. The result will be the existence of too many largely empty

boxes. Hence the algorithm will be burdened by storage requirements and by the processing of empty boxes. Additionally in the case of the uniform subdivision scheme the ratio of the volume of the hypercube to that of the inscribed sphere grows exponentially with dimensionality. In such case the points will have a high probability of falling inside the cube and outside the sphere, which results in the larger truncation error in the Hermite and Taylor expansions.

5.3 Improved Fast Gauss Transform

Yang et al. [Yang03] argue that these limitations are due to the blind application of the FMM approach to the FGT. While the original FMM was developed for singular potential functions with long range forces, the Gaussian functions are C^{∞}. This essentially means that we do not need to perform multi-pole expansions accounting for far-field contributions. They propose an alternative algorithm called the Improved Fast Gauss Transform (IFGT). The algorithm is based on a simple new factorization and on an intelligent space subdivision scheme.

5.3.1 The multivariate Taylor expansion

If we have N sources $\{s_i\}$ centered at s_0 and M target points $\{t_j\}$, the exponential term can be expressed as:

$$\exp(-\frac{\|t_j - s_i\|^2}{\sigma^2}) = \exp(-\frac{\|\Delta t_j\|^2}{\sigma^2})\exp(-\frac{\|\Delta s_i\|^2}{\sigma^2})\exp(\frac{2\Delta t_j \Delta s_i}{\sigma^2}) \tag{5.14}$$

where $\Delta t_j = t_j - s_0$ and $\Delta s_i = s_i - s_0$.

The first two exponential terms in (5.14) can be evaluated separately at either the source or the target points. The only problem is then to evaluate the last term

containing non-separable source and target coordinates. The approach adopted in the IFGT method is to expand this term into the series:

$$\exp(\frac{2\Delta t_j \Delta s_i}{\sigma^2}) = \sum_{n=0}^{\infty} \Phi_n(\Delta t_j)\Psi_n(\Delta s_i) \tag{15}$$

using the expansion functions Φ_n and Ψ_n. By denoting $\phi(\Delta t_j) = \exp(-\frac{\|\Delta t_j\|^2}{\sigma^2})$ and $\psi(\Delta s_i) = \exp(-\frac{\|\Delta s_i\|^2}{\sigma^2})$ we can rewrite the Gaussian mixture of (5.6) as:

$$G(t_j) = \sum_{i=1}^{N} f_i \phi(\Delta(t_j))\psi(\Delta(s_i)) \sum_{n=0}^{\infty} \Phi_n(\Delta t_j)\Psi_n(\Delta s_i) \tag{5.16}$$

If the infinite series (5.15) absolutely converges we can truncate after p terms. By rearranging (5.16) we get

$$G(t_j) = \phi(\Delta(t_j)) \sum_{n=0}^{p-1} C_n \Psi_n(\Delta s_i) + \varepsilon(p) \tag{5.17}$$

with $C_n = \sum_{i=1}^{N} f_i \psi(\Delta(s_i))\Psi_n(\Delta s_i)$

This factorization is the basis of the IFGT, which attempts to reduce the factor $O(p^d)$ that hinders the speed of the original FGT. This factor arises from the way the multivariate Gaussian is considered as the product of univariate Gaussian functions and expanded along each dimension. The idea proposed by Yang et al [Yang03] is to consider the dot product in (5.14) as a scalar variable that can be expanded using Taylor expansion. In this case the expansion functions Φ_n and Ψ_n are expressed as multivariate polynomials.

We will use the fact that the power of the dot product of two vectors x and y is expanded as:

$$(x.y)^n = \sum_{|\alpha|=n} \binom{n}{\alpha} x^\alpha y^\alpha \tag{5.18}$$

Where $\binom{n}{\alpha} = \frac{n!}{\alpha_1!\alpha_2!...\alpha_d!}$

From this we can derive the multivariate Taylor expansion of the Gaussian functions:

$$\exp(2x.y) = \sum_{\alpha \geq 0} \frac{2^{|\alpha|}}{\alpha!} x^\alpha y^\alpha \tag{5.19}$$

using (5.19) in addition to (5.15) and (5.17) we obtain the new factorization and expansion around s_0

$$G(t_j) = \sum_{\alpha \geq 0} C_\alpha \exp(-\frac{\|t_j - s_0\|^2}{\sigma^2})(\frac{t_j - s_0}{\sigma})^\alpha \tag{5.20}$$

with

$$C_\alpha = \frac{2^{|\alpha|}}{\alpha!} \sum_{i=1}^{N} f_i \exp(-\frac{\|s_i - s_0\|^2}{\sigma^2})(\frac{s_i - s_0}{\sigma})^\alpha \tag{5.21}$$

Once the series are truncated after a degree of p-1 the number of terms will be: $r_{p-1,d} = \binom{p+d-1}{d}$, which for higher dimensions is much less than p^d. In fact for $d \to \infty$ the number of terms tends towards $O(d^p)$. In addition to the new

factorization [Yang03] suggests the use Graded lexicographic ordering for the multivariate polynomial coefficients for faster evaluation of (5.20).

5.3.2 Space subdivision

The second important feature in the IFGT framework is efficient space subdivision into cells. The influence of the points in each cell will be collected using the Taylor series expansion. To achieve such subdivision (and the box scheme of the original FGT), the IFGT method transforms the task into a k-center problem, which given n points and a predefined number of clusters k finds a partition of the points into clusters $S_1,...,S_k$ and recovers the cluster centers $c_1,...,c_k$ so as to minimize the maximum radius of the clusters: $\max_i \max_{v \in S_i} \|v - c_i\|$. This problem was addressed using the very simple algorithm proposed by Gonzalez [Gonzalez85]. This greedy algorithm called farthest point clustering works as follows:

1. Pick an initial arbitrary point v_0 as the center of the first cluster and add it to the center set C.

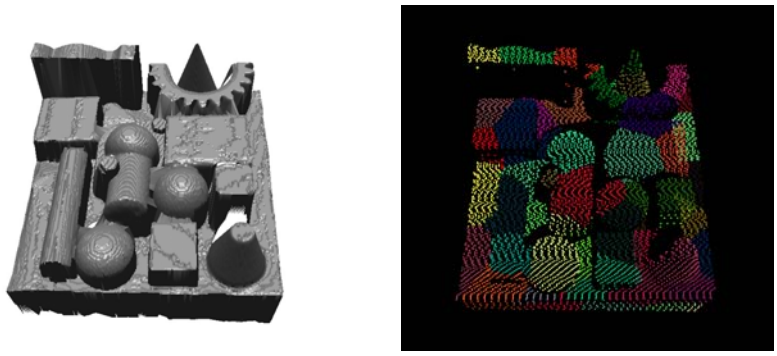
2. Then for $i = 1,...,k$ do:

2.1. For every point compute its distance to the set C: $d_i(v.C) = \min_{c \in C} \|v - c_i\|$. Let v_i the point that is the farthest away from C (i.e. $d_i(v_i.C) = \max_v d_i(v.C)$).

2.2. Add v_i to the set C.

3. Report the points $v_0, v_1,..., v_{k-1}$ as the cluster centers and assign each point to its nearest center.

This algorithm was shown to result in a partition with a maximum radius at most the optimum. This holds for any metric space. The computational complexity for a direct implementation of the algorithm is $O(nk)$, but using Feder and Greene [Feder85] proposed a two phase optimal algorithm with an optimal complexity of $O(n \log k)$. In any case given that the number of clusters will be small the subdivision operation will not be expensive. We found this algorithm to be particularly effective for addressing our 3D datasets. Fig 5.2 shows an example where we applied the scheme to one of our datasets. We used Euclidean distance over the five dimensional space embedding 3D points augmented by the two local feature descriptor presented in the previous chapter.



(a) **(b)**

Fig. 5.2. Result of applying the farthest point clustering algorithm to the 3D model (a). The obtained point clusters (the points of each cluster are labeled with a different color) shown in (b) correspond to regions that are roughly homogeneous with respect to our generalized distance measure (including position and attributes).

5.3.2 IFGT and its workloads

The structure of the IFGT algorithm itself as given in [Yang03] is as follows:

1. Assign the N sources to K clusters using the farthest-point clustering algorithm such that the radius is less than $\sigma\rho_s$.

2. Choose p sufficiently large such that the Taylor series truncation error estimate given in (22) is sufficiently low.

3. For each cluster S_k with center c_k, compute the coefficients given by the expression (21).

4. For each target t_j find its neighbor clusters whose centers lie within the range $\sigma\rho_t$, then the sum of Gaussians can be evaluated by the expression (5.20):

$$G(t_j) = \sum_{\|t_j - c_k\|} \sum_{|\alpha|<p} C_\alpha^k \exp\left(-\frac{\|t_j - s_0\|^2}{\sigma^2}\right)\left(\frac{t_j - s_0}{\sigma}\right)^\alpha$$

As mentioned previously step 1 requires $O(NK)$ operations for direct implementation. For large K it can be optimized using the Feder and Greene [Feder85] algorithm to $O(N \log K)$. The workload for step 3 is $O(Nr_{pd})$, with $r_{pd} = \binom{p+d}{d}$. The cost of step 4 will be $O(Mnr_{pd})$ where n is the maximum number of neighbor clusters for each target. In most applications the number both the precision and the number of clusters can be chosen small, hence K and r_{pd} will be low which leads to the linear complexity

performance of the IFGT algorithm. The algorithm requires only the storage of K coefficients of size r_{pd} .

An error bound on the truncation after order p of the Taylor series (5.20) was given by Yang et al [Yang03] as:

$$|E(t)| \leq F\left(\frac{2^p}{p!}\rho_s^p \rho_t^p + e^{-\rho_t^2}\right) \tag{5.22}$$

where $F = \sum_j f_j$

In summary the IFGT approach presents an attractive alternative to the standard technique. Given that in our case we sre dealing with a dimensionality ranging from at least 3, in the case we are using only the coordinates in the Gaussian criterion, to a dimensionality of 5 if we use the moment invariant J_1 in addition to our saliency measure S the IFGT is clearly more suitable. The application to our task of the algorithm is straightforward and does not require any other modification, except that we choose to include a the de-correlation step before using the generalized distance in both 3D coordinate and attribute space. The substantial gains in the computational cost of criterion evaluation, for different sizes of the point-sets, are shown in Fig. 5.3. One other noteworthy point is the case of the gradient of the Gaussian criterion, which boils down to the computation of a weighted version of the Gaussian function which is similar to the one in (5.6).

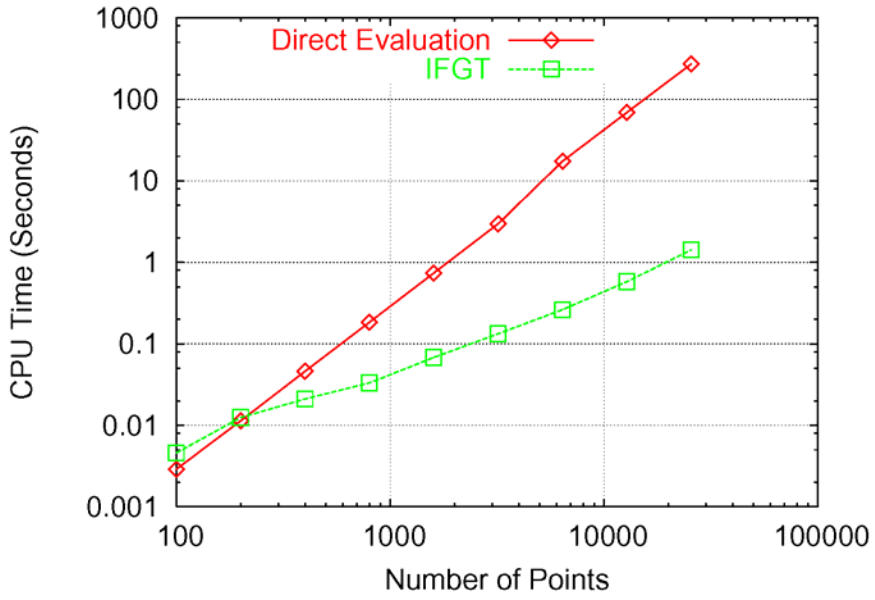


Fig. 5.3. Plot showing the CPU time (Pentium IV, 2.3 GHz) required for the evaluation of the Gaussian criterion for a given number of points in the datasets. The gains in computational costs increase dramatically with the size of point-sets.

6 RESULTS

The Bulk of this chapter is devoted to an experimental quantitative and analytic study of the Gaussian Fields registration method when applied to 3D registration. We also show results for multimodal image registration, 3D object modeling, and face reconstruction.

6.1 Three Dimensional Analysis, Objectives, and Methodology

One of the main advantages of the Gaussian Fields registration method is the relatively limited number of free parameters used. The only parameter that can change during the registration process is the force range σ which is, as we stressed in chapter 4, of fundamental importance to the overall algorithm. Hence, a large part of the analysis will be devoted to it. Other parameters on which the method depends are generally computed only once and in most cases are derived in a process similar to sensor characterization. For instance the main purpose of the de-correlation matrix Σ is to create the orthogonal features necessary for effective fusion. In addition a confidence parameter C_a, which we will discuss shortly, is used to control the effects of noise on the features used in the Gaussian criterion. The setting of its value is also based on our knowledge of the noise characteristics of a given sensor.

The experimental analysis that we have undertaken in this chapter attempts to be thorough. It will include the study of several aspects that were not jointly examined in a single work in the context of free-form shape registration. We recall the full

expression of the Gaussian criterion in (6.1). A short description of the parameters that appear in this expression is shown in Table 6.1.

$$E(Tr) = \sum_{\substack{i=1\ldots N_M \\ j=1\ldots N_D}} \exp\left(-\frac{d^2(P_i, Tr(Q_j))}{\sigma^2} - \frac{(S(P_i) - S(Tr(Q_j)))^T \Sigma^{-1} (S(P_i) - S(Tr(Q_j))))}{C_a^2}\right) \tag{6.1}$$

In addition to studying the effects of noise on the algorithm we will investigate the important issue of the role of local descriptors, studying their combination and fusion as well as the effect of the size of the areas over which they are computed. Another important aspect that we examine is the effect of the amount of overlap between the datasets to be registered which is crucial to all registration techniques. We also consider the complexity of shapes to be registered and the robustness of the Gaussian fields method to low levels of surface complexity. This first set of experiments is conducted on synthetic datasets to isolate and focus on the different factors being studied.

Table 6.1. Summary of the parameters used in the Gaussian Fields registration algorithm.

Parameter	**Description**
σ	The force range parameter, controlling the decay of the Gaussian Field
Σ	Covariance matrix of the feature descriptors used in the algorithm. This matrix is computed from the data
ρ	Radius of the ball in which local feature descriptors are computed
C_a	Confidence factor associated with the descriptors when dealing with noisy datasets
Tr	The actual registration parameter that we are recovering, our analysis is focused on rigid transformations

Several 3d datasets acquired by various sensors, which will be described in the following sections, are used to study the performance of the algorithm in one of its primary areas of applications, namely in the case of scene reconstruction from multiple 3D point-sets obtained from range images. The main focus in this second set of experiments is on the parameter σ and on designing an optimal scheme for accurate registration without close initialization. This required studying the registration error and basins of convergence of the algorithm for several datasets. A comparison of the region of convergence with the standard Iterative Closest Point algorithm was also undertaken. In the context of performance evaluation, the robustness of the algorithm to low levels of sampling was investigated on all the datasets. The set of 3D experiments conducted is summarized in Table 6.2.

Table 6.2. List of experiments conducted for the 3D analysis of the registration algorithm.

3D Synthetic experiments	**Experiments on real 3D datasets**
• Effects of noise • Analysis of local descriptors • Effect of the amount of overlap • Shape complexity and registration	• Sample plots around registered position • Study of the parameter σ • Analysis of the convergence: basins and schedule • Comparison of the basin of convergence with ICP • Effects of sampling

6.2 3D Synthetic Experiments

6.2.1 Noise effects

In the Gaussian criterion framework noise will influence both the position of the point-sets and consequently the descriptors that are computed from them. When considering very high levels of noise local shape descriptors can become so corrupted by noise that they are practically useless. It is for this reason that we added to our criterion the confidence factor C_a which will balance the contribution of the descriptors with respect to the coordinates. Given that the descriptors are scaled so that there variance is 1, the confidence parameter is set to a low value (typically less than 10^{-3}) for datasets with low-to-moderate noise levels, and will be higher than unit value for very high levels of noise.

We will focus our experimental analysis on uniform noise. Dealing with uniform noise constitutes studying a worst case scenario given that in practical applications noise is more biased in one direction (usually the radial component with respect to the camera's coordinate frame is dominant). In this experiment we use a 3D model of a head which we divide into two partially overlapping sections (Fig. 6.1). To each of these datasets we add uniform noise of amplitude going up to $\pm 12\%$ of the length of the head. Each of the head sections has about 3500 points. The main purpose of the experiment is to study the drift in the maximum of the criterion under the effect of noise. This is achieved by initializing the algorithm close to the ground truth registered position and starting the optimization scheme. The plots shown in Fig. 6.2 give the resulting rotation error in degrees and translation error as a fraction of the length of the model. They show the rate of increase in the error for two confidence values $C_a = 0.001$ and $C_a = 1.0$.

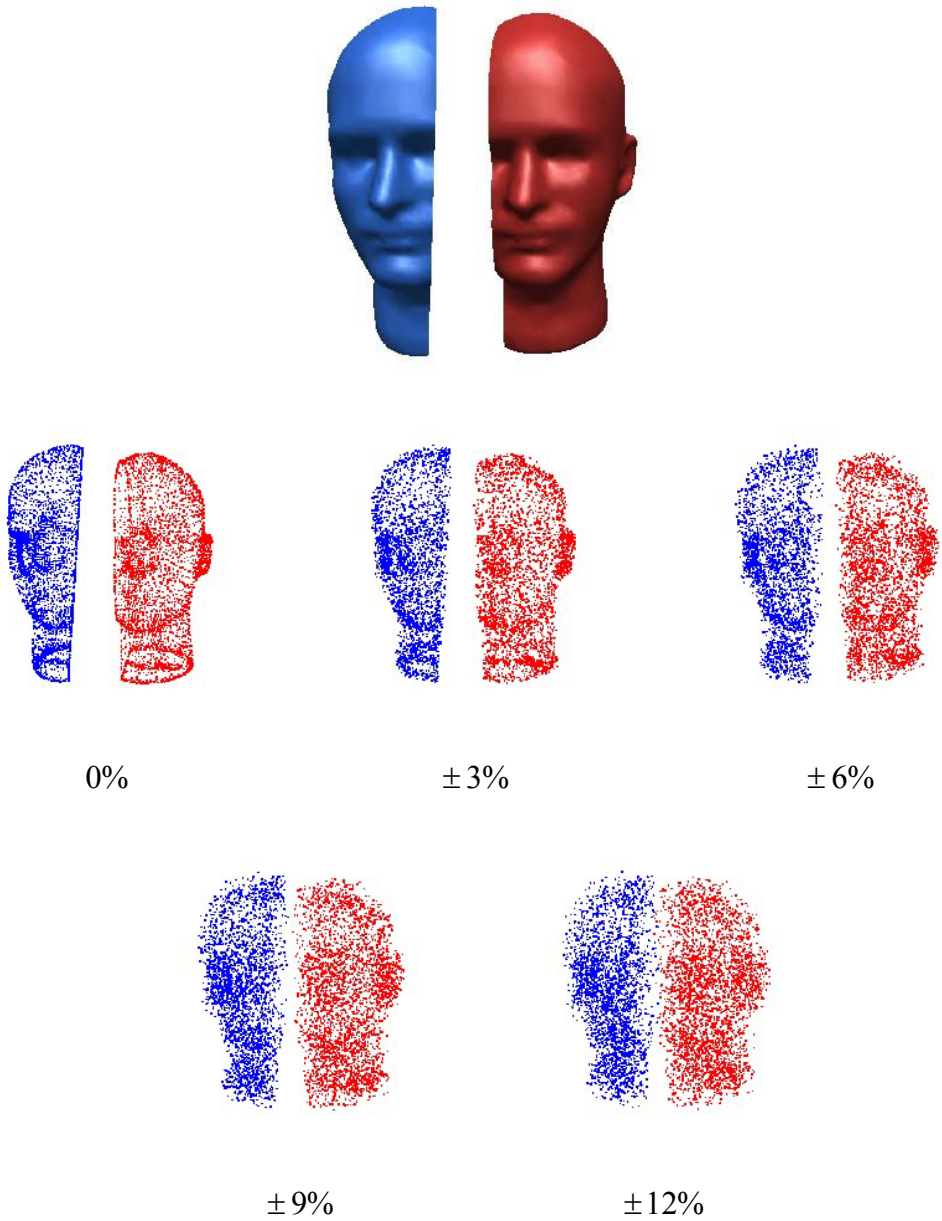
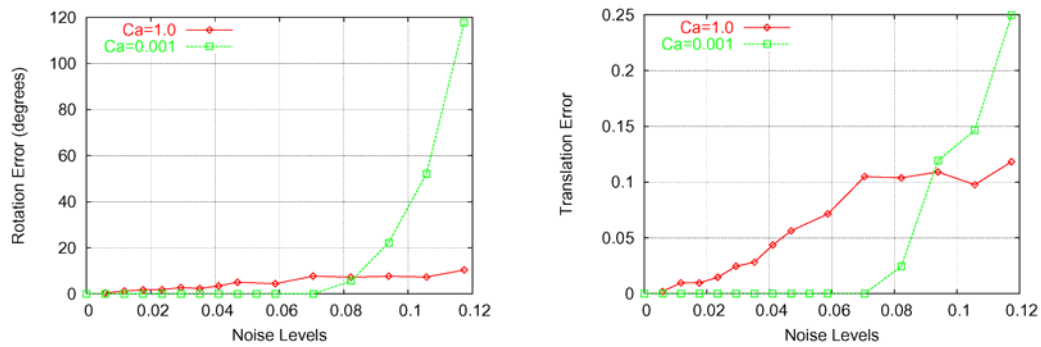


Fig. 6.1. Datasets used for the noise analysis. Two overlapping sections of a head model are used to study the effects of uniform noise. The point sets corrupted with increasing levels of uniform noise are shown with the noise value expressed as a fraction of the model's length.



(a) (b)

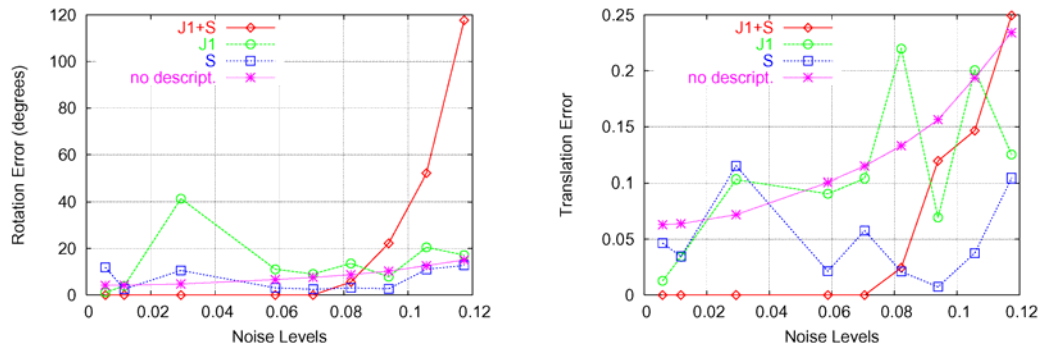
Fig. 6.2. Registration error versus uniform noise: (a) rotation error in degrees, (b) translation error as a fraction of the length the head model. We show plots for two values of the confidence parameter, low and high.

The first conclusion that we draw from this experiment is that the algorithm is stable for levels of uniform noise up to $\pm 8\%$ which is by any practical standard very high. In the plots we clearly see the role that the confidence parameter plays in moderating the sharp increase in registration error for the higher levels of noise. This, predictably, comes at the expense of a more accurate registration in the low to moderate noise range. Such behavior is caused by the forfeiting of part of the discriminatory power that the descriptors add to the algorithm. Note that we have employed here the two descriptors J_1 and S computed within a ball of radius 5% the length the head. These same settings are also used for the experiment that we describe next.

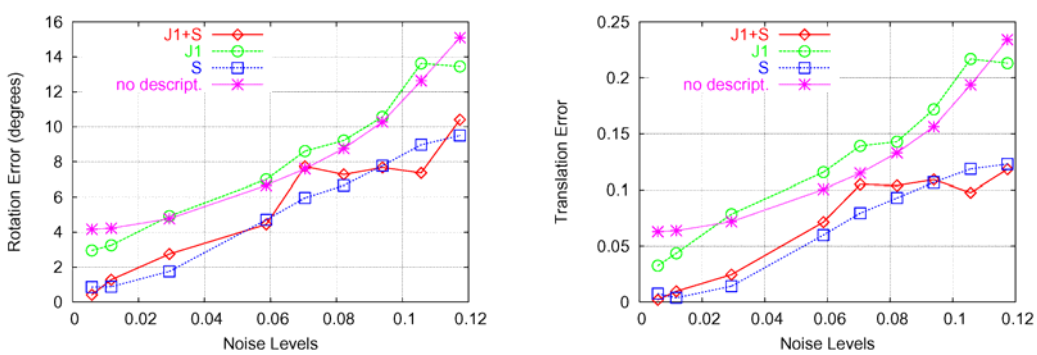
6.2.2 Local descriptors and fusion

One of the strengths of using the Gaussian criterion for free form shape registration is the ease with which multiple features can be fused. As discussed earlier for the method to benefit from these features a de-correlation step needs to be performed. The covariance matrix is computed in nearly flat regions and used to scale the features so that they have a unit variance, and so that they are independent. To study the quality of the features that we employed we use the same head dataset along with uniform noise. Our purpose is to assess the increase in registration error in the cases: (1) without attributes, (2) for each attribute being used alone, and (3) for the fused case. Plots of Fig. 6.3 show the evolution of rotation and translation errors with levels of uniform noise. We show results for both a low confidence factor ($C_a = 0.001$) in Fig. 6.3(a) and for a high value ($C_a = 1.0$) in Fig. 3(b). In the first case the fused criterion, as expected, outperforms the other cases in the considered range with an error that is practically zero and that increases fast for noise levels higher than $\pm 8\%$. For the experiment without attributes we notice a steady but slow increase in the registration error. The algorithm without any attributes actually outperforms the algorithm in the case we use J_1 alone; not surprising given that this feature is just an average of Euclidean distances to neighboring points, which does not account well for local shape variations. The second best performance after the fused case is obtained by the saliency descriptor S which as we saw was derived from the robust tensor voting framework.

This experiment shows clearly the usefulness of fusing multiple features which leads to a much better results than what each feature will give alone. Also we notice that fusion works well even if we use features that are coarse descriptors such as J_1. For $C_a = 1.0$ the gap in performance between the different criteria is reduced, while the error for the moderate noise level is higher. The fused criterion will slightly outperform the others only for very high levels of noise.



(a)



(b)

Fig. 6.3. Shown are the rotation and translation errors versus levels of uniform noise for different descriptors and for the fused case. In (a) we have $C_a = 0.001$, and for (b) $C_a = 1.0$. The same head dataset was employed.

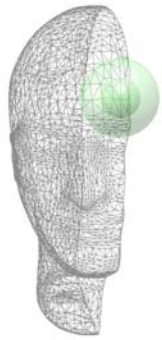
6.2.3 Local neighborhood size and features

One other aspect that we have investigated, which, to our knowledge, was rarely studied rigorously before is the effect on the registration algorithm of the size of the neighborhood over which the features are computed. This hole in the literature maybe due to the fact that this size can be deducted heuristically from the scale and noise characteristics of the datasets used. The size of the neighborhood will depend on the resolution and on the information content of the datasets. If the noise level is very low then one must choose the smallest neighborhood possible. This will spread the values of the descriptors over a larger spectrum allowing for more accurate matching. On the other hand in the noisy case features computed on such small domains will not be reliable enough to allow for accurate correspondence, hence the need to increase the size of the local area used for computation.

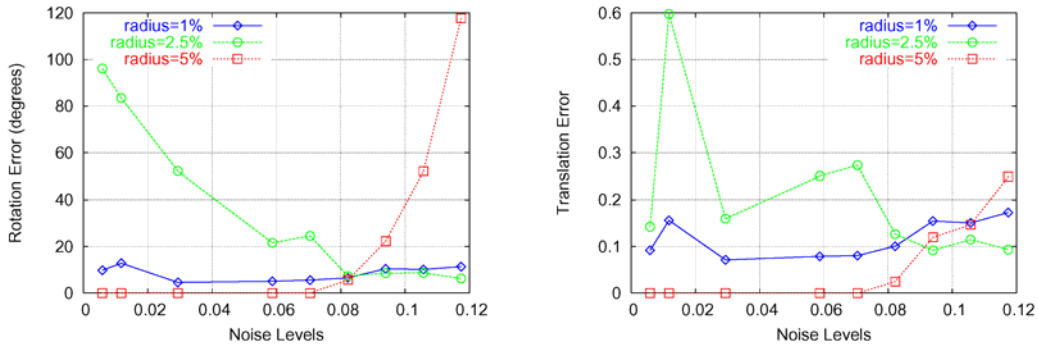
When dealing with point-sets the local neighborhood is usually a ball of radius ρ centered at the point for which the descriptor is computed. Using the same head dataset we examined the effect of uniform noise with three different values of ρ. Here again we study both high and low confidence levels. From the plots shown in Fig. 6.4 we deduced that for the used dataset a best behavior is obtained for $\rho = 5\%$, applying noise levels considered practical in our earlier discussion. This experiment gives an idea on the empirical study we can conduct to calibrate the value of ρ in actual applications with real datasets. In such applications, for a given sensor, a similar analysis involving runs of the registration algorithm with different values of the neighborhood radius on a typical sample will provide the optimal ρ.

6.2.4 Amount of overlap

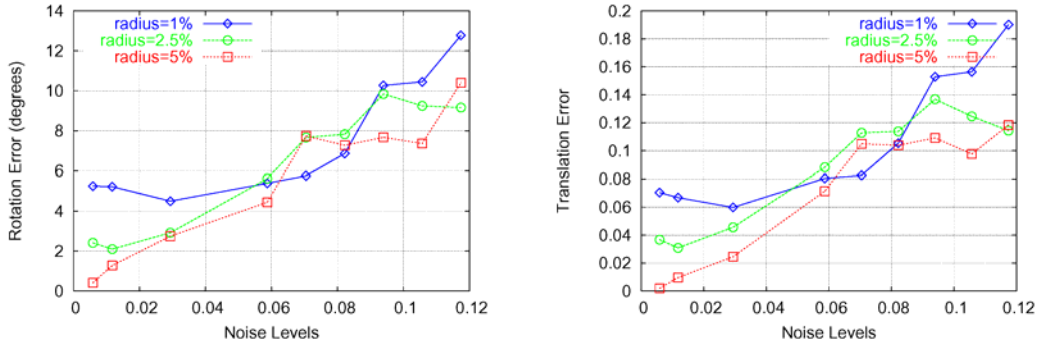
One of the fundamental issues in registration, and pattern matching in general, is the effect of outliers, by which we mean areas of the datasets that are not shared by them.



(a)



(b)



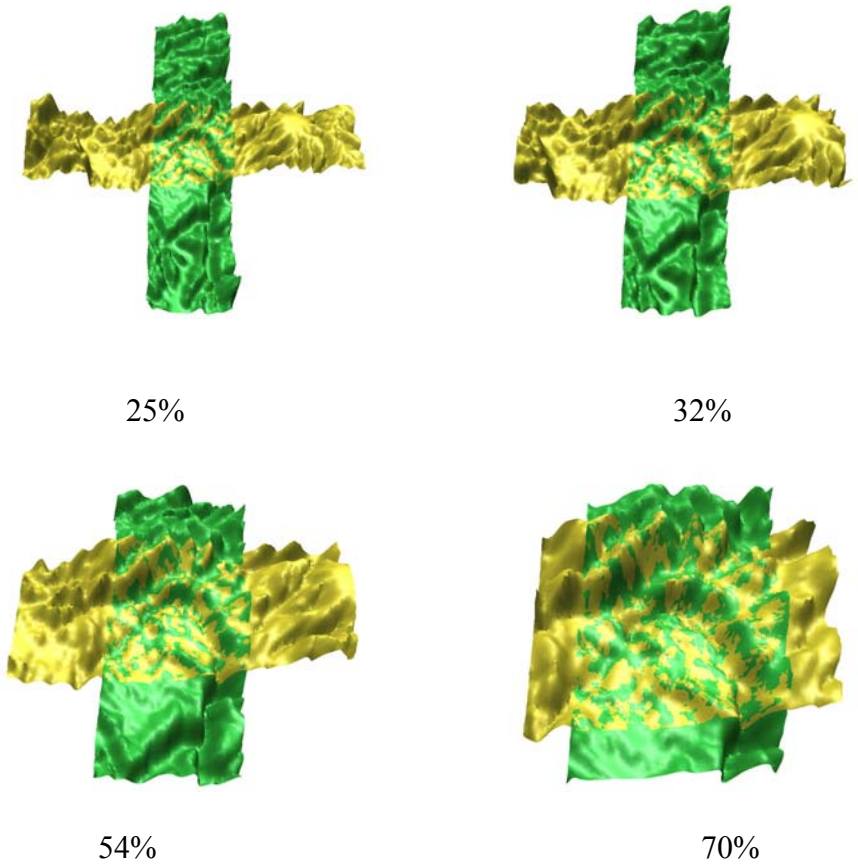
(c)

Fig. 6.4. Effects of the size of the local neighborhood used in feature computations. Rotation and translation errors are shown for different values of the radius of the local ball. The inner and outer balls (a) have radii of 5% and 10% of model's length respectively. In (b) we used $C_a = 0.001$, and in (c) $C_a = 1.0$.

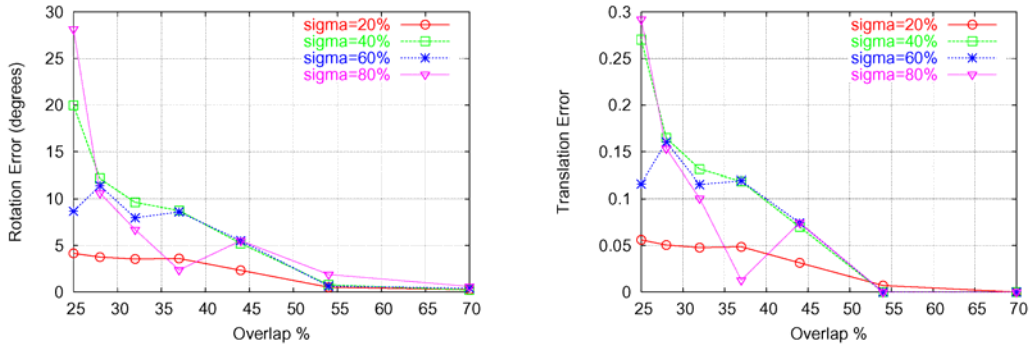
The lower the relative overlap between the point-sets the harder the correspondence problem becomes. Even for human beings. This is well accounted for in the formulation of our criterion, as we showed in chapter 4. Outliers will result in a drift of the maximum of the Gaussian criterion away from the correct position. This drift will only be moderated by a decrease in the value of σ (which itself results in a smaller region of convergence). The error can also be moderated by high complexity in the datasets. To study the effect of overlap on our registration algorithm we used a surface reconstructed from a digital elevation map (DEM). Several partially overlapping surfaces are generated from this DEM. Fig. 6.5(a) shows four surfaces in the registered position with an amount of overlap ranging from approximately 25% to 70%. In this analysis we keep the same common area and increase the outliers. The drift of the criterion's maximum caused by the outliers is studied for several pairs and for four values of σ (20%, 40%, 60%, and 80%). The results are summarized in the plots of Fig. 6.5(b). These show that up to about 50% of overlap the criterion maintains a very low localization error, which starts to increase rapidly for values of overlap less than 30%. As expected the slowest drift is the one corresponding to the lowest value of σ. The other curves evolve closely for most of the range, with the ones corresponding to the two higher values showing a more oscillatory behavior. This experiment gives an idea about the setting of force range parameter to minimize the effect outlier. The analysis also shows that for practical applications it is suitable to have at least around 40% to 50% overlap. Nevertheless the algorithm is able to handle lower values, at the cost of paying more attention to the way values of the force range parameter σ are adapted during the optimization scheme.

6.2.5 Surface complexity

The issue of surface complexity was studied before in the context of 3D registration, mostly with the aim of introducing, physically, in the scene objects that allow current algorithms to function properly in the case of low levels of shape information [Pito97].



(a)



(b)

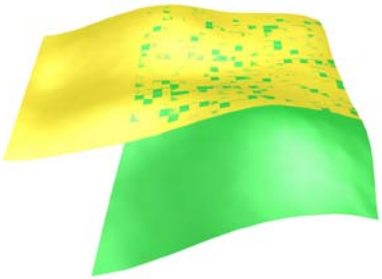
Fig. 6.5. The amount of overlap between two surfaces and its effect on the registration accuracy: Four of the DEM pairs used in this experiment are shown in (a). Plots of the rotation and translation errors versus the percentage of overlap are shown in (b).

This approach is not practical in many real applications. Our theoretical analysis of the Gaussian criterion addressed the point of shape complexity and showed that a high amount of surface information results in a decrease in the influence of the outliers. We devised a simple experiment to study the effects of surface shape variation on the registration accuracy. In this experiment we use pairs of partially overlapping DEMs (with about 60% of overlap) that were generated from a larger terrain model. In the overlapping we have a point to point mapping. Several pairs were created at different levels of averaging of the height of the DEM's (Fig. 6.6(a)). The evolution of the localization error of the criterion as it varies with the variance of the height, taken here as a very simple measure of surface complexity, is shown in the plots of (Fig. 6.6(b)). For this noiseless setting and with a value of σ equal to 10% of the length of the datasets we have found that the drift of the maximum of the Gaussian criterion is extremely low. This shows that the algorithm is able to handle low levels of surface complexity.

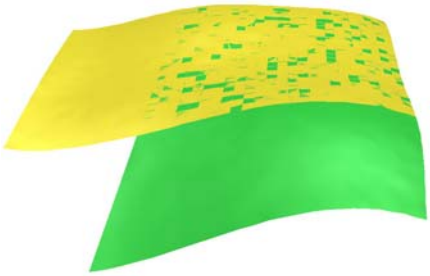
6.3 Experiments with 3D Real Datasets

In our second set of experiments we used 3D models reconstructed from range maps. To show the scope of the algorithm we employed several 3D range sensors [Besl88] operating at largely different scales. The first sensor employed in this analysis, the Riegl LMS Z-210 (Fig. 6.7(a)), is a laser-based system that uses the time of flight principle. The Riegl's operating range goes from about 2 m to 350 m, with an estimated noise level of about 20 mm and a range resolution of 25 mm. The Riegl is mostly used for scanning large scale scenes such as buildings and for monitoring construction and mining projects. The second scanner is the IVP Ranger 2200 (Fig. 6.7(b)), also a laser scanner, but one that is based on the principle of triangulation. The Ranger operates by acquiring several profiles of a scene.

Z-variance = 25.0

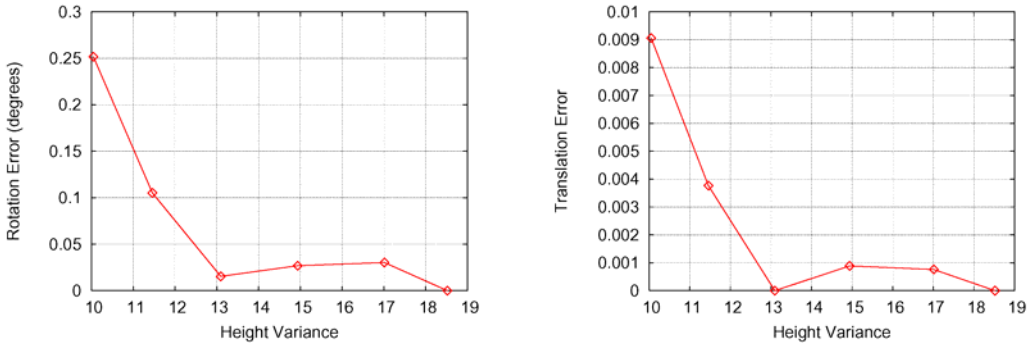


Z-variance = 18.52



Z-variance = 17.0

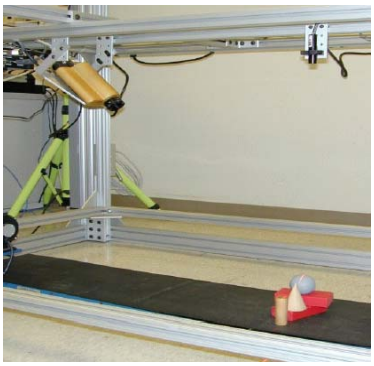
(a)



(b)

Fig. 6.6. Study of the registration error versus surface complexity: a pair of DEMs with increasing levels of averaging is shown on (a) with their height variance (the terrain size is 100×150). Plots of the rotation and translation errors for different height variance levels are shown in (b).

(a) Riegl LMS Z-210



(b) IVP-Ranger2200

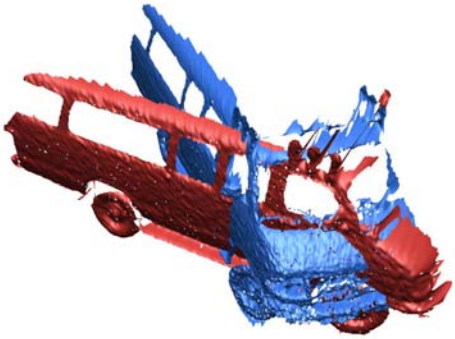
(c) Confocal Microscope Leica SP2 LSCM

Fig. 6.7. The three range scanning systems used in our registration experiments. The Ranger (a) system is mostly used for the acquisition of medium sized objects such as automotive parts. The Riegl (b) is primarily employed for large scale scenes such as buildings, and the confocal Microscope in the sub-millimeter domain.

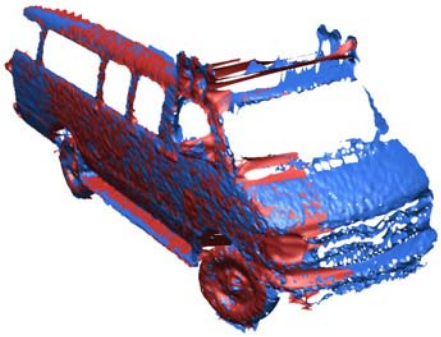
The scanner has a camera with a two-dimensional CMOS sensor that captures a laser sheet of light which illuminates the imaged objects. Knowledge about the geometry of the system, such as the relative location of the camera and the laser source as well as the camera's internal and external parameters, allows the 3D reconstruction of a given profile. Each profile contains 512 samples. A controlled relative motion between the camera and the scanned objects allows for the acquisition of a 2D range map. The resolution of this scanner is of about 2mm, and it is mostly used for mapping objects measuring several tens of centimeters in length.

For smaller objects we acquired 3D imagery with a Leica SP2 LSCM confocal microscope (Fig. 6.7(c)) that resolves objects as small as 1 μm. The confocal microscope uses depth from focus to build several layers of the scanned objects. A full three dimensional model is reconstructed by assembling these layers. The system is commonly used to inspect small scale components and for both material science and life sciences applications. We acquired six datasets using these sensors, two per scanner. The datasets consist of a pair of 3D scans of various objects corresponding to the typical operating ranges of the cameras. The two scans are acquired from significantly different viewing points. To build a ground truth registration the surface is reconstructed from the original high resolution scans, corresponding points are picked by hand and the registering transformation is computed using the classic absolute orientation SVD-based technique [Horn87]. A refinement step is performed using an extended ICP version that takes into account normal information. In our actual analysis we use lower resolution sub-sampled models (less then 10000 points), and all the results are obtained in the context of point-set registration. Figures 6.8 to 6.13 show the different datasets used in the experiments. The first pair of 3D views acquired by the Riegl is of a 14 passenger Van. The reference 3D view (blue) has a bounding box of dimensions $5.28 \times 4.42 \times 2.53$ m.

(a)



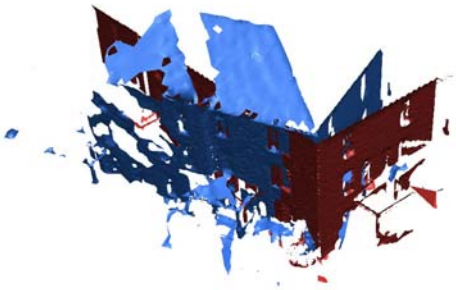
(b)



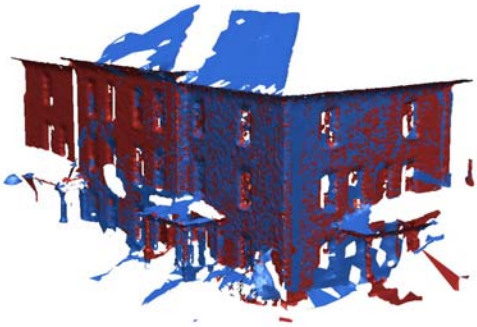
(c)

Fig. 6.8. Van Dataset: Color image (a), 3D scans in unregistered (b) and registered (c) positions. Scans acquired by the Riegl system. The dimensions of the van are $5.28 \times 4.42 \times 2.53$ m.

(a)

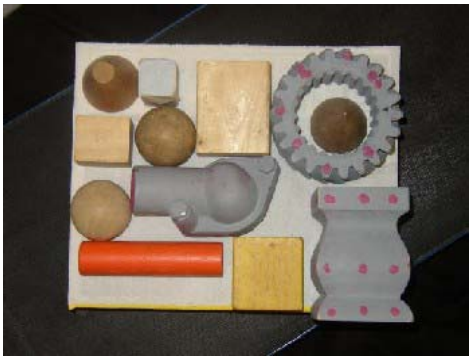


(b)

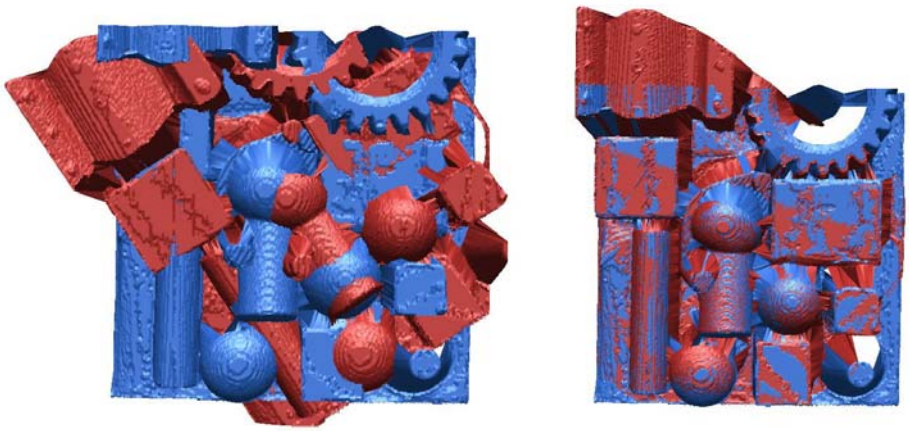


(c)

Fig. 6.9. Building Dataset: Color image (a), 3D scans in unregistered (b) and registered (c) positions. Data acquired by the Riegl scanner. The dimensions of the building scene are $33.59 \times 21.87 \times 12.82$ m.



(a)

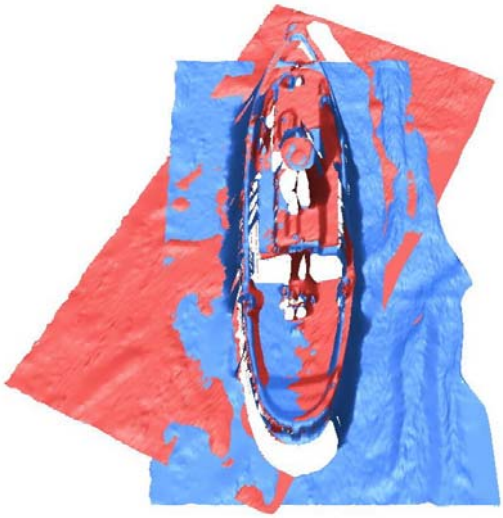


(a) (b)

Fig. 6.10. Parts Dataset: shown are the color image of the scanned objects (a) and two reconstructed scans in unregistered (b) and registered (c) positions. The Ranger scanner was used to acquire this dataset. The dimensions of the scene are $237 \times 235 \times 127$ mm.

(a)

(b)



(c)

Fig. 6.11. Boat Dataset: Color image of the model boat (a) and two reconstructed scans shown in unregistered (b) and registered (c) positions. Scans acquired by the Ranger system. The dimensions of the scene are $462 \times 273 \times 140$ mm.

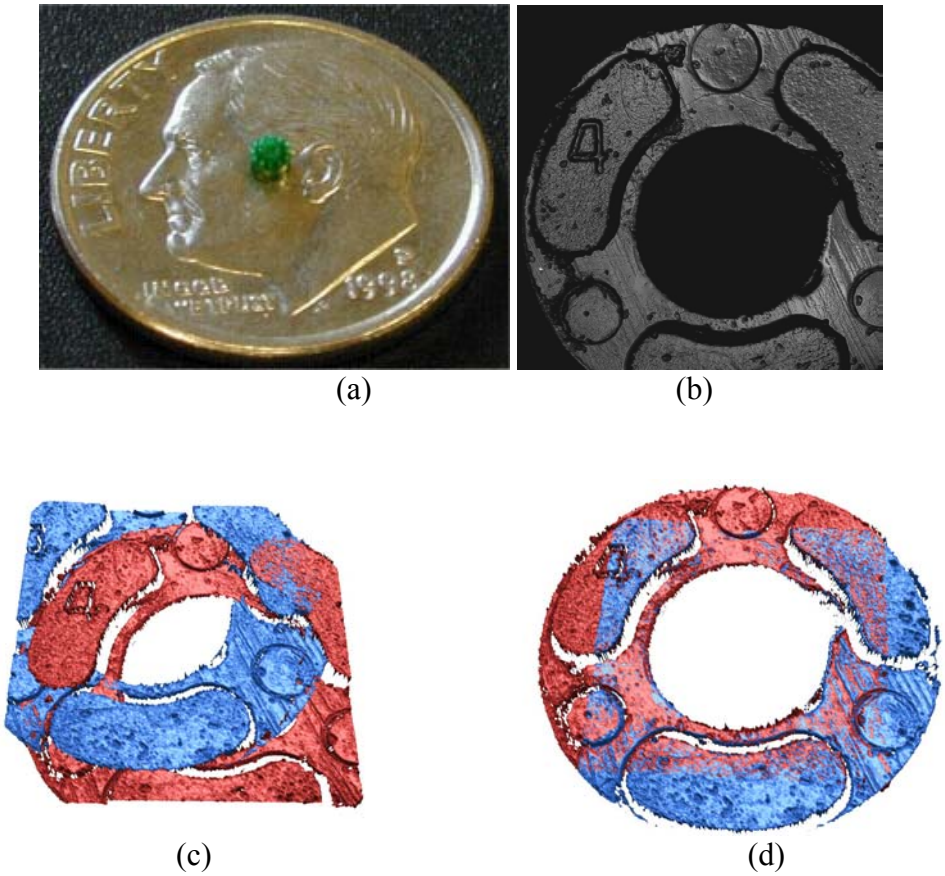


Fig. 6.12. Gear Dataset: image of the green gear showing a scale comparison with a dime (a). In (b) we show the intensity image obtained by the Leica confocal microscope. Two 3D reconstructions from confocal slices are shown in unregistered (c) and registered (d) positions. The dimensions of the merged views are $1385 \times 1462 \times 125\,\mu m$.

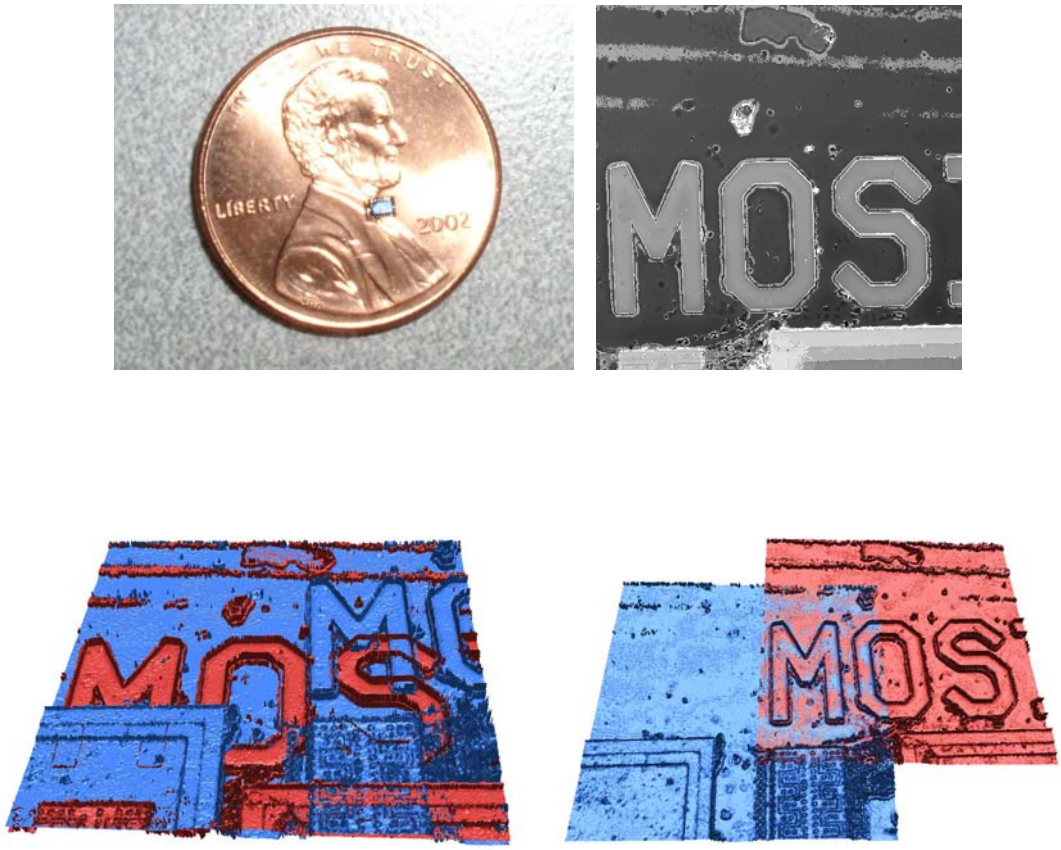
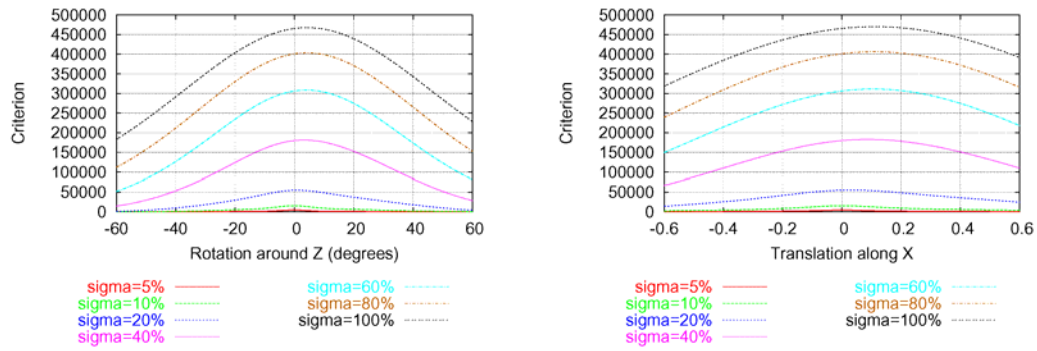


Fig. 6.13. Microchip Dataset: image of the microchip showing a scale comparison with a penny (a). In (b) we show the intensity image of part of the surface of the microchip obtained by the confocal microscope. Two 3D reconstructions from confocal slices are shown in unregistered (c) and registered (d) positions. The dimensions of the merged views are $149 \times 149 \times 8\ \mu\text{m}$.

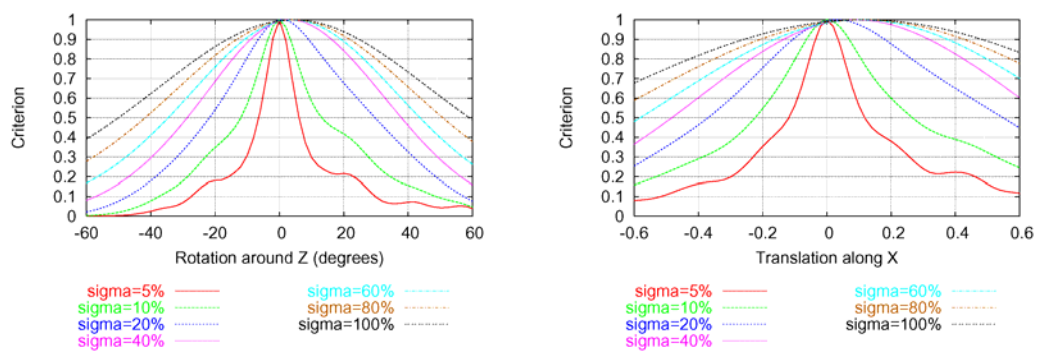
In addition to this 'Van' dataset (Fig. 6.8), we acquired a 'Building' datasets with size $33.59 \times 21.87 \times 12.82$ m (Fig. 6.9). The first dataset captured by the Ranger 2200 consists of a set of parts and objects that we call 'Parts' dataset (Fig. 6.10), the reference view has dimensions $237 \times 235 \times 127$ mm. The second pair of scans is of a model boat. The scene has dimensions $462 \times 273 \times 140$ mm and will be called 'Boat' dataset (Fig. 6.11). Using the Leica confocal microscope a further two pairs of views were obtained. First a couple of scans of a small gear shown in Fig. 6.12 (referred later as 'Gear' dataset $1385 \times 1462 \times 125$ μm). Fig. 6.13 shows the second set of confocal microscope 3D views focused on a small area of a Microchip ('Microchip' dataset $149 \times 149 \times 8$ μm).

6.3.1 Plots around the ground truth

We start by plotting the profile of the Gaussian criterion around the ground truth registration for all the datasets. The behavior is quite similar for all the six pairs of views, with the usual dominant mode resembling the one of the simple 2D experiment shown in chapter 4. We show in Fig. 6.14 sample plots obtained using the 'Parts' dataset. The plots are generated for various values of the Gaussian force range parameter σ ranging from 5% to 100% of the length of the datasets. The original plots for one translation parameter along with the plots for one rotation parameter are shown in Fig. 6.14(a). The increase in sigma results in an increase of the amplitude (dimensionless) of the criterion. It also leads to a drift in the position of the maximum away from the correct position, a behavior that we explained in the theory. To emphasize the increase in the width of the dominant mode of the Gaussian criterion we show the same plots scaled by the maximum in Fig. 14(b). In these latter plots we can see an increase that is almost proportional to the increase in the values of σ for the lower ranges of this parameter but which is slower for the higher ones. This behavior is discussed in the next sections.



(a)



(b)

Fig. 6.14. In (a) are shown the plots of the Gaussian criterion versus the rotation around the z-axis (perpendicular to the plan of view of the datasets in Fig. 6.8), and the translation along the x-axis (the horizontal axis in Fig. 6.8) for seven values of σ (as a % of the length). The plots in (b) are the scaled version (of unit maximum) of those in (a) emphasizing the relative increase in the width of the dominant mode.

6.3.2 Effects of varying sigma

From the plots generated for all six datasets, with the different values of σ we can measure the drift of the dominant mode in the different dimensions of the registration parameter space and plot their evolution. This is shown in Fig. 6.15 which renders the evolution of localization error with σ. The overall behavior is similar for the six datasets, in the sense that it starts with an almost linear increase in the drift as a function of σ. For larger values this drift is much slower and tends toward an asymptotic limit. This can be explained by the fact that, as shown earlier, the force range parameter controls the influence of outliers, hence the relatively rapid increase in the lower range in particular for a dataset with low amount of overlap such as the 'Microchip'. The asymptotic stabilization is explained by the fact that as σ exceeds the average distance between points in the datasets the exponential can be approximated by its first order development:

$$\exp\left(-\frac{d^2(Tr(P_i),Q_j)}{\sigma^2}\right) \approx 1-\frac{d^2(Tr(P_i),Q_j)}{\sigma^2} \tag{6.2}$$

The actual optimization problem will not in fact depend on σ anymore. We can easily show that in the case of large force range parameter the problem of maximizing the Gaussian criterion is equivalent to minimizing the sum of average distances from the points of one dataset to the other dataset (6.3).

$$\min_{Tr} \sum_{i=1\ldots N_M} \bar{d}^2(Tr(P_i),D) \tag{6.3}$$

with $\bar{d}^2(Tr(P_i),D) = \frac{1}{N_D} \sum_{j=1\ldots N_D} d^2(Tr(P_i),Q_j)$.

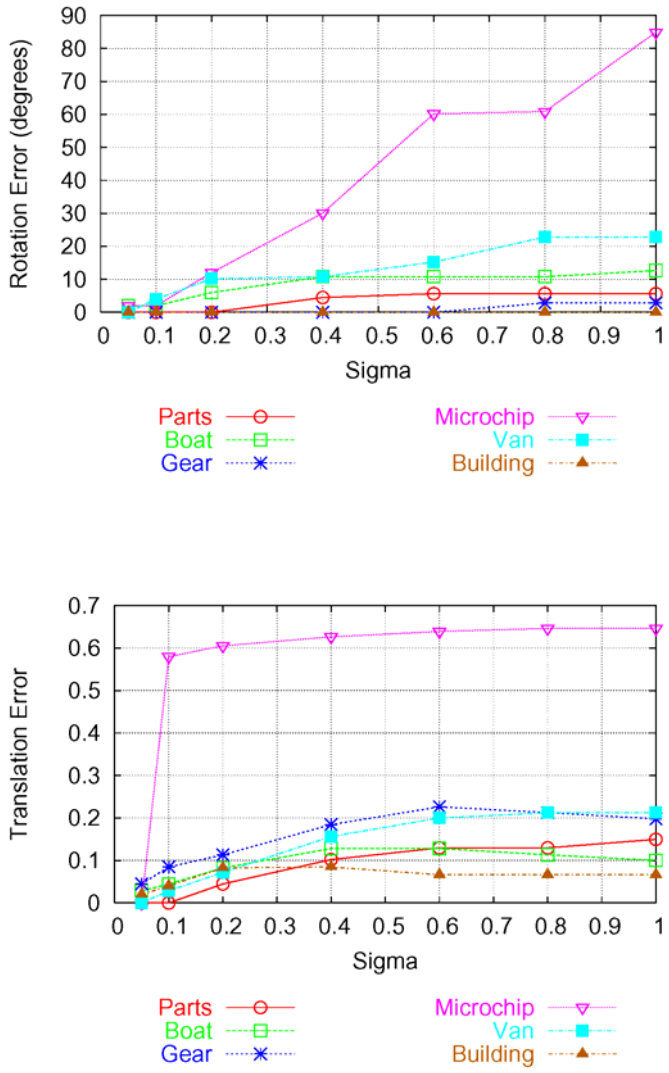
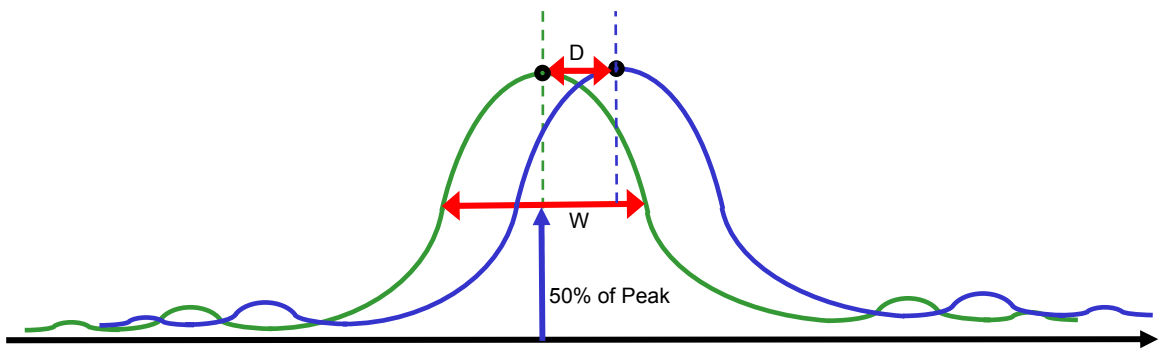


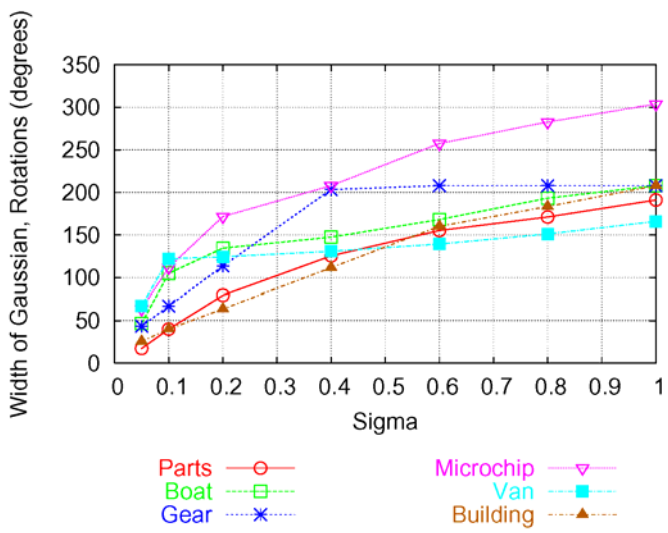
Fig. 6.15. Plots showing the evolution of the rotation and translation errors as a function of the parameter σ (as a fraction of the length).

This problem is independent of the value of σ, whence the asymptotic behavior. This fact is quite useful in our algorithm since it insures that the registration error is bounded from above, for the plots shown the maximum error for most datasets is between 7% and 10% for translation and between 2º and 20º for the rotation, excluding the 'Microchip' dataset which has an overlap area that we consider too small and which is very flat. Such dataset presents great difficulty for registration algorithms.

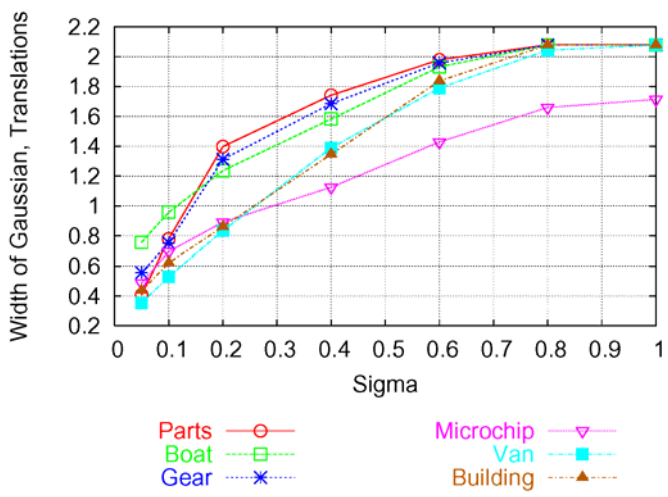
An important issue that we discussed in chapter 4 was the optimization scheme and the adequate choice of a schedule for σ, so that we can enlarge the basin of convergence while obtaining accurate alignment. A rigorous condition can be developed to avoid falling into local maxima when we reduce σ between two runs of the basic optimization routine. A study of the profiles of the criterion suggests that we can impose a simple constraint to avoid this problem. The dominant peak of the criterion is safely convex starting from about 50% its height. We will require that the drift of the maximum does not exceed half the width of the criterion at 50% of the peak. As shown in Fig. 6.16 this empirical condition will ensure that we avoid the local extrema. Fig. 6.17 shows plots of the width of the dominant peak at 50% as a function of σ. By comparing these with the plots of drift of maximum with respect to σ, we conclude that this condition is practically always fulfilled. These can be visualized also by directly observing the profiles of the criterion in Fig 6.14 where the drifting peak for higher σ does not fall into the rugged areas of the plots corresponding to the lower values of this parameter. It emerges from this that we will need only a two steps algorithm, starting with an initial rough alignment with large σ that will be followed by refinement step where σ is decreased sharply.



6.16. Finding a constraint on the schedule of σ between two runs of the optimization scheme. Making sure that the drift between the two steps (D) does not exceed half the width (W) of the dominant peak at 50% of its amplitude will avoid the local exterma.



(a)



(b)

6.17. Variation of the width of the Gaussian criterion at 50% of the peak (as explained in Fig. 6.16) as a function of σ (as a fraction of the length) for rotation (a) and translation (b) parameters.

6.3.3 Basins of convergence and comparison with ICP

The effect of σ on the region of convergence (ROC) can be clearly seen from the analysis of the two previous sections. But to have a better understanding of the scale of these Regions of Convergence, for our different real datasets, and to determine the nature of the relationship with the force range parameter we analyzed the basins of convergence of the algorithm for our datasets (Fig. 6.18). The plots of Figures 6.19 and 6.20 show the relationship between the initial value of the transformation parameters provided to the algorithm and the residual registration error at the end of the process. These so-called basins of convergence were obtained for several values of σ.

What these plots confirm is the tradeoff between a large basin of convergence for a large value of σ associated with a large residual error as well, and a smaller basin of convergence for lower values of σ that come with a better registration accuracy. This fact argues again for the two-steps scheme discussed before. We note that the width of the basins will grow fast first but then does not increase much after a certain value of the force range parameter which was already deduced from the profiles of the criterion. Also the width of these basins is significantly larger than the value of σ (generally around 10 times for values less then 5%). When these basins are compared with those of the point-based ICP algorithm (Figures 6.21 to 6.26) we notice that they are wider for all datasets even for low values of σ. This is to be expected, since we know that ICP is a close-range locally convergent scheme. On the other hand ICP has a smaller residual error except when compared with the algorithm tuned for close range Gaussian Fields. A balance between residual error and ROC size is clearly achieved by the adaptive optimization strategy. The resulting basin of convergence for a typical two-steps approach is shown in Fig. 6.27, showing the both the increase of the basin of convergence combined with the reduction of the residual error.

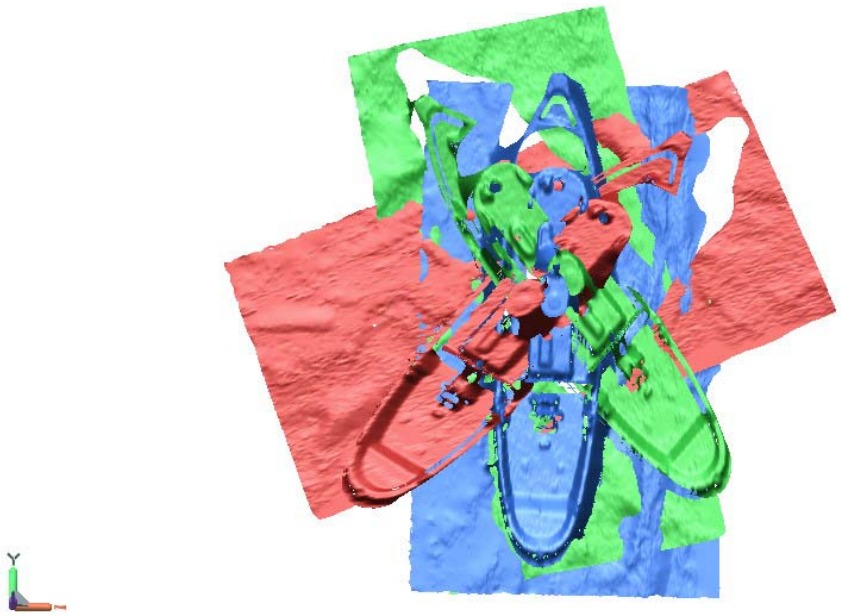


Fig. 6.18. For our convergence experiments we initialize one of the views in several positions around the registered position then we run the algorithm. The obtained residual registration error is used to measure the quality of the convergence. In this Figure we have two initializations (Red and Green) obtained by rotation around the z-axis of one of the views around the fixed reference one (Blue).

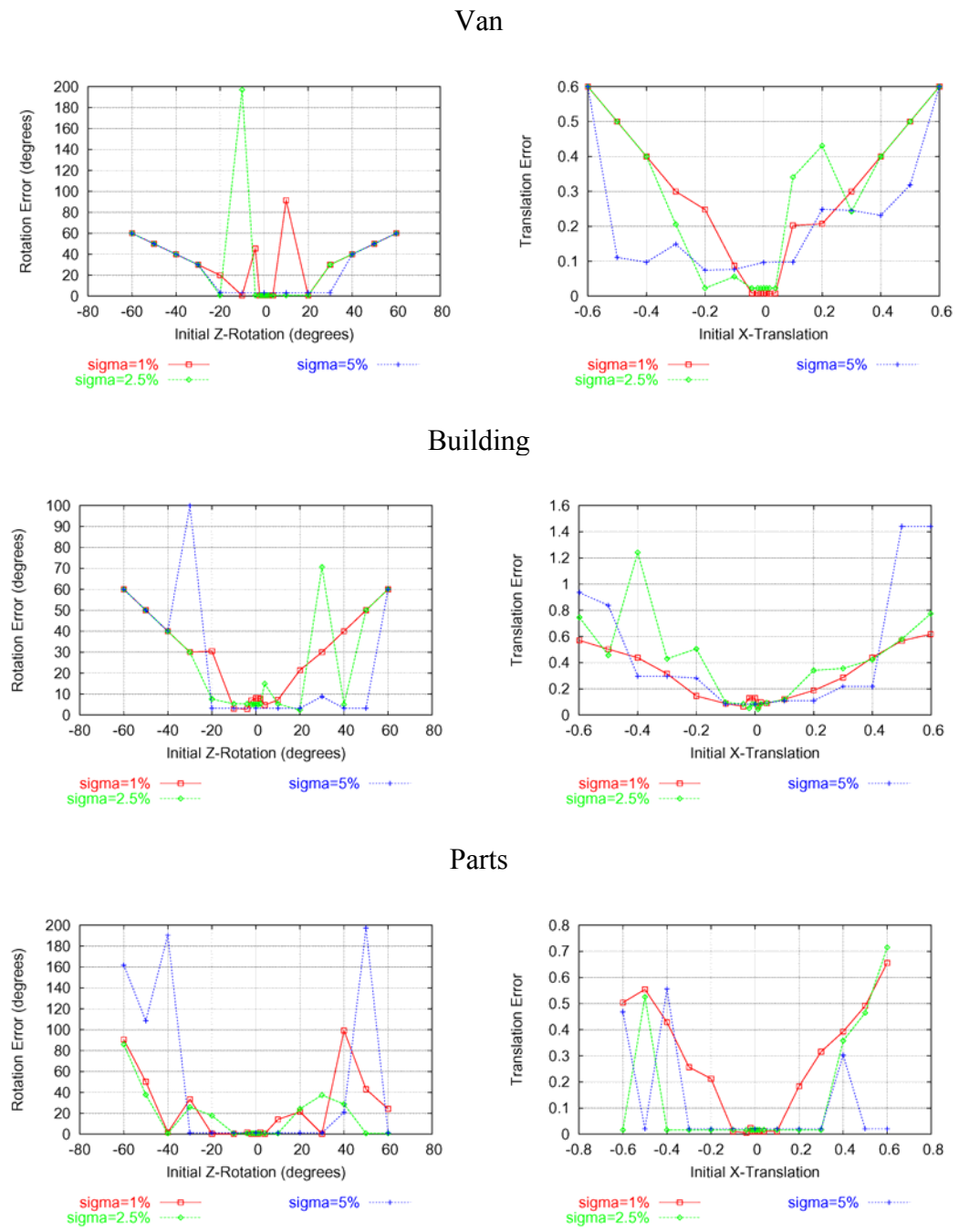


Fig. 6.19. Basins of convergence of the method showing residual registration error, as a function of initialization, for rotation around z-axis and translation along the x-axis. Datasets from top to bottom: Parts, Boat, and Gear.

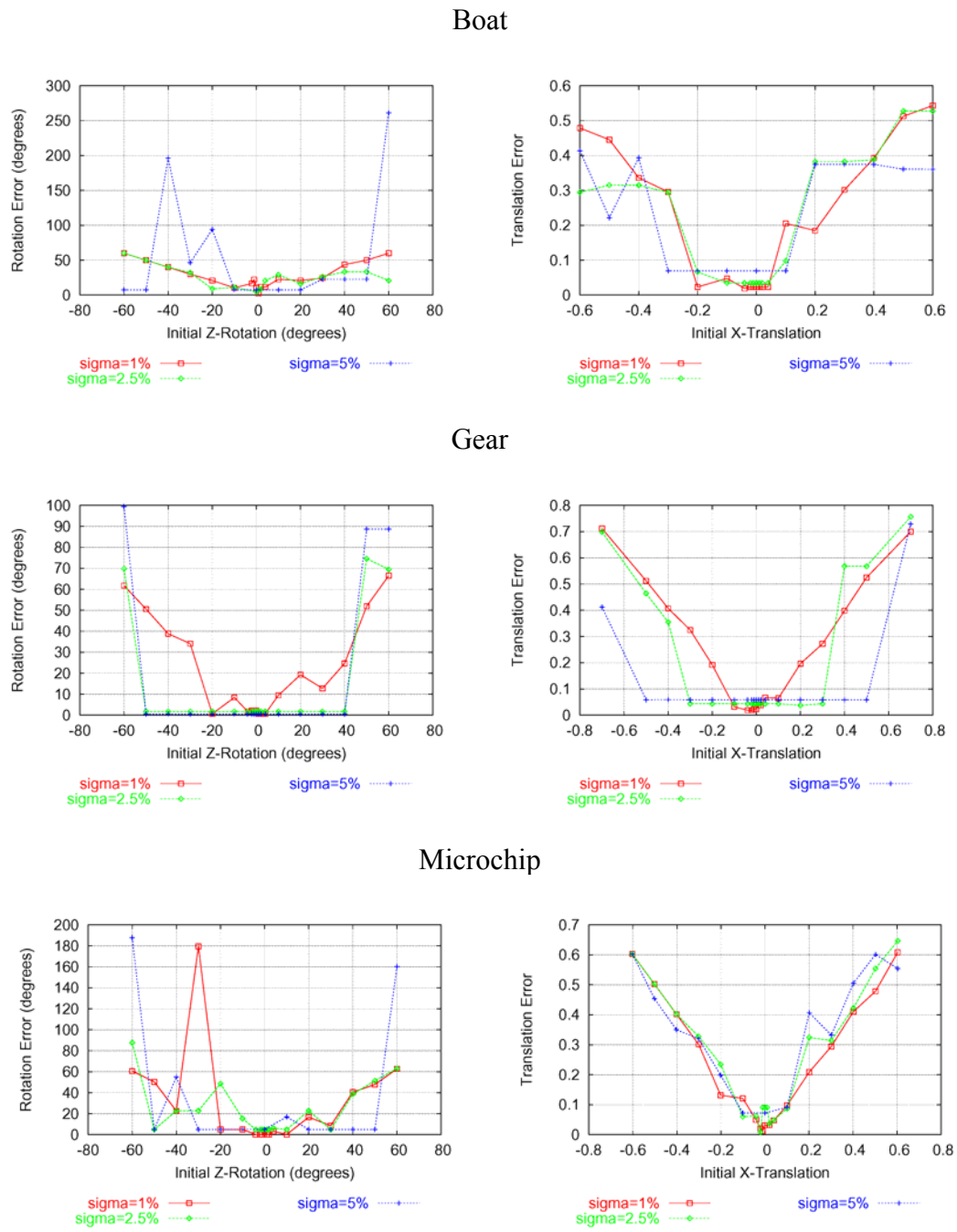


Fig. 6.20. Basins of convergence of our method showing residual registration error, as a function of initialization, for rotation around z-axis and translation along the x-axis. Datasets from top to bottom: Microchip, Van, and Building.

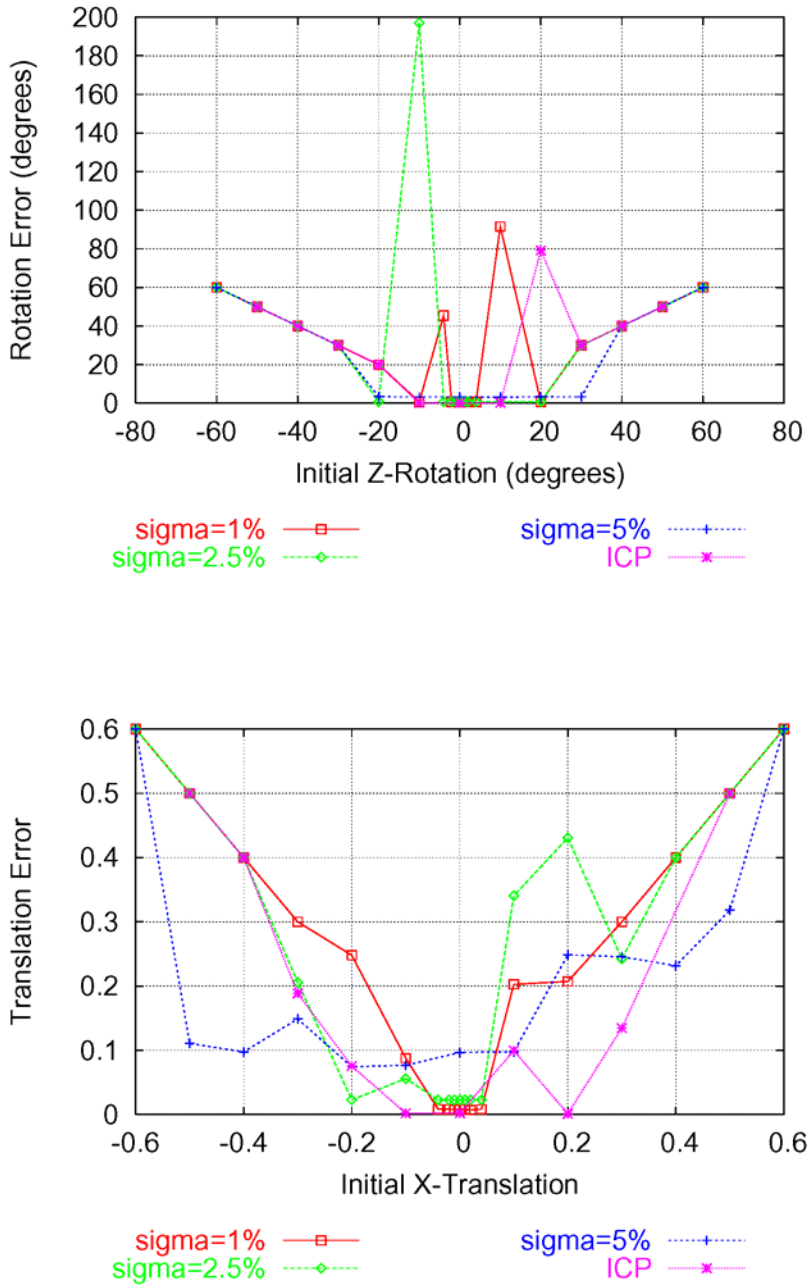


Fig. 6.21. Comparison of the method's Basins of convergence with the ICP basins: Van dataset.

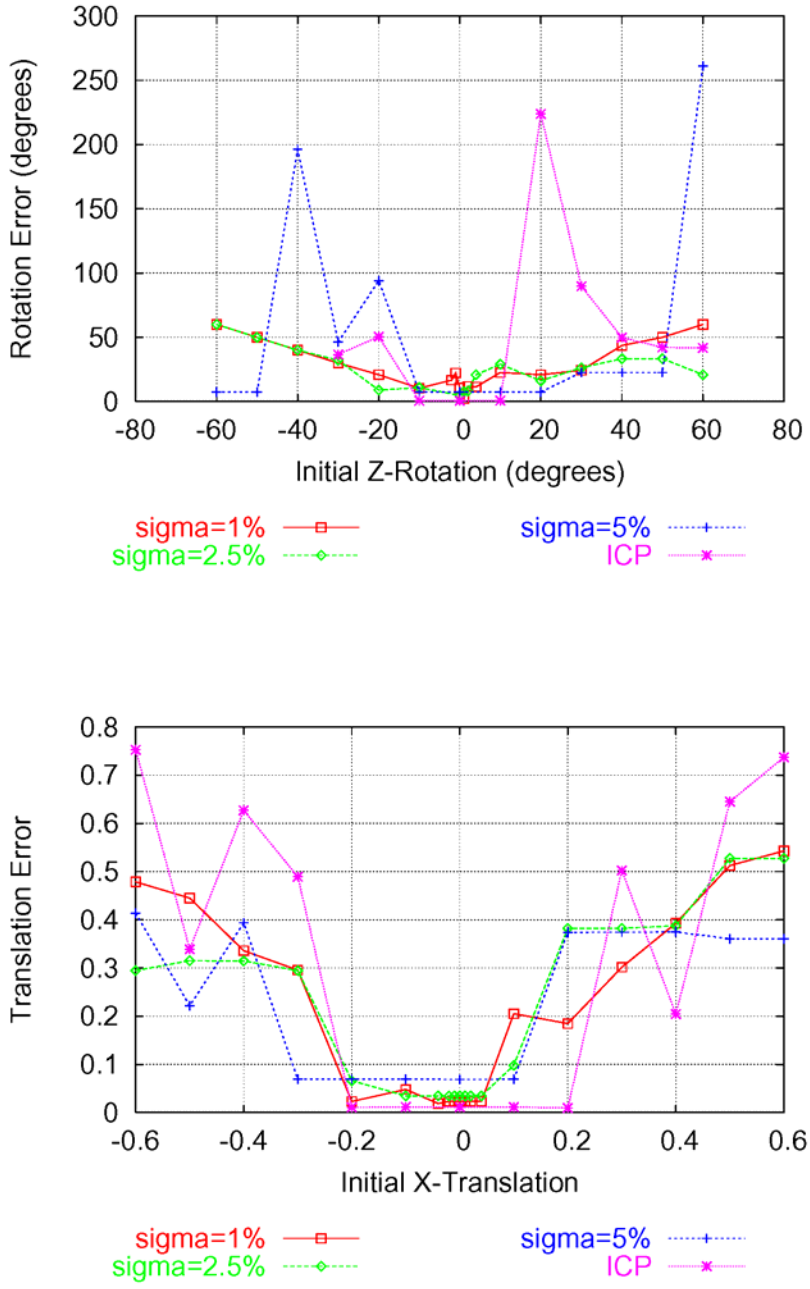


Fig. 6.22. Comparison of the basins of convergence with the ICP basins: Building dataset.

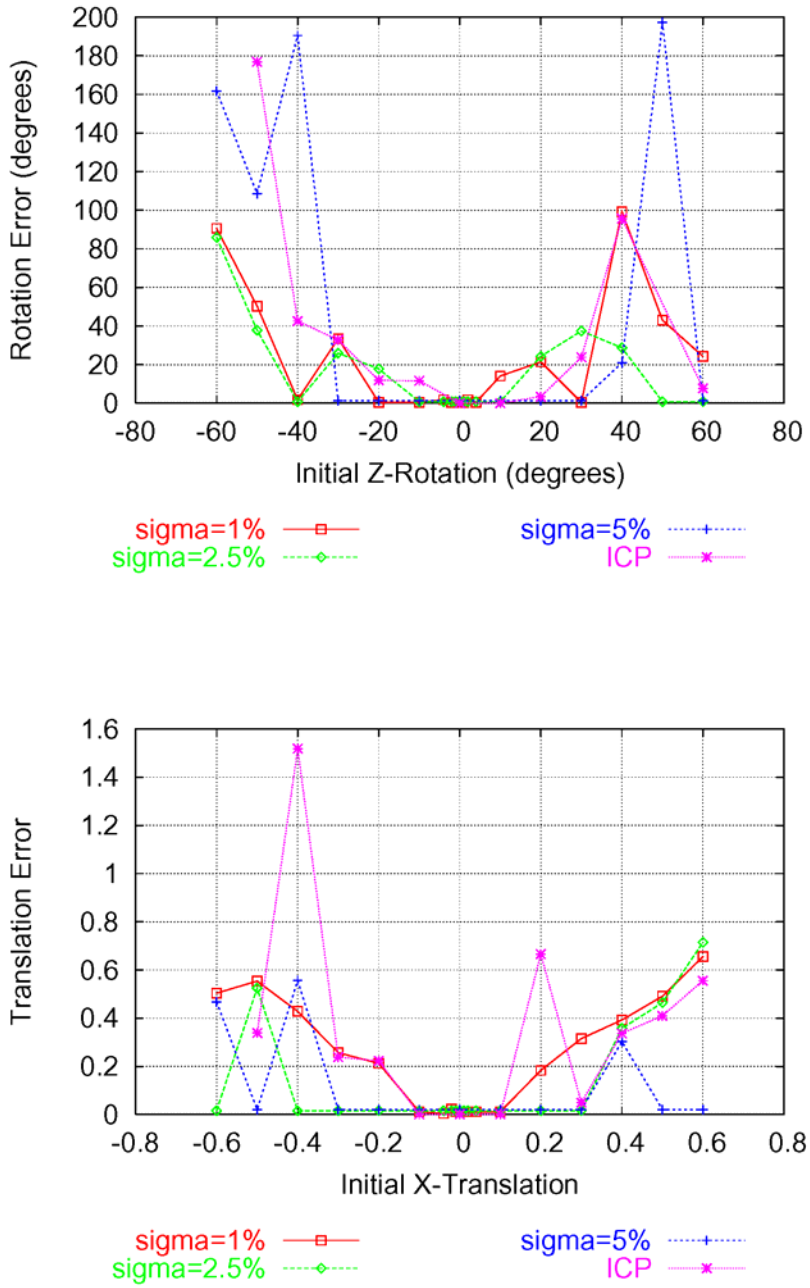


Fig. 6.23. Comparison of the method's basins of convergence for several values of σ with the ICP basin of convergence. Parts dataset.

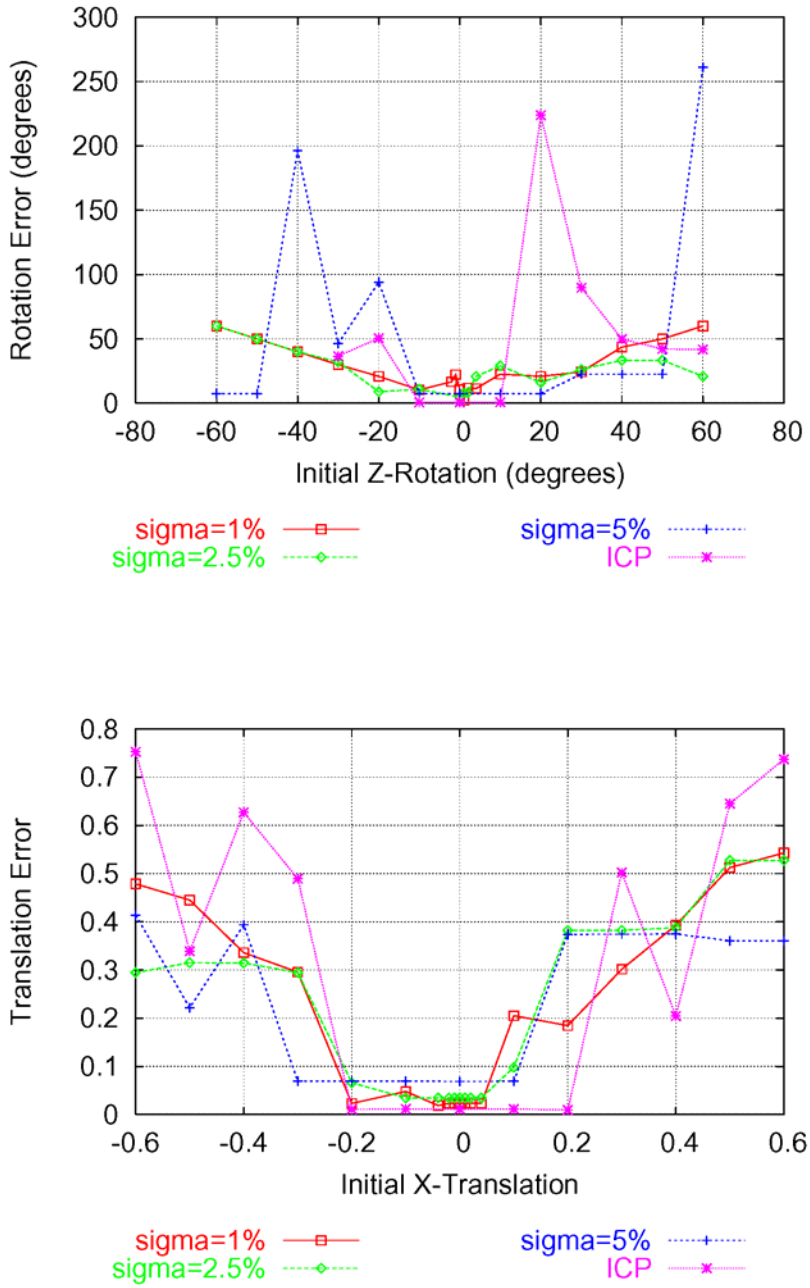


Fig. 6.24. Comparison of the method's basins of convergence with the ICP basins: Boat dataset.

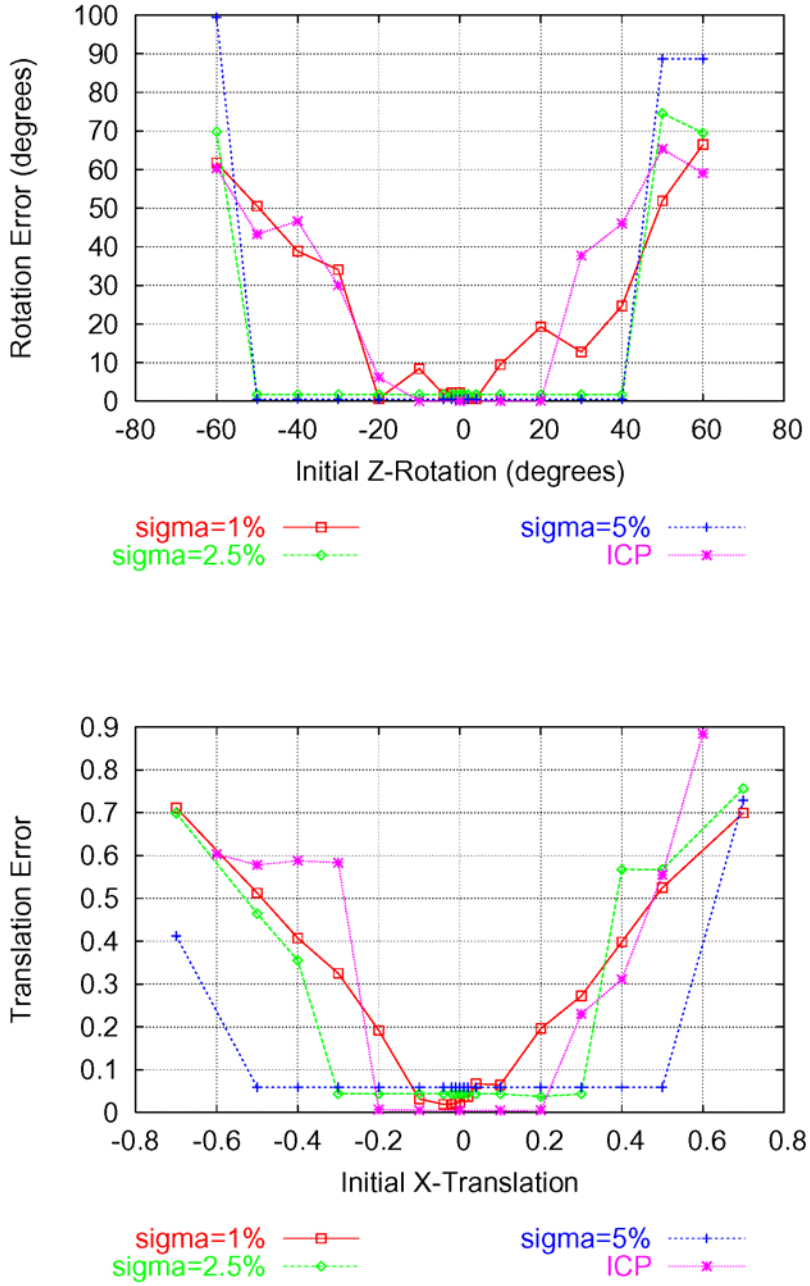


Fig. 6.25. Comparison of the method's basins of convergence with the ICP basins: Gear dataset.

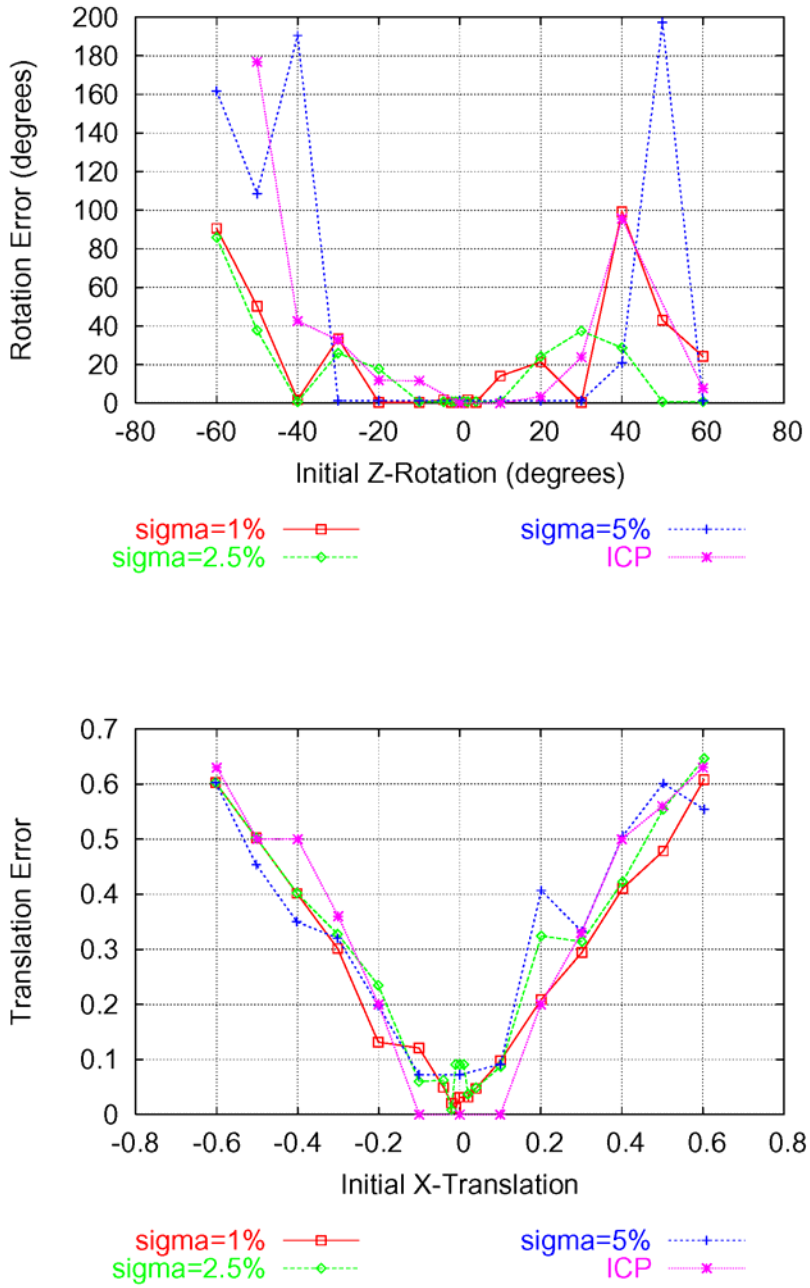


Fig. 6.26. Comparison of the method's basins of convergence with the ICP basins: Microchip dataset.

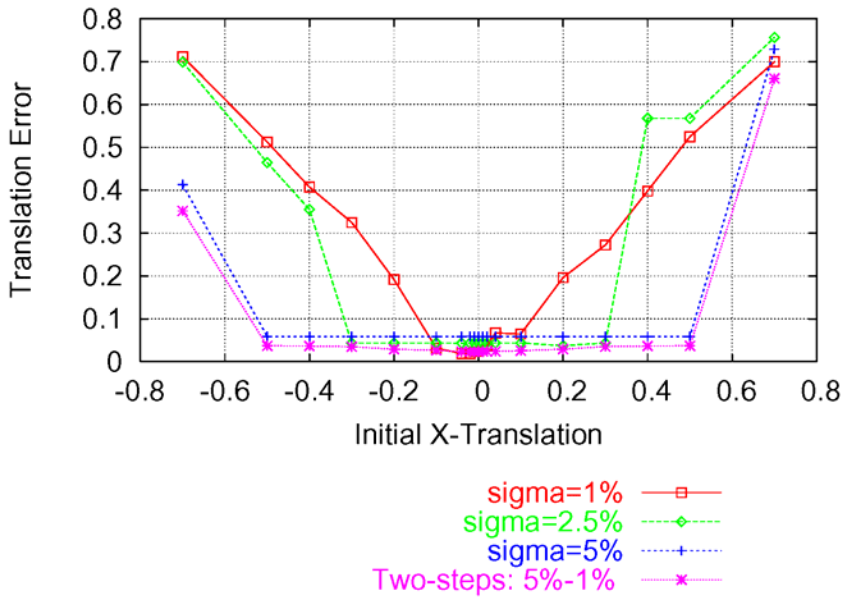


Fig. 6.27. Example of basins of convergence with the two-step algorithm for the Gear dataset compared with basins obtained with the single step runs for different values of σ, and compared with the ICP basin.

6.3.4 Mean Squared Error, Bias and Variance, and Comparison with ICP

In this section we will show: (1) that our method has an optimum parameter σ which balances the accuracy of registration and the region of convergence with respect to the initial starting parameters of the Gaussian Fields algorithm and (2) that for a large range of σ values our algorithm outperforms the ICP algorithm with respect to the Mean Squared Error (MSE) criterion. We start by decomposing the MSE into the Bias and Variance terms. The MSE criterion offers a natural benchmark for quantifying the quality of registration across a range of perturbations. Since we are primarily interested in the sensitivity to initialization (which is the main weakness of current algorithms) we will compute the MSE with respect to a distribution of initial starting points of the algorithm. Furthermore the MSE will give as a way of directly comparing the performance of our method with the ICP algorithm. Let the actual registration parameters for a pair of datasets be Tr^* (the ground truth in our case), and let $Tr^\sigma(x)$ the result of the registration for a given initial guess x and for a given σ. The MSE with respect to the distribution of initial parameters is:

$$MSE = E_x((Tr^\sigma(x) - Tr^*)^2) \tag{6.4}$$

$$= E_x((Tr^\sigma(x) - E_x(Tr^\sigma(x)))^2) + E_x(Tr^\sigma(x) - Tr^*)^2$$

The term $E_x((Tr^\sigma(x) - Tr^*)) = B_\sigma{}^2$ represents square of the Bias of the registration, while $E_x((Tr^\sigma(x) - E_x(Tr^\sigma(x)))^2) = V_\sigma$ is the Variance of the registration parameters with respect to the initial relative positions of the datasets. Hence:

$$MSE = V_\sigma + B_\sigma{}^2 \tag{6.5}$$

We have shown that if σ is large the criterion will become independent of this parameter. The criterion will flatten and tend toward a constant function resulting in a unique result for all initializations. Therefore $\lim_{\sigma\to\infty} V_\sigma = 0$ and $\lim_{\sigma\to\infty} B_\sigma = B$, a constant value. For the lower values of σ the Bias will decrease given that the maximum of the criterion will be closer to the correct position as discussed earlier but the narrowing of the width of the dominant mode and the appearance of local maxima will result in the increase of the variance of the algorithm with respect to initialization. This behavior is illustrated by the plots of Fig. 6.28. In the case of ICP we know that the method is precise locally, hence characterized by a local low bias. Given this it is natural to expect a high variance over a large range of initializations. The question that we will answer experimentally is: how does the two methods compare with the respect to the MSE (Bias+Variance) criterion?

For a quantitative comparison of the performance of the Gaussian Fields method as compared to ICP we use the same six real-world datasets. We illustrate the difference between the two methods by using a uniform distribution of initial translations (along the *x*-axis), in the same way that we obtained the basins of convergence. For each dataset we compute $MSE = V_\sigma + B_\sigma{}^2$ over the different initializations. The results are obtained for several values of σ set as a fraction of the size of the datasets (ranging from 0.5% to 30%). The plots of Figures 6.29 to 6.31 show the variation of MSE criterion with respect to sigma as compared to the MSE of the ICP algorithm. For all datasets, except for the Van dataset, where the difference is still relatively small, there is a point at which the Gaussian method will outperform the ICP algorithm. We clearly see also that a minimum of the MSE is obtained with respect to σ corresponding to an optimal behavior balancing the Bias and Variance constraints. The threshold below which the Gaussian method is better than ICP as well as the optimal σ are inevitably dependant on the datasets, as can be clearly seen from the different plots.

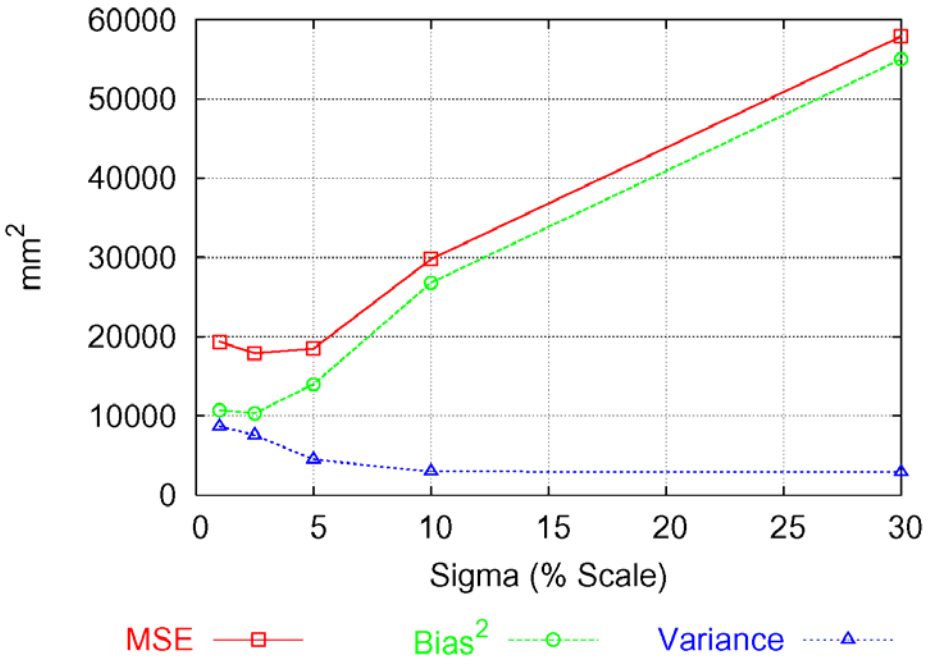
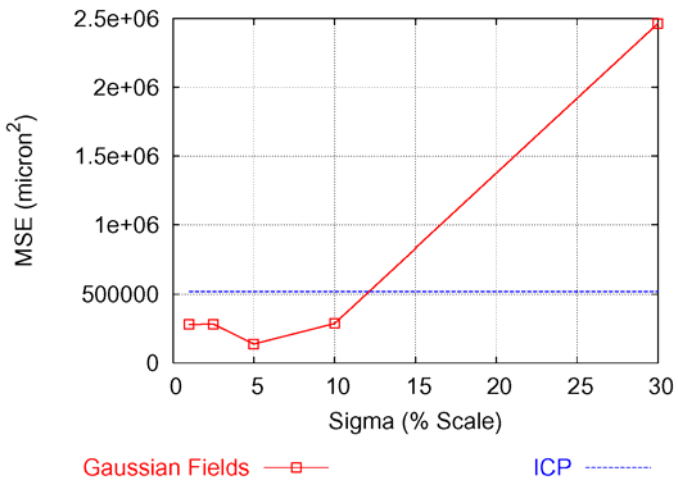


Fig. 6.28. Evolution of the Bias and Variance components of the Mean Squared Error (MSE) computed with respect to a uniform distribution of initial transformations for the Gaussian Fields method, and for increasing values of σ (boat dataset). The Bias will increase as σ increases while the Variance becomes smaller.

Gear



Microchip

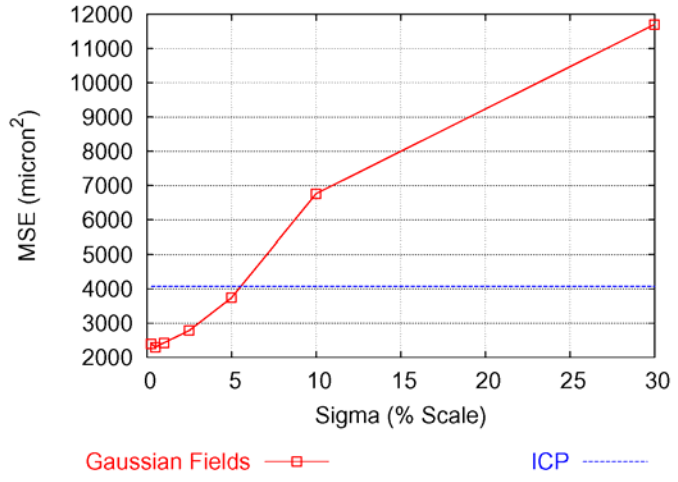
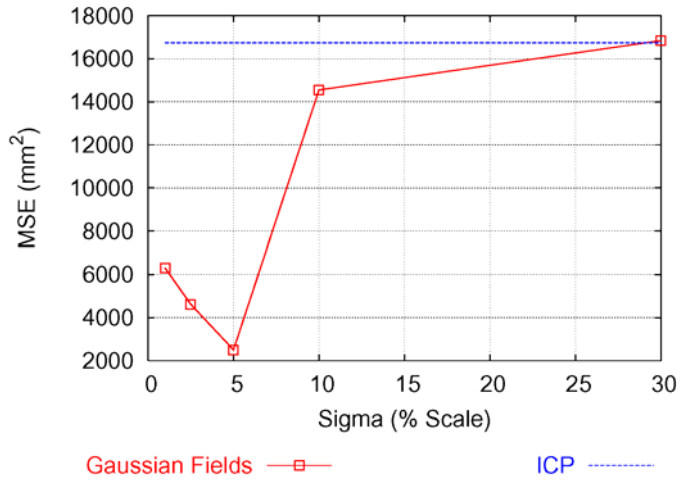


Fig. 6.29. Comparison of the MSE of the Gaussian Fields method with the MSE of ICP (MSE computed with respect to initialization), Gear and Microchip datasets. Two points emerge: (1) there is a threshold for σ below which the Gaussian Fields method will have a lower MSE than the ICP algorithm, and (2) there exists an optimum σ that minimizes the MSE uncertainty criterion.

Parts



Boat

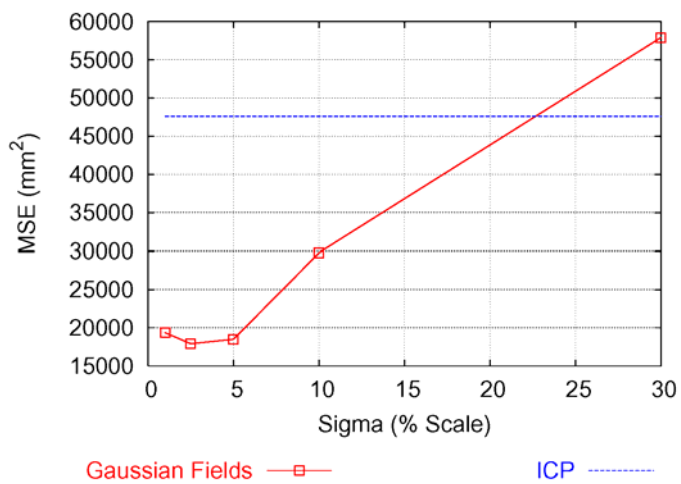
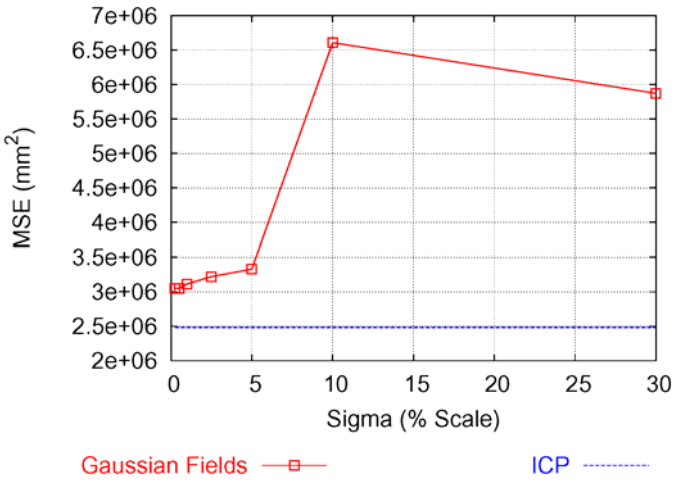


Fig. 6.30. Comparison of the MSE of the Gaussian Fields method with the MSE of ICP, Parts and Boat datasets. Notice that the value of σ below which the MSE of the Gaussian Fields method is lower than ICP will change from one dataset to the other, also the value of σ minimizing the MSE is data which is to be expected.

Van



Building

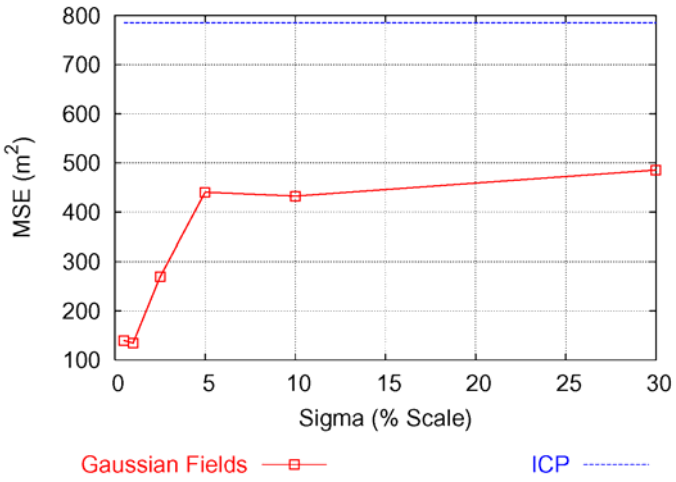


Fig. 6.31. Comparison of the MSE of the Gaussian Fields method with the MSE of ICP, Van and Building datasets. Two extreme cases: (1) for the Van dataset the ICP algorithm has a lower MSE than the Gaussian Fields method although for the optimum σ the difference is relatively small, (2) for the Building dataset the MSE of the Gaussian Fields method is safely below that of ICP for the entire range of σ considered.

6.3.5 Sampling effects

The interest of studying the effects of sub-sampling on the registration algorithm stems from two important issues: (1) reducing the computational burden, and (2) studying the robustness of the algorithm to the reduction in shape information resulting from using less vertices to describe the same object. While the first of these two issues can be addressed by using the FGT framework it is also interesting to see to what level can accuracy be preserved while the number of points is reduced with, keeping in mind the inherent speed gains of combined sub-sampling and FGT techniques. To study the influence of the reduction of resolution we sub-sampled our six datasets in two different ways: (1) uniform sampling, and (2) curvature-based sampling, this last method puts the selected number of points in areas with high shape variations while the first distributes the point in a uniform manner (Fig. 6.32). We start with the relatively low number of 2800 points for each view then sample by two to obtain the next pairs until we reach 400 points.

The two ways of sampling offer both advantages and disadvantages for the method. In the case of uniform sampling we have on one hand a better spatial distribution of the points which will be scattered on a larger area, a fact that is usually helpful for computing rigid transformations, on the other hand the constraint on the distribution of the points results on a coarser description of the object. As for curvature-based sampling it provides for a good visual description of the object, by investing the available point in high-information regions, but the clustering of the points in relatively small areas can result in ambiguities and degenerate solutions. In the plots showing registration error versus the number of points (Figures 6.33 and 6.34) we see no consistent advantage of either sampling scheme. But we notice also that the algorithm handles very high level of sampling and degenerates only when the datasets are described by just few hundred points. This promises further reduction in the computational burden through the use of multi-resolution optimization strategy that initializes at coarser levels.

Uniform Sampling Curvature-based Sampling

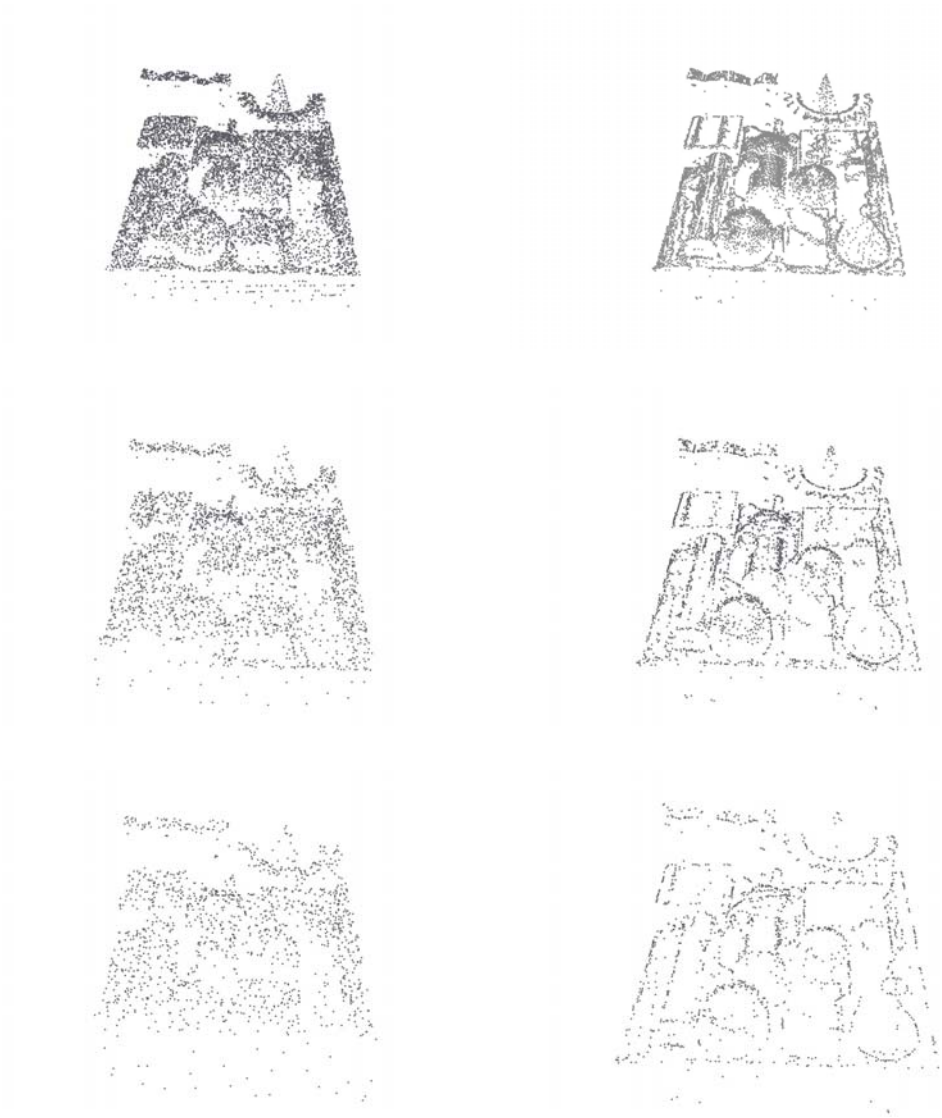


Fig. 6.32. The sub-sampling of the points of Parts dataset using uniform and curvature based sampling.

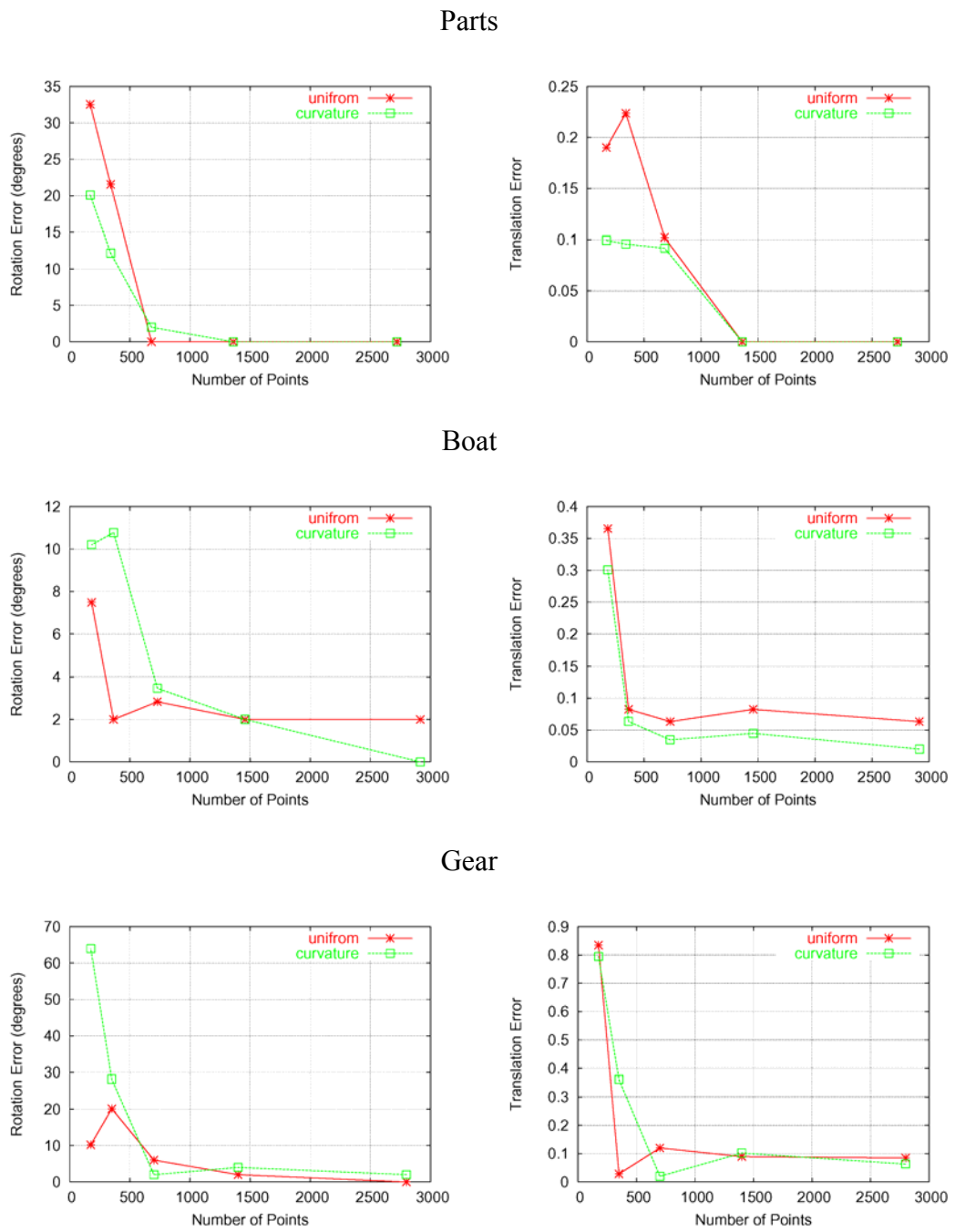
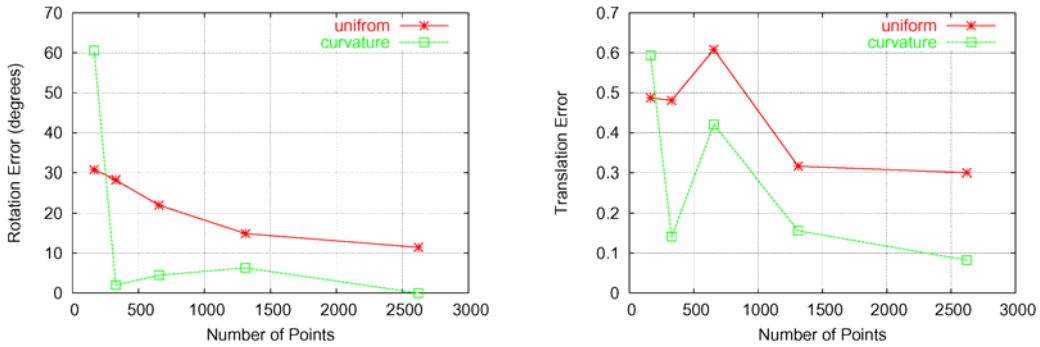
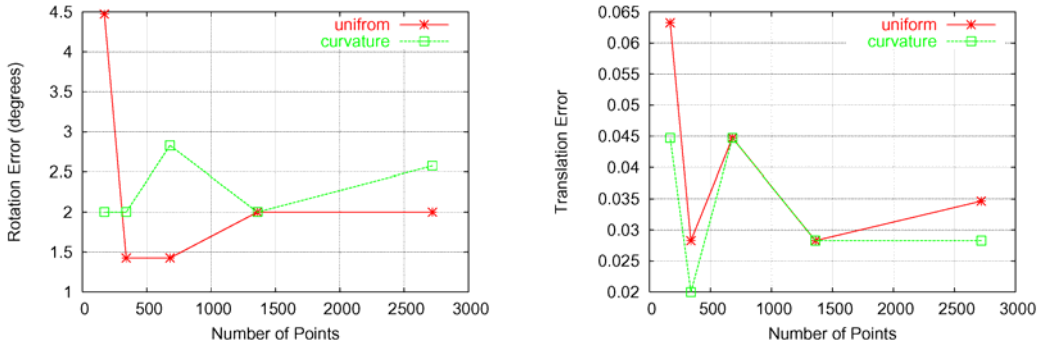


Fig. 6.33. Rotation and translation errors for the Gaussian method as a function of the number of points for uniform and curvature-based sampling.

Microchip



Van



Building

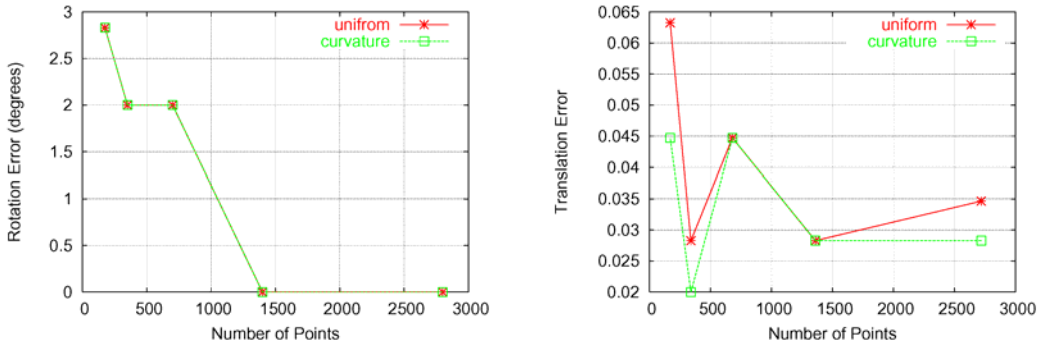


Fig. 6.34. Rotation and translation errors for the Gaussian method as a function of the number of points for uniform and curvature-based sampling. The remaining datasets.

6.4 Multimodal Image Registration and Integration for Modeling and Recognition

So far we have studied the workings of our registration method in the case of 3D point sets registration. The fundamental theory behind our alignment framework also extends, as we have argued in chapter 4, to the case of image registration. Single-modality image registration and tracking in particular should be a good application area of this work. Due to our interest in some specific applications in biometrics and 3D object modeling, we mainly investigated the applicability of the Gaussian Fields method to the task of multimodal 2D image alignment. In addition we have also worked on the integration of the registered range, color, and Infra-Red imagery for object reconstruction.

6.4.1 2D Matching under non-rigid transformations

The main change for the case of image registration from the previous work will come from the need to consider warps which are usually much less constrained than the rigid ones addressed so far. As with any other non-rigid registration criterion we will need in this case to impose enough constraints as to avoid spurious results. The question of local descriptors and attributes will be important here. In fact in the case of non-rigid registration adding more attributes will help constrain the registering transformations, acting as a regularizing factor. In several cases of relatively simple warps such as affine ones it will be useful to employ attributes that are invariant to these transformations. Otherwise we need to make sure that these attributes keep the criterion's nice properties of differentiability and local convexity intact.

Some of the preliminary tests that we conducted to verify the usefulness of the Gaussian Fields algorithm for 2D registration were simple synthetic experiments where we applied an affine transformation to a reference shape, then tried to use the algorithm to align the deformed shape with the original one. In the case illustrated in

Fig. 6.35 we modified slightly the algorithm by using as attributes affine invariant moments. Given that rigid transformations are a special case of affine warps the modification of our algorithm was straightforward. The registration process converged after a few iterations to the accurate alignment. This showed that in principle we can apply the method to non-rigid transforms.

6.4.2 IR-Color alignment

Our study of the Gaussian Fields registration method for IR-Color images registration focused mostly on multimodal face recognition applications. The fusion of IR and color imagery for face recognition was shown to increase significantly the recognition rates of currently deployed systems [Heo04]. An improvement that is due to the important amount of independent information available in these modalities. In this context, shortcomings of color imagery such as sensitivity to illumination changes can be compensated for by fusion with IR data. Furthermore, while color imagery provides information about the surface of the face, IR images show the blood vessel and heat emission patterns unique to every person. Another important application of IR-color fused images is the location of the eyes for recognition purposes.

Our purpose was to examine the behavior of the Gaussian criterion for this particularly challenging task. The difficulty of this task stems from the wide difference between the salient patterns and features in the two images. First, the two images need to be brought to a common space by a feature extraction step. Given the large differences between the two modalities, we will have a much larger disparity of the values of local feature descriptors when we go from one image to another than in the single modality case. Therefore the algorithm will rely mostly on the global visual similarity to achieve the alignment.

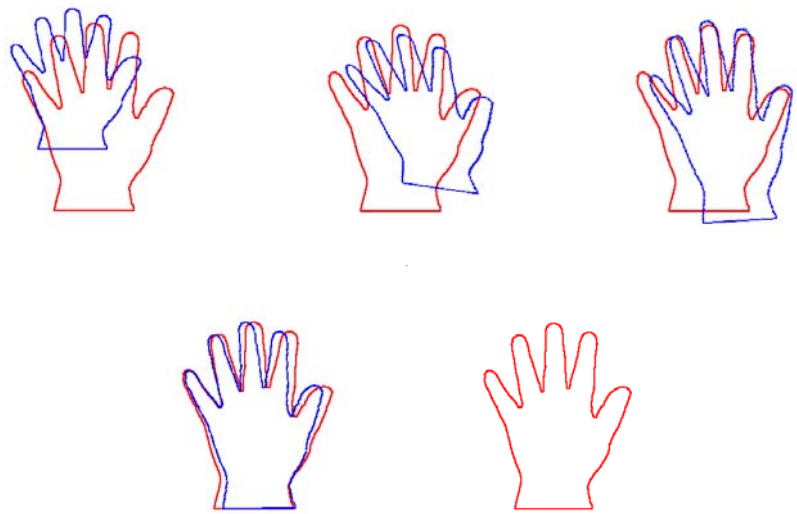


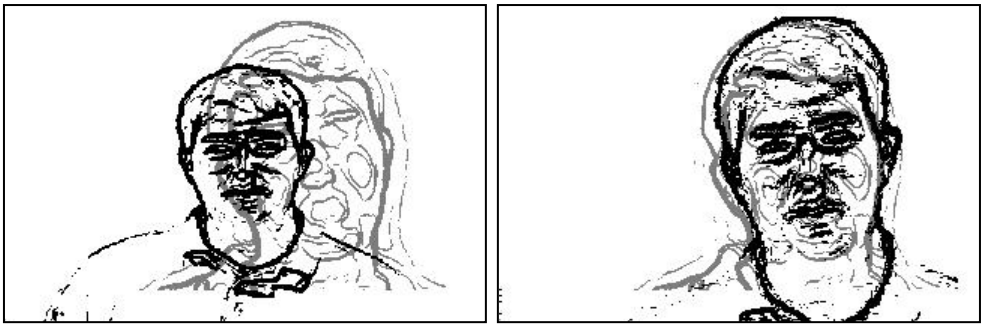
Fig. 6.35. Several steps of the iterative alignment of hand contours which are related by an affine transformation using the Gaussian Fields method.

The registration algorithm was applied to pairs of IR-Color images such as the ones shown in Fig. 6.36. Edge maps were extracted using a generic canny edge detector. Given that for non-rigid transformations initialization is an important issue, the parameters and relative position of the imagers can be set so that we have a significant overlap between the faces in the image pair. In real application this will not be a hindering constraint, since we can have a fairly controlled setup for identification purposes in places such as ports of entry. To further improve the convergence of the method we start by employing a transformation constrained to a similarity (rigid + scale). After the initial registration further refinement is obtained by using the projective 8-parameter transformation:

$$Tr\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} p_1 + p_2 x + p_3 y + p_7 x^2 + p_8 xy \\ p_4 + p_5 x + p_6 y + p_7 xy + p_8 y^2 \end{bmatrix} \tag{6.6}$$

(a) (b)

(c)



(d) (e)

Fig.6.36. Thermal (a) and Color (b) images of faces are registered as shown in the composite image in (c) by maximizing the Gaussian registration criterion to align the edges shown in the unregistered (d) and registered (e) cases.

The method allowed for the registration of the widely dissimilar images by maximizing the overlap between the edge maps and by implicitly matching the locally computed shape descriptors. The next step in this particular task is to study several more sophisticated feature extraction techniques that can provide better matches between the images. It will be also useful to investigate the use of more elaborate motion models both parametric and dense motion fields which are expected to provide more accurate results than those obtained so far.

6.4.3 Integrating range and color imagery

We turn now to the alignment of range maps with color images, a task needed in applications such as virtualized reality where photo-realistic models are required. The commonly adopted approach to the registration of color and range images, in the context of 3D modeling, is the formulation of the problem in terms of camera pose estimation [El-Hakim98][Sequiera99][Stamos00]. Assuming that the system is calibrated and that the relative position of the 3D scanner and of the color camera is fixed during the imaging process, the registration will need to be done only once. In some other applications where accurate calibration is not available we can adopt an approach based on image warping to map the color image onto the range map We also think that it is straightforward to extend the Gaussian Fields registration method to this task by employing edge maps extracted from both color and range imagery. The optimization process will have to recover the 3D to 2D projective transformation, which is in fact much more constrained than the non-rigid warps that we discussed in the case of IR-color alignment. Going back to the standard pose estimation approach we need first to establish point correspondences between the range and color images. Once the image-to-image matches found a set of 3D-2D matches can be inferred using the scanner's calibration parameters.

Using the now commonplace 3D editing tools these matches can be established directly between the 3D model reconstructed from the range maps and the color images. The list of correspondences is used to recover the 3×4 camera projection matrix P mapping the world coordinate system to the retinal plan of the color camera as follows:

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{6.7}$$

This matrix is computed using a two-stage approach, starting with a linear over-constrained system. The solution of this system will be a rough approximation of the mapping that will be refined by non-linear methods widely studied in camera calibration literature [Tasi87][Faugeras93][Hartley]. After the recovery of the model-to-image mapping, texture maps can be generated. These are images used by 3D rendering engines for overlay on top of polygon models [Weinhaus97]. For each triangle in the 3D mesh, the associated texture is determined by finding the texture-coordinates of each vertex, which are the 2D coordinates of the projected vertices in the texture image. In Fig. 6.37 we show an example of a model that we reconstructed from range and color imagery. Some areas of the reconstructed scene that are not visible from the camera's point of view are determined using z-buffering techniques, associating a depth buffer to each texture element so that when several 3D surface samples are mapped into that element only the closest one will be considered as non-occluded.

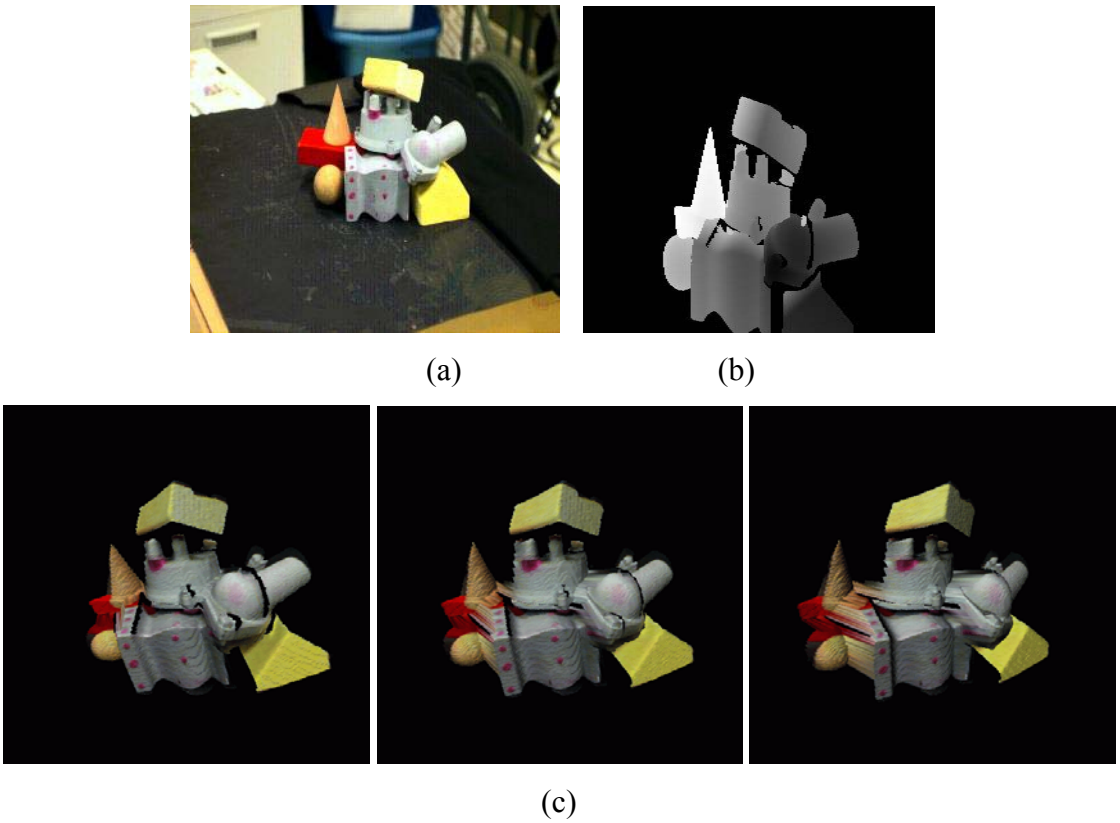


Fig. 6.37. The input of a photo-realistic modeling system: (a) color image, (b) range map. The reconstructed geometry with the registered texture map is rendered from different viewpoints in (c).

6.4.4 Multimodal 3D face reconstruction

In another illustration of multimodal object modeling we reconstructed 3D faces from image sequences, Infra-Red imagery, and range data. Face modeling has several applications in the fields of computer vision, biometrics, and graphics, where there is a need for 3D face recognition and facial animation [Pighin02][Liu99]. A preliminary step will be to register the 2D multimodal imagery using the method described earlier. Following this a generic face model (shown in Fig. 6.38(c)) was reconstructed from dense range scans of the mannequin's head shown in Fig. 6.38(a). Strategically located landmark points were chosen in the range image (Fig. 6.38(b)). An image sequence of the moving face of interest was captured, and the feature-points were tracked throughout the sequence (Fig. 6.38(d)). To perform the tracking, pre-defined facial feature templates could be used and matched with the images. In our experiment we picked the features manually for the purpose of illustrating the overall approach. The tracked points will be reconstructed using SFM. It is clear that the few selected features will result in a very coarse structure of the face, as can be seen in Fig. 6.38(f). We will fill in the structure with some generic face shape information provided by the dense range data. For this purpose the face model is warped using 3D thin-plate splines [Toga99] interpolation to fit the SFM-reconstructed points. With this approach, a far more realistic 3D model of the face is obtained. Color texture is mapped into this model using the color images as shown in Fig. 6.38(g).The registered thermal imagery (6.38(e)) is also overlaid in a similar way (Fig 6.38(h)). From these results, we can see that the method was able to recover a significant level of detail of the modeled object. In particular, the use of thin-plate splines mapping allowed for a good approximation of both global and local shape variations. The technique works well because the generic model embeds already an important amount of face shape information. Hence, the combination of dense range data and SFM resulted in a fast video-IR-Range face reconstruction pipeline. Recently several publications adopting a warping approach to face modeling appeared including the work of Chowdhury et al [Chowdhury02], Hwang and Lee [Hwang02], and Zhang and Cohen [Zhang02].

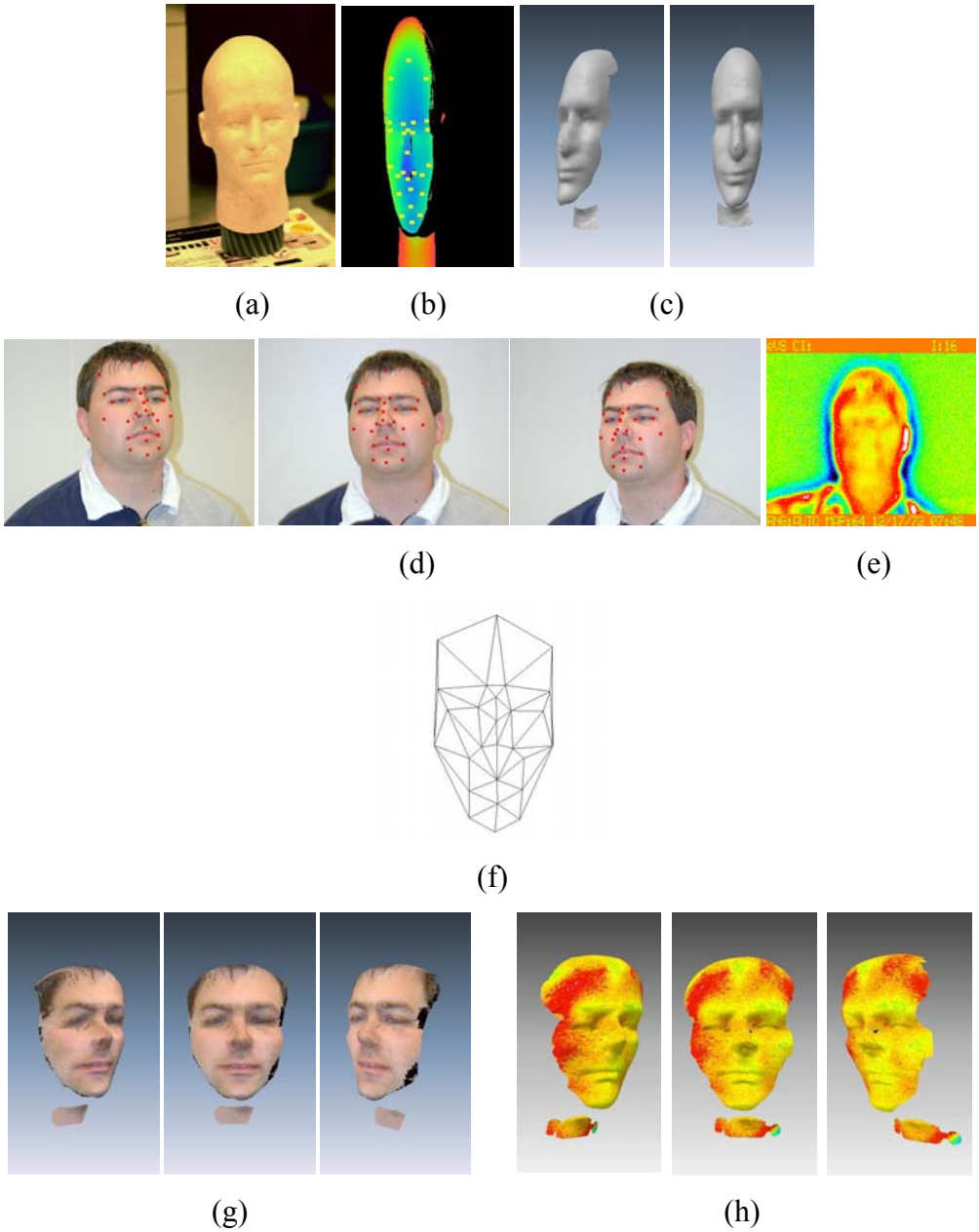


Fig. 6.38. Multi-modal Face reconstruction from range, color sequences and thermal images. The input imagery used will be: range imagery (b) of a generic face (mannequin head (a)), from which a 3D model is reconstructed (c); color image sequences from which features are extracted and matched (d); and thermal images. A skeleton of the face is reconstructed using SFM (f), then a more accurate model is obtained by aligning the generic face model with it. The resulting reconstruction is texture mapped with registered color (g) and thermal images.

7 CONCLUSIONS AND FUTURE WORK

In summary, this work is an attempt at further automating the registration task, which is the main bottleneck in modeling and fusion pipelines. The need for operator intervention is very clear in commercial 3D modeling and editing software platforms that are increasingly popular for different applications. The scanning systems and their associated interfaces require point-picking by the operators, use special fiducial markers to find the correspondences, or rely on the controlled motion of either the scanning platform or the scanner itself to register the different views. The ICP algorithm is the basis for close-range accurate registration, a task for which it is well suited. In this dissertation we have stressed the fundamental limits of this method pointing to the non-differentiable nature of the criterion that required a specialized optimization heuristic. The question that was at the core of our efforts is: can we design a new criterion which balances the two conflicting goals of registration accuracy and large range of convergence? We believe that the Gaussian Fields method developed in this dissertation offers a good answer to this question.

7.1 Dissertation Key Points

The key points forming the backbone of this research are the following

A robust feature saliency descriptor for general point-sets

We developed the saliency measure first in the context of camera motion recovery but soon it proved a good feature descriptor for noisy point-sets. For this we relied on the

efficient framework of tensor voting to robustly infer the nature of point-features. Most other local descriptors use differential surface information, and invariant moment invariants that can be computed for point-sets are very computationally expensive. Furthermore, higher order invariants are very sensitive to noise. The local saliency measure strikes a compromise between good discrimination and robustness. Structure information is embedded in an efficient way without any need for explicit extraction. Additionally the descriptor is analytically expressed in terms of the point-sets which will be very helpful later for criterion optimization.

The Gaussian Fields criterion for registration

The Gaussian criterion constitutes the most important contribution of this research. It was developed in our effort to overcome several shortcomings of ICP-based algorithms. The method has the following advantages over current techniques:

1. The criterion stems from a clear and rigorous formulation of the registration task as a search for the maximum overlap between the datasets. We extend the overlap to the multi-dimensional space of both position and local attributes. This formulation allows for the easy incorporation of intensity and shape information in the registration framework. Our formulation derives from the use simple combinatorial matching principles, along with mollification and relaxation techniques.

2. The Gaussian criterion gives a straightforward understanding of the effect of outliers as well and provides a mechanism for their control.

3. An intuitive physical interpretation of the method can be provided by analogy to particle physics where points are subject to some exponentially decaying force fields.

4. The criterion has the nice properties of continuous differentiability and can be also shown to be locally convex in the neighborhood of the registered position. Such properties allow for the use of the well developed gradient based optimization techniques.

5. In addition to easy optimization the criterion allows for the increase of the range of convergence by tuning the Gaussian smoothing parameter to have a two-stage global scheme. Hence we have with the same framework both initialization and refinement without need for explicit point-feature matching.

The Fast Gauss Transform

This powerful numerical technique saves our method from what was going to be its main drawback which is the nearly $O(N^2)$ computational complexity. By employing clever analytic 'tricks' it clusters several data points substituting them with a small number of field sources and targets, thus achieving a remarkable linear $O(N)$ complexity. Computational burden was also in our minds when we chose the local descriptors, ruling out some expensive high-order moments.

Multimodal image registration

By extracting local feature maps we are able to employ the method for single and multi-sensor image registration. The application of Gaussian framework needs no major modifications to extend to this important task.

Applications

Our targeted applications are virtualized reality for simulation reverse engineering and design verification, remote inspection and hazardous waste management, and

biometrics. In the dissertation we showed examples of 3D multimodal reconstruction from color, IR, and range imagery.

7.2 New Questions, Other applications, and Future Research

Of course this research while claiming an important contribution to the state of the art in point-sets registration does not pretend to 'solve' the problem in any definitive way. Many areas in the 3D registration framework could be improved, especially in the optimization part. Gaussian convolution was used in other applications as a way of regularizing noisy data and of stabilizing energy minimizing methods. This comes usually at the price of a loss to accuracy, and while we showed here that we can empirically design a multi-step global scheme to strike a balance between the range of convergence and accuracy we cannot actually prove that global convergence is a certainty.

This work opens several promising research opportunities. We are particularly interested in the application of the Gaussian fields approach to non-rigid alignment tasks. In several applications, such as in medical imaging, addressing this problem is of great importance. Our work on feature-based multi-modal image registration offered a first step in this direction but so far the transformations used were rather simple. More complex mappings will need to be considered in many other cases. The use of the saliency descriptor will be helpful in this task given its differentiability properties which are more important in this case then not invariance, which is not possible in case of complex transformations. We think that our framework is well suited to video frame-to-frame registration and to in video sequences tracking. For these applications the Gaussian Fields method offers a good technique that uses both shape and color information to find the correspondences. The formulation of our criterion as a multidimensional matching energy function extends beyond our

immediate image processing context to general pattern recognition. We think that we are far from exhausting the possibilities that this paradigm offers.

BIBLIOGRAPHY

BIBLIOGRAPHY

[Andersen01] P. R. Andersen, M. Nielsen, “Non-rigid registration by geometry-constrained diffusion,” Medical Image Analysis, Vol. 5, pp. 81-88, 2001.

[Azarbayejani95] A. Azarbayejani and A. P. Pentland, “Recursive Estimation of Motion, Structure, and Focal Length,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 6, pp. 562-574, 1995.

[Bjcsy89] R. Bajcsy, S. Kovacic, “Multiresolution elastic matching,” Computer Vision, Graphics and Image Processing, Vol. 46, pp. 1-21, 1989.

[Baker01] S. Baker, T. Sim, and T. Kanade. A Fundamental Theorem of Stereo? Tech. Report CMU-RI-TR-01-07, Robotics Institute, Carnegie Mellon University, June, 2001.

[Bargevin96] R. Bergevin, M. Soucy, H. Gagnon, and D. Laurendeau, "Towards a General Multi-view Registration Technique," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 18, pp. 540-547, May 1996.

[Baxter02] B. J. C. Baxter and G. Roussos, “A new error estimate of the fast Gauss transform,” SIAM Journal on Scientific Computing, Vol. 24, No. 1, pp. 257–259, 2002.

[Besl88] P.J. Besl. Active Optical Range Imaging Sensors, Springer-Verlag, New York, 1988.

[Besl92] 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, No. 2, pp. 239-256, 1992.

[Blais95] G. Blais and M. D. Levine, “Registering Multiple Range Data to Create 3D Computer Objects,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 17, No. 8, pp. 820-824, 1995.

[Boughorbel04a] F. Boughorbel, B. Abidi, and M. Abidi, "Registration of Infra-Red and Color Images for Multimodal Face Recognition," The Biometric Consortium Conference (BC2004), Crystal City, VA, September 20-22, 2004.

[Boughorbel04b] F. Boughorbel, A. Koschan, B. Abidi, and M. Abidi, "Gaussian Fields: a New Criterion for 3D Rigid Registration," Pattern Recognition, Vol. 37, No. 7, pp. 1567-1571, July 2004.

[Boughorbel04c] F. Boughorbel, A. Koschan, B. Abidi, and M. Abidi, "Gaussian Energy Functions for Registration without Correspondences," 17th International Conference on Pattern Recognition (ICPR), Cambridge, UK, August 2004.

[Boughorbel03] F. Boughorbel, P. Crilly, A. Koschan, and M. Abidi, "Estimating 3D camera motion without correspondences using a search for the best structure," Pattern Recognition Letters, Vol. 24, No. 1-3, pp. 327-337, January 2003.

[Bookstein89] F. L. Bookstein, "Principal Warps: Thin-Plate Splines and the Decompostion of Deformations," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 11, No. 6, pp. 567-585, 1989.

[Borgefors88] G. Borgefors, "Hierarchical chamfer matching: A parametric edge matching algorithm," IEEE Trans. PAMI, Vol. 10, pp. 849-865, Nov. 1988.

[Brujic96] D. Brujic and M. Ristic, "Analysis of Free-form Surface Registration," Proceedings of the IEEE International Conference on Image Processing, Vol. 2, pp. 393-396, Lausanne, Switzerland, 1996.

[Campbell01] R. Campbell and P. Flynn, "A Survey of Free-form Object Representation and Recognition Techniques," Computer Vision and Image Understanding, Vol. 81, No. 2, pp. 166-210, 2001.

[Charpiat03] G. Charpiat, O. Faugeras, and R. Keriven, "Shape Metrics, Warping and Statistics," in Proc. International Conference on Image Processing, Vol. 2, pp. 627-630, Barcelona, 2003.

[Chen92] Y. Chen and G Medioni, "Object Modeling by Registration of Multiple Range Images," Image and Vision Computing, Vol. 10, No. 3, pp. 145-155, 1992.

[Chen94] Q. Chen, M. Defrise, and F. Deconinck, "Symmetric phase-only matched filtering of Fourier–Mellin transform for image registration and recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 16, pp. 1156-1168, 1994.

[Chen99] C. Chen, Y. Hung, and J. Cheng, "RANSAC-Based DARCES: A New Approach to Fast Automatic Registration of Partially Overlapping Range Images," IEEE PAMI, Vol. 21, No. 11, 1999.

[Chiba98] N. Chiba and T. Kanade, “A Tracker for Broken and Closely Spaced Lines,” CMU-CS-97-182 1997, or ISPRS paper vol. XXXII, Part 5, pp. 676-683, Hakodate, 1998.

[Chou97] G. Chou and S. Teller, “Multi-Image Correspondence Using Geometric and Structural Constraints,” in Proc. 1997 Image Understanding Workshop, pp. 869-874, May, 1997.

[Chowdhury02] A. R. Chowdhury, R. Chellappa, S. Krishnamurhty, and T. Vo, “3D Face Reconstruction from Video using a Generic Model,” Proceedings of the International Conference on Multimedia, Vol. 2, pp. 449-452, Switzerland, 2002.

[Chua96] C. Chua and R. Jarvis, "Point Signatures: a New Representation for 3D Object Recognition," International Journal of Computer Vision, Vol. 25, No. 1, pp. 63-85, 1996.

[Costeira95] J. Costeira and T. Kanade, “A Multi-body Factorization Method for Motion Analysis,” in Proceedings of ICCV'95, pp. 1071-1076, 1995.

[Dalley02] G. Dalley and P. Flynn, “Pair-Wise Range Image Registration: A Study in Outlier Classification,” Computer Vision and Image Understanding, Vol. 87, No. 1-3, pp. 104-115, July 2002.

[De Castro87] E. De Castro, C. Morandi, “Registration of translated and rotated images using finite Fourier transform,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 9, pp. 700-703, 1987.

[Dellaert00] F. Dellaert, S. M. Seitz, C. E. Thorpe, and S. Thrun, “Structure from Motion without Correspondence,” In IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Press, June 2000.

[Devroy85] L. Devroye, F. Machell, “Data structures in kernel density estimation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 7, No. 3, pp. 360–366, 1985.

[Dorai97] C. Dorai, J. Weng, and A. K. Jain, “Optimal Registration of Object Views using Range Data,” IEEE Transactions on Pattern Analysis Machine Intelligence, Vol. 19, No. 10, pp. 1131-1138, 1997.

[Dyer01] C. R. Dyer, “Volumetric Scene Reconstruction from Multiple Views,” in Foundations of Image Analysis, L. S. Davis, ed., Kluwer, Boston, 2001.

[Eggert96] D. Eggert, A. W. Fitzgibbon, R. B. Fisher, "Simultaneous Registration of Multiple Range Views for use in Reverse Engineering," in Proceedings of International Conference on Pattern Recognition, pp 243-247, Vienna, Aug. 1996.

[Elgammal03] A. Elgammal, R. Duraiswami, and L. Davis, "Efficient Kernel Density Estimation using the Fast Gauss Transform with Applications to Color Modeling and Tracking," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No. 11, pp. 1499-1504, 2003.

[El-Hakim98] S. F. El-Hakim, C. Brenner, G. Roth, "A Multi-sensor Approach to Creating Accurate Virtual Environments," ISPRS Journal for Photogrammetry and Remote Sensing, 53(6), December 1998.

[Faugeras86] O. Faugeras and M. Hebert "The Representation, Recognition, and Locating of 3-D Objects," International Journal of Robotic Research, Vol. 5, No. 3, 1986.

[Faugeras90] O. D Faugeras, S. Maybank, "Motion from point matches: Multiplicity of Solutions," International Journal of Computer Vision, Vol. 4, pp. 225-246 1990.

[Faugeras92] O. D Faugeras, S. Maybank, "What can be Seen in Three Dimensions with an Uncalibrated Stereo Rig," in Proceedings of European Conference on Computer Vision, LNCS 588, pp. 321-334, Springer-Verlag, 1992.

[Faugeras93] O. Faugeras. Three-Dimensional Computer Vision. The MIT Press, Cambridge Massachusetts, 2001.

[Faugeras98] O. Faugeras and R. Keriven, "Complete Dense Stereovision using Level Set Methods," In Proc. EECV, pp. 379-393, 1998.

[Faugeras01] O. Faugeras, Q. Luong, and T. Papadopoulo, The Geometry of Multiple Images, The MIT Press, Cambridge Massachusetts, 1993.

[Feder88] T. Feder, D. Greene, "Optimal algorithms for approximate clustering," In Proc. 20th ACM Symposium on the Theory of computing, pp. 434–444, Chicago, Illinois, 1988.

[Fitzgibbon03] A. W. Fitzgibbon, "Robust registration of 2D and 3D Point Sets," Image and Vision Computing, Vol. 21, pp. 1145-1153, 2003.

[Fruh01] C. Früh, A. Zakhor, “3D Model Generation for Cities using Aerial Photographs and Ground Level Laser Scans,” in Proceedings of Computer Vision and Pattern Recognition, pp. 31-38, 2001.

[Fukunaga89] K. Fukunaga, R. R. Hayes, “The Reduced Parzen Classifier,” IEEE Trans. Pattern Anal. Mach. Intell., Vol. 11, No. 4, pp. 423-425, 1989.

[Fusiello00] A. Fusiello, “Uncalibrated Euclidean reconstruction: A review,” Image and Vision Computing, Vol. 18, No. 6-7, pp. 555-563, May 2000.

[Geman92] S. Geman, E. Bienenstock, Rene Doursat, “Neural Networks and the Bias/Variance Dilemma,” Neural Computation, Vol. 4, pp. 1-58, 1992.

Goldberg02] B. Goldberg, M. Maimone, and L. Matthies, “Stereo Vision and Rover Navigation Software for Planetary Exploration,” IEEE Aerospace Conference Proceedings, March 2002, Volume 5, Big Sky, Montana, USA, pp. 2025-2036.

[Gonzalez85] T. Gonzalez, “Clustering to minimize the maximum intercluster distance,” Theoretical Computer Science, Vol. 38, pp. 293–306, 1985.

[Greengard87] L. Greengard, V. Rokhlin. “A fast algorithm for particle simulations,” J. Comput. Phys., Vol. 73, No. 2, pp. 325-348, 1987.

[Greengard88] L. Greengard, The Rapid Evaluation of Potential Fields in Particle Systems, Cambridge, Massachusetts, MIT Press, 1988.

[Greengard91] L. Greengard and J. Strain, “The Fast Gauss Transform,” SIAM J. Scientific Computing, Vol. 12, pp. 79-94, 1991.

[Gumerov03] N. A. Gumerov, R. Duraiswami, E. A. Borovikov, “Data structures, optimal choice of parameters, and complexity results for generalized multilevel fast multipole methods in d-Dimensions,” Technical Report UMIACS-TR-2003-28, UMIACS, University of Maryland, College Park, 2003.

[Guy97] G. Guy and G. Medioni, "Inference of Surfaces, 3-D Curves and Junctions from Sparse, Noisy 3-D Data,” IEEE PAMI, Vol. 19, No. 11, pp. 1265-1277, 1997.

[Han99] M. Han and T. Kanade, The Factorization Method with Linear Motions, tech. report CMU-RI-TR-99-23, Robotics Institute, Carnegie Mellon University, Oct. 1999.

[Hartley00] R. Hartley and A. Zisserman. Multiple View Geometry in Computer Vision, Cambridge University Press, June, 2000.

[Hastie01] T. Hastie, R. Tibshirani, J. Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer, New York, 2001.

[Hebert95] M. Hebert, K. Ikeuchi, and H. Delingette, "A Spherical Representation for Recognition of Free-form Surfaces," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 17, No. 7, pp. 681-690, 1995.

[Heo04] J. Heo, S. Kong, B. Abidi, and M. Abidi, "Fusion of Visual and Thermal Signatures with Eyeglass Removal for Robust Face Recognition," IEEE Workshop on Object Tracking and Classification Beyond the Visible Spectrum in conjunction with CVPR 2004, pp.94-99, Washington D.C., July 2004.

[Hermosillo02] G. Hermosillo, C. Chefd'hotel, and O. Faugeras, "Variational methods for multimodal image matching," International Journal of Computer Vision, Vol. 50, No. 3, pp. 329-343, 2002.

[Higgins81] H. C. Longuet-Higgins, "A Computer Program for Reconstructing a Scene from Two Projections," Nature, Vol. 293, pp. 133-135, September 1981.

[Higuchi95] K. Higuchi, M. Hebert, and K. Ikeuchi, "Building 3-D Models from Unregistered Range Images," Graphical Models and Image Processing, Vol. 57, No. 4, pp. 315-333, 1995.

[Horn87] B. K. P. Horn, "Closed-form Solution of Absolute Orientation using Unit Quaternions," Journal of the Optical Society of America A, Vol. 4, No. 4, pp. 629-642, April 1987.

[Horn90] B.K.P. Horn, "Relative Orientation," International Journal of Computer Vision, Vol. 4, No. 1, pp. 59-78, January, 1990.

[Huber03] D. Huber and M. Hebert, "Fully automatic registration of multiple 3D data set," Image and Vision Computing, Vol. 21, No. 7, pp. 637-650, July 2003.

[Huber96] P. Huber, Robust Statistical Procedures, SIAM, 2nd Edition, 1996.

[Huttenlocher93] D. P. Huttenlocher, G. A. Klanderman, W. J. Rucklidge, "Comparing images using the Hausdorff distance," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 15, pp. 850-863, 1993.

[Hwang02] B. W. Hwang, and S. W. Lee, "Face Reconstruction with a Morphable Face Model," Proceedings of the 16th International Conference on Pattern Recognition (ICPR 2002), Vol. 2, pp. 366-369, August 2002.

[Jebara99] T. Jebara, A. Azarbayejani and A. Pentland, "3D Structure from 2D Motion," in IEEE Signal Processing Magazine, Vol. 16, No. 3, pp. 66-84, 1999.

[Jeon94] B. Jeon and D. A. Landgrebe, "Fast Parzen density estimation using clustering-based branch and bound," IEEE Trans. Pattern Anal. Mach. Intell., Vol. 16, No. 9, pp. 950–954, 1994.

[Jia03] J. Jia, C.-K. Tang, "Image repairing: robust image synthesis by adaptive ND tensor voting," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. 643-650, June 2003.

[Jia04] J. Jia, C.-K. Tang, "Inference of segmented color and texture description by tensor voting," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 26 , No. 6 , pp. 771 – 786, June 2004.

[Johnson97] A. E. Johnson and M. Hebert, "Using Spin Images for Efficient Multiple Model Recognition in Cluttered 3-D Scenes," IEEE PAMI, Vol. 21, No. 5, pp. 433-449, 1999.

[Irani98] M. Irani and P. Anandan, "Robust multi-sensor image alignment," Proceedings of the sixth International Conference on Computer Vision, 959-965, 1998.

[Kanade97] T. Kanade and D. D. Morris, "Factorization Methods for Structure From Motion," Philosophical Transactions of the Royal Society of London, Series A, 356(1740), pp.1153-1173, 1998.

[Kanatani00] K. Kanatani, D. and Morris, "Gauges and Gauge Transformations in 3-D Reconstruction from a Sequence of Images," Proceedings of the Fourth Asian Conference on Computer Vision (ACCV2000, Taiwan, pp. 1046-1051, 8-11 January 1999.

[Kapoutsis99] C. A. Kapoutsis, C. P. Vavoulidis, and I. Pitas, "Morphological iterative closest point algorithm," IEEE Transactions on Image Processing, Vol. 8 , No. 11 , pp.1644-1646, November 1999.

[Kirkpatrick83] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by Simulated Annealing," Science, Vol. 220, No. 4598 , pp.671-680, May 1983.

[Kornprobst00a] P. Kornprobst and G. Medioni, "Tracking Segmented Objects using Tensor Voting," in Proceedings of CVPR00, pp. 118-125, 2000.

[Liu99] Z. Liu, Z. Zhang, C. Jacobs, and M. Cohen, “Rapid Modeling of Animated Faces from Video,” Technical Report, Microsoft Research 99-21, April 1999.

[Liu02] J. Liu, B.C. Vemuri, and J.L. Marroquin, “Local frequency representations for robust multimodal image registration,” IEEE Transactions on Medical Imaging, Vol. 21, pp. 462-469, 2002.

[Lucchese02] L. Lucchese, G. Doretto, and G. M. Cortelazzo, “A frequency domain technique for range data registration,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 11, pp.1468-1484, November 2002.

[Maimone98] M. Maimone, L. Matthies, J. Osborn, E. Rollins, J. Teza, and S. Thayer. "A photorealistic 3-D mapping system for extreme nuclear environments: Chornobyl," Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, October, 1998.

[Maes97] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, P. Suetens, “Multimodality image registration by maximization of mutual information,” IEEE Transactions on Medical Imaging, Vol. 16, pp. 187-198, 1997.

[Maintz98] J. B. A. Maintz and M. A. Viergever, “A survey of Medical Image Registration Techniques,” Medical Image Analysis, 2(1), pp. 1-36, 1998.

[Marr82] D. Marr, Vision, Freeman, San Francisco, 1982.

[Masuda95] T. Masuda and N Yokoya, "A Robust Method for Registration and Segmentation of Multiple Range Images,” Computer Vision and Image Understanding, Vol. 61, No. 3, pp. 295-307, 1995.

[Medioni00] G. Medioni, M. S. Lee, C. K. Tang, A Computational Framework for Segmentation and Grouping, Elsevier, Amsterdam, 2000.

[Meinguet84] J. Meinguet, "Surface Spline Interpolation: Basic Theory and Computational Aspects,” in Approximation Theory and Spline Functions, (Editors S. P. Singh et al.), pp.127-142, Dordrecht, The Netherlands: Reidel, 1984.

[Morris01] D. D. Morris, K. Kanatani and T. Kanade, “Gauge fixing for accurate 3D estimation,” Proc. IEEE Conf. Comput. Vision Pattern Recognition, Vol. 2, pp. 343-350, Kauai, Hawaii, December 2001.

[Morris98] D. D. Morris and T. Kanade, “A Unified Factorization Algorithm for Points, Line Segments and Planes with Uncertainty Models,” in Proceedings of Sixth

IEEE International Conference on Computer Vision (ICCV'98), pp. 696-702, January, 1998.

[Mundy92] J. Mundy and A. Zisserman, Geometric Invariance in Computer Vision, MIT Press, 1992.

[Murio93] D. A. Murio, The Mollification Method and the Numerical Solution of Ill-Posed Problems, J. Wiley & Sons, New York, 1993.

[Neugebauer97] P. J. Neugebauer, "Geometrical Cloning of 3D Objects via Simultaneous Registration of Multiple Range Images," in Proceedings of Int. Conf. on Shape Modeling and Application, pp. 130-139, Aizu-Wakamatsu, Japan, March 1997.

[Nicolescu03] M. Nicolescu, G. Medioni, "Layered 4D representation and voting for grouping from motion," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25 , No. 4, pp. 492-501, April 2003.

[Park04] S. Y. Park and M. Subbarao, "Automatic 3D Reconstruction based on Novel Pose Estimation and Integration Techniques," Image and Vision Computing, Vol. 22, No. 8, pp. 623-635, August 2004.

[Pennec97] X. Pennec and J.P. Thirion, "A Framework for Uncertainty and Validation of 3D Registration Methods based on Points and Frames," International Journal of Computer Vision, Vol. 25, No. 3, pp. 203-229, 1997.

[Penny98] G. P. Penney, J. Weese, J. A. Little, P. Desmedt, D. L. G. Hill, D.J. Hawkes, "A comparison of similarity measures for use in 2D–3D medical image registration," IEEE Transactions on Medical Imaging, Vol. 17, pp. 586-595, 1998.

[Pighin02] F. Pighin, R. Szeliski, and D. H. Salesin, "Modeling and Animating Realistic Faces from Images," International Journal of Computer Vision, Vol. 50, No. 2, pp. 143-169, 2002.

[Pito97] R. Pito, "A Registration Aid," Proceedings of International Conference on Recent Advances in 3-D Digital Imaging and Modeling, pp. 85-92, Ottawa, 1997.

[Poelman94] C. J. Poelman and T. Kanade, "A Paraperspective Factorization Method for Shape and Recovery," in Proceedings of ECCV94, pp. 97-108, May, 1994.

[Pollefeys99] M. Pollefeys, Self-calibration and Metric 3D Reconstruction from Uncalibrated Image Sequences, Ph.D. Thesis, ESAT-PSI, K.U.Leuven, 1999.

[Postaire82] J. G. Postaire and C. Vasseur, "A fast algorithm for nonparametric probability density estimation," IEEE Trans. Patt. Anal. Mach. Intell., Vol. 4, No. 6, pp. 663-666, 1982.

[Pottmann04] H. Pottmann, S. Leopoldseder, and M. Hofer, "Registration without ICP," Computer Vision and Image Understanding, Vol. 95, No. 1, pp. 54-71, July 2004

[Press92] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. Numerical Recipes in C: The Art of Scientific Computing, Cambridge University Press, second edition, 1992.

[Qi99] H. Qi, A High-Resolution, Large-Area, Digital Imaging System, Ph.D. Dissertation, North Carolina State University, August, 1999.

[Quan96] L. Quan, and T. Kanade, "A Factorization Method for Affine Structure from Line Correspondences," in Proceedings of CVPR, pp. 803-808, San Francisco, June 18-20, 1996.

[RapidForm02] 3D Reverse Modeler Rapid Form 2002 (Software Manual), Inus Technology Inc., 2002.

[Roche00] A. Roche, G. Malandain, and N. Ayache, "Unifying maximum likelihood approaches in medical image registration," International Journal of Imaging Systems and Technology, Vol. 11, pp. 71-80, 2000.

[Roche98] A. Roche, G. Malandain, X. Pennec, and N. Ayache, "The correlation ratio as a new similarity measure for multimodal image registration," Proceedings of the First International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'98), Lecture Notes in Computer Science, Cambridge, USA, vol. 1496, pp. 1115–1124,1998.

[Rodrigues02] M. Rodrigues, R. Fisher, and Y. Liu, "Special Issue on Registration and Fusion of Range Images," Computer Vision and Image Understanding, Vol. 87, No. 1-3, pp. 1-7, July 2002.

[Rusinkiewicz01] S. Rusinkiewicz01, and M. Levoy, "Efficient Variants of the ICP Algorithm," in Proc. of 3D Digital Imaging and Modeling, pp. 145-152, IEEE Computer Society Press, 2001.

[Sadjadi80] F. A. Sadjadi and E. L. Hall, "Three-Dimensional Moment Invariants," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 2. No. 2, pp. 127-136, 1980.

[Schmid00] C. Schmid, R. Mohr and C. Bauckhage, “Evaluation of Interest Point Detectors,” International Journal of Computer Vision, 37(2), pp. 151-172, 2000.

[Sequiera99] V. Sequeira, K. Ng, E. Wolfart, J. G. M Goncalves, D. Hogg, “Automated Reconstruction of 3D Models from Real Environments,” ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 54, pp. 1-22, 1999.

[Sharp02] G. C. Sharp, S. W. Lee, and D. K. Wehe, “ICP Registration using Invariant Features,” IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 24, No. 1, pp. 90-102, January 2002.

[Silverman82] B.W. Silverman, “Algorithm AS 176: Kernel density estimation using the fast Fourier transform,” Applied Statistics, Vol. 31, No.1, pp. 93-99, 1982.

[Slama80] C. C. Slama, Manual of Photogrammetry, Fourth Edition, American Society of Photogrammetry, Falls Church, VA, 1980.

[Soatto98] S. Soatto and P. Perona, “Reducing Structure From Motion, a General Framework for Dynamic Vision part 1: Modeling,” IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 20, No. 9, September 1998.

[Stamos00] I. Stamos and P. Allen, “3D Model Construction Using Range and Image Data,” Proceedings of CVPR00, vol. I, pp. 531-536, 13-15 June 2000.

[Stein92] F. Stein and G. Medioni, “Structural Indexing: Efficient 3-D Object Recognition,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14, No. 2, pp. 125-145, February 1992.

[Stoddart96] A. J. Stoddart, A. Hilton, "Registration of multiple point sets,” Proceedings 13th International Conference on Pattern Recognition, pp. 40-44 Vienna, Austria, 1996.

[Szeliski99] R. Szeliski and R. Zabih, “An Experimental Comparison of Stereo Algorithms,” in Proc. International Workshop on Vision Algorithms, pp. 1-19, Springer, Kerkyra, Greece, September 1999.

[Tang99] C. K. Tang, G. Medioni and M. S. Lee, "Epipolar Geometry Estimation by Tensor Voting in 8D,” in Proceedings of ICCV 99, Greece, Sep. 1999.

[Thirion96] J.-P. Thirion, “New Feature Points Based on Geometric Invariants for 3D Image Registration,” International Journal of Computer Vision, Vol. 18, No. 2, pp. 121-137, 1996.

[Thirion98] J.-P. Thirion, “Image matching as a diffusion process: an analogy with Maxwell’s demons,” Medical Image Analysis, Vol. 2, pp. 243-260, 1998.

[Tang01] C.-K. Tang; G. Medioni; M.-S. Lee, N-dimensional tensor voting and application to epipolar geometry estimation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, No. 8 , pp. 829-844, August 2001.

[Toga99] A. W. Toga, Brain Warping, Academic Press, San Diego, 1999.

[Tomasi92] C. Tomasi and T. Kanade, “Shape and Motion from Image Streams under Orthography-a Factorization Method,” International Journal on Computer Vision, Vol. 9, No. 2, pp. 137-154, Nov. 1992.

[Tong01] W.S. Tong, C.K. Tang, and G. Medioni, “First Order Tensor Voting, and Application to 3-D Scale Analysis,” in Proceedings of CVPR01, pp. 175-182, 2001.

[Tong04] W.-S. Tong, C.-K. Tang, G. Medioni, “Simultaneous Two-View Epipolar Geometry Estimation and Motion Segmentation by 4D Tensor Voting,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 26, No. 9, pp.1167-1184, Sept. 2004.

[Trucco99] E. Trucco, A. Fusiello, and V. Roberto, “Robust motion and correspondence of noisy 3-D point sets with missing data,” Pattern Recognition Letters, Vol. 20, pp. 889-898, 1999.

[Tsai87] R. Y. Tsai, “A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology using Off-the-Shelf TV Cameras and Lenses,” IEEE Journal of Robotics and Automation, Vol. 3, No. 4, pp. 323-344, August 1987.

[Turk94] G. Turk and M. Levoy, "Zippered Polygon Meshes from Range Images,” in Proceedings of Siggraph 94, Orlando, Fla., pp. 311-318, 1994.

[Viola97] P. Viola and W.M. Wells, “Alignment by maximization of mutual information,” International Journal of Computer Vision, Vol. 24, pp. 137-154, 1997.

[Weinhaus97] F. M. Weinhaus and V. Devarajan, "Texture Mapping 3-D Models of Real-world Scenes,” ACM Computing Surveys, Vol. 29, pp. 325-365, December 1997.

[Wolberg90] G. Wolberg, Digital Image Warping, IEEE Computer Society Press, 1990.

[Wollny02] G. Wollny, F. Kruggel, “Computational cost of nonrigid registration algorithms based on fluid dynamics,” IEEE Transactions on Medical Imaging, Vol. 21, pp. 946-952, 2002.

[Wyngaerd02] J. V. Wyngaerd and L. Van Gool, “Automatic Crude Patch Registration: Toward Automatic 3D Model Building,” Computer Vision and Image Understanding, Vol. 87, No 1-3, pp. 8-26, July 2002.

[Yang03] C. Yang, R. Duraiswami, N. A. Gumerov, L. Davis, “Improved fast gauss transform and efficient kernel density estimation,” Proc. Ninth International Conference on Computer Vision, pp. 464-471, October, Nice, France, 2003.

[Zhang94] Z. Zhang, "Iterative Point Matching for Registration of Free-Form Curves and Surfaces,” International Journal of Computer Vision, Vol. 13, No. 2, pp 119-152, 1994.

[Zhang98] Z. Zhang, “Determining the Epipolar Geometry and its Uncertainty – a Review,” International Journal of Computer Vision, Vol. 27, No. 2, pp. 161-195, March 1998.

[Zhang02] C. Zhang and F. Cohen, “3D Face Structure Extraction and Recognition From Images using 3D Morphing and Distance Mapping,” IEEE Transactions on Image Processing, Vol. 11, No. 11, pp. 1249-1259, November 2002.

[Zinsser03] T. Zinsser, J. Schmidt, H. Niemann, “A refined ICP algorithm for robust 3-D correspondence estimation,” Proceedings of the International Conference on Image Processing, Vol. 2, pp. 695-698, September 2003.

[Zitova03] B. Zitova., J. Flusser, “Image registration methods: a survey,” Image and Vision Computing, Vol. 21, No. 11, pp. 977-1000, 2003.

[Zhu02] Y. M. Zhu, “Volume image registration by cross-entropy optimization,” IEEE Transactions on Medical Imaging, Vol. 21, pp. 174-180, 2002.

Vita

Faysal Boughorbel was born in Tunis, Tunisia on September 14, 1974. After earning the Baccalaureate in Sciences (National High School Diploma) degree in June 1992, he joined the Preparatory School for Engineers in Nabeul. In June 1994 he passed the National Engineering Schools Entry Exam and joined The National School of Engineers of Tunis (ENIT), where he specialized in Electrical Engineering. In July 1997 he graduated from ENIT with the Engineering Diploma and joined the Master's program at the same school. From January 1999 to August 2000 he was a Visiting Research Scholar at the Imaging Robotics and Intelligent Systems Laboratory (IRIS) at the University of Tennessee. Meanwhile he earned his Master's degree from ENIT for his research on photo-realistic 3D reconstruction from range and color images. In August 2000 he enrolled in the Ph. D. program at the department of Electrical and Computer Engineering at the University of Tennessee. Since then he has worked with the IRIS Lab as a Graduate Research Assistant, conducting research in 3D Imaging, Computer Vision, and Pattern Recognition. He earned his Doctor of Philosophy degree in May 2005.