Studienrichtung Vermessungswesen
Technische Universität Wien

GEOWISSENSCHAFTLICHE
MITTEILUNGEN

Heft 51

A Similarity Measure for Global Image Matching Based on the Forward Modeling Principle

von

Christian Öhreneder

Veröffentlichung des Institutes für Photogrammetrie und Fernerkundung

Geowiss. Mitt.
51, 1999

Wien, im April 1999

Studienrichtung Vermessungswesen
Technische Universität Wien

GEOWISSENSCHAFTLICHE
MITTEILUNGEN

Heft 51

A Similarity Measure for Global Image Matching Based on the Forward Modeling Principle

von

Christian Öhreneder

Veröffentlichung des Institutes für Photogrammetrie und Fernerkundung

Geowiss. Mitt.
51, 1999

Wien, im April 1999

Herausgeber und Verleger: o. Prof. Dr.-Ing. Karl Kraus
Vorstand des Institutes für Photogrammetrie und Fernerkundung
der Technischen Universität Wien
A-1040 Wien, Gußhausstraße 27-29

Die Kosten für den Druck wurden aus eigenen Einnahmen des Institutes für Photogrammetrie und Fernerkundung der Technischen Universität Wien getragen.

Diese Arbeit wurde an der Technisch-Naturwissenschaftlichen Fakultät der Technischen Universität Wien (Karlsplatz 13, A-1040 Wien, Austria) zur Erlangung des akademischen Grades eines Doktors der technischen Wissenschaften eingereicht.

Begutachter:
Ao. Univ.-Prof. Dr. Josef Jansa
Inst. f. Photogrammetrie und Fernerkundung, Technische Universität Wien
Gußhausstraße 27-29, A-1040 Wien, Austria
Ao. Univ.-Prof. Dr. Hans Georg Feichtinger
Numerical Harmonic Analysis Group (NUHAG), Inst. f. Mathematik, Universität Wien
Strudlhofg. 4, A-1090 Wien, Austria

Druck: Laser-Print-Express
Auflage: 300 Stück
ISBN 3-9500791-2-2

Die Hinterseite des Mondes sieht man meistens nicht.

One rarely sees the backside of the moon.*

(Anonymous)

*It is not known whether this is meant as an allegory for the problem of cognition or whether it refers to the issue of occlusions ☺.

Abstract

For the purpose of surface reconstruction from two or more digital photographs, it is necessary to determine points in the images that portray the same surface point. The process that seeks to retrieve the correspondences is called image matching. For the calculation, one image, which may be a virtual image like the orthophoto or the cyclopean image, serves as reference domain. Admitting virtual images, it is possible to consider various constellations, such as image-to-image matching and surface reconstruction by orthophoto-to-image matching, in a common framework. The reference domain serves for the parameterization of the surface, abstractly of the geometric connection. An important aspect is, whether it leaves parts of the geometric connection unmodeled. Global image matching seeks to calculate maps from the reference domain to the image domains. In contrast, local methods consider point-to-point correspondences only. Parameterizations of the maps by scalar matching functions, e.g. disparity or height, are used to enforce geometric constraints. The reference domain serves as support for geometric and radiometric properties of the surface model.

The proposed method is intensity-based. It follows the energy minimization paradigm. The energy comprises image energy terms, one for each processed image, and a deformation energy term, which accounts for the a priori probability of the matching function. In the course of modeling, the images are treated as continuous rather than sampled functions. The energy terms are integrals and the minimization problem is a variational problem. The image energy of a single image, also called similarity measure, indicates the conformance of the image with the model in terms of intensity values. The formulation follows the principle of forward modeling. This means that the similarity measure is based on a model that describes the dependency of the image data on the parameters of the model of the surface. The resulting similarity measure exhibits the following advantages with respect to related approaches: Regions of an image contribute to the similarity measure according to their size in image space and not according to their extent in the model. The similarity measure is invariant with respect to the choice of reference domain. For multi-image matching, the images that portray a particular surface area best, that is with best spatial resolution, dominate the reconstruction process. The similarity measure supports the modeling in the presence of occlusions and ensures continuity in that case. If for the contour lines a regularity assumption holds, it is also differentiable. Treating the images as continuous functions is important in order to reveal the structure of the problem. Employing Finite-Elements methods for the discretization leads to a high computational effort. Approximation by finite sums and finite differences is possible in two variants: using a single grid only or using a pair of dual grids.

The similarity measure is not sufficient to determine a reasonable matching function, as it does not preclude arbitrary irregular solutions. The regularization term or deformation energy is used to penalize non-smooth matching functions. Another means is to restrict the solution to a subspace of smooth functions. For the optimization, a method, called the Gauss-Newton-Jacobi method, is proposed, which is closely related to the Gauss-Newton method. It resembles a relaxation scheme that is popular in the course of image matching, but it exhibits better convergence properties. A coarse-to-fine scheme is employed to enlarge the region of convergence.

An image-to-image matching constellation is successfully used for a measurement system for facial surgery. The reconstruction of the shape of a human face from two images takes 3 to 7 min. on a standard PC. For this application, the matching function is composed of bilinear basis elements. 65×65 node values covering a region of 513×513 pixels are determined simultaneously.

Kurzfassung

Für die Rekonstruktion von Objektoberflächen aus zwei oder mehreren digitalen Photographien sucht man Punkte in den Bildern, die dem selben Oberflächenpunkt entsprechen. Dieser Prozeß wird Bildzuordnung genannt. Ein Bild, das auch ein virtuelles Bild sein kann, wie z.B. das Orthophoto oder ein Zyklopen-Bild, dient als Referenzbereich. Läßt man virtuelle Bilder zu, dann kann man unterschiedliche Konstellationen, wie die Bild-zu-Bild Zuordnung oder die Oberflächenrekonstruktion durch Orthophoto-zu-Bild Zuordnung, gemeinsam behandeln. Der Referenzbereich dient der Parametrisierung der Objektoberfläche, abstrakt gesprochen des geometrischen Zusammenhangs. Entscheidend bezüglich der Wahl des Referenzbereiches ist dabei, inwieweit eine vollständige Parametrisierung erzielt wird. Bei der globalen Bildzuordnung wird nach Abbildungen des ganzen Referenzbereichs in die Bildbereiche gesucht. Im Gegensatz dazu ermitteln lokale Methoden nur Punkt-zu-Punkt Zuordnungen. Die Abbildungen werden durch skalare Zuordnungsfunktionen, wie etwa durch Parallaxen oder Höhen, parametrisiert, um geometrische Bedingungen in das Modell zu integrieren. Der Referenzbereich ist der Definitionsbereich der geometrischen und radiometrischen Eigenschaften des Modells.

Das vorgeschlagene Verfahren ist intensitätsbasiert und stützt sich auf das Prinzip der Energieminimierung. Der Energieterm umfaßt Beiträge der einzelnen Bilder und einen Regularisierungsterm. Die Bilder werden als kontinuierliche Funktionen behandelt. Die Energieminimierung führt auf ein Variationsproblem. Der Energiebeitrag eines Bildes wird als Ähnlichkeitsmaß bezeichnet, weil er die Übereinstimmung von Modell und Bild im Bezug auf die Intensitätswerte angibt. Die Formulierung basiert auf dem Prinzip der Vorwärtsmodellierung, das heißt auf einem Modell, das die Abhängigkeit der Bilddaten von den Modellparametern beschreibt. Gegenüber anderen Ansätzen ergeben sich folgende Vorteile: Bereiche eines Bildes tragen zum Ähnlichkeitsmaß entsprechend ihrer Größe im Bildbereich und nicht entsprechend ihrer Ausdehnung im Referenzbereich bei. Das Ähnlichkeitsmaß ist invariant im Bezug auf die Wahl des Referenzbereichs. Bei der Mehrbildzuordnung dominieren jene Bilder die Rekonstruktion, die einen bestimmten Teil der Oberfläche mit bester geometrischer Auflösung darstellen. Verdeckungen lassen sich auf natürliche Weise berücksichtigen, wobei keine Unstetigkeiten des Energiemaßes auftreten. Gilt für die Konturlinien eine Regularitätsbedingung, so bleibt auch die Differenzierbarkeit erhalten. Für die Analyse der Problemstruktur ist die Modellierung mit Hilfe kontinuierlicher Funktionen vorteilhaft. Die Diskretisierung mit Hilfe von finiten Elementen ist rechnerisch sehr aufwendig. Die Approximation mit Hilfe von finiten Summen und Differenzen kann mit einem Gitter oder mit einem Paar dualer Gitter erfolgen.

Das Ähnlichkeitsmaß allein bestimmt die Lösung nur ungenügend, weil beliebig unregelmäßige Lösungen nicht ausgeschlossen werden. Der Regularisierungsterm hilft glatte Lösungen zu finden, indem diese nur einen kleinen Beitrag zur Gesamtenergie liefern. Eine andere Möglichkeit besteht darin, die Lösung auf einen Teilraum von glatten Funktionen einzuschränken. Für die Optimierung wird das Gauß-Newton-Jacobi Verfahren vorgeschlagen, das dem Gauß-Newton Verfahren eng verwandt ist. Es ähnelt einem innerhalb der Bildzuordnung verbreiteten Relaxationsverfahren, weist aber bessere Konvergenzeigenschaften auf. Bildpyramiden werden verwendet, um den Konvergenzradius zu vergrößern.

Die Methode wird in einem Meßsystem für die Gesichtschirurgie eingesetzt. Es wird eine Bild-zu-Bild Zuordnung durchgeführt. Die Rekonstruktion eines Gesichtes aus zwei Bildern benötigt 3 bis 7 min. auf einem Standard-PC. Dabei wird die Zuordnungsfunktion aus bilinearen Elementen zusammengesetzt. 65×65 Knoten in einem Gebiet von 513×513 Pixel werden simultan berechnet.

Acknowledgements

This work has been carried out at the Inst. of Photogrammetry and Remote Sensing (IPF) at the Vienna University of Technology. The IPF is truly remarkable for its positive and encouraging atmosphere. This thesis profited from discussions with numerous colleagues primarily of the IPF. I would like to thank all who have contributed to this work in whatever way. Special thanks I owe my supervisor Ao. Univ.-Prof. Dr. Josef Jansa. He always took time for my concerns, but gave me large liberty to determine the central points of my research and to develop the concept of this thesis. The second supervisor was Ao. Univ.-Prof. Dr. Hans Georg Feichtinger of the Numerical Harmonic Analysis Group at the University of Vienna. He supported me as contact person for mathematical questions. His contributions kept practical relevance always in mind. Further, I would like to thank O. Univ.-Prof. Dr.-Ing. Karl Kraus the head of the IPF. He encouraged me to start and to finish this work. The latter is of equal value as the former, as nothing is ever finished completely.

The primary financial support for this work came from the Doctoral Scholarship of the Austrian Academy of Sciences. Additionally I received financial support from the IPF and the Austrian Science Fund, project no. S7004 Theory and Application of Digital Image Processing and Pattern Recognition/Stereovideometry and Spatial Object Recognition.

My wife will heave a sigh of relief when this doctoral thesis is completed, and so will I. Her, my parents, and all friends I would like to thank for their support especially in the last heavy year.

Contents

Chapter 1

Introduction

The extraction of information from visual stimuli is a task solved by the human vision system with superb efficiency and reliability. At the front end, the receptors of the eyes translate the incident light into a spatial array of stimuli, an image of the world around us. Subsequently, a network of neurons performs a variety of operations such as the fusion of the images of left and right eye, the perception of depth and motion, and the recognition of objects. We can rely on our ability to perform all these tasks without thinking. Only with the help of experiments and optical illusions, we get aware of the rich functionality. Computer Vision is the discipline that tries to find methods for performing these vision tasks on computers. This work is concerned with one such problem, the reconstruction of the 3D geometry of objects from two or more images. Contributions to this topic come from researchers with diverse background. The determination of geometrical quantities from photographs is however a central area of Photogrammetry. What has been done on analog and analytical plotters using conventional photographs, is replaced increasingly by automated digital processes using digital images. Applications range from the derivation of terrain models from aerial imagery, to quality control tasks in industrial production lines, and medical applications. The results of this work are used in a measurement system for facial surgery, which has been developed in the course of a collaboration with the University Clinic of Oral and Maxillofacial Surgery at the General Hospital of Vienna. The system serves for the determination of the geometric shape of human faces. Material from this application is used in Chapter 5 for testing purposes. The method has however not been tailored specifically for this particular case, but can be applied under other circumstances, too.

The principle for determining the 3D geometry of objects from images on a computer is the same as used by the human vision system. To enable surface reconstruction, the images have to portray the object from different viewing angles. The central problem for the reconstruction process is the detection of homologous points in the images. Homologous points are points in different images that correspond to the same point of the object surface. The retrieval problem is called the *correspondence* problem or *image matching* problem. Given information about the imaging geometry, the positions of homologous or corresponding points in the images can be used to infer the 3D coordinates of the respective surface point.

The human vision system is able to perceive depth and relative distances from monocular cues, such as the perceived size of known objects, too. Stereopsis is however the most precise tool for the determination of spatial relations and the geometric shape of anything in the surrounding. The well-known example to try touching the tips of two pencils with outstretched arms is a perfect demonstration. Doing it with one eye closed is a non-trivial task. Using both eyes, the relative depths of the tips of the pencils are clear and the problem is merely

physically controlling the arms.

1.1 Image Matching Methods

Commonly the methods for image matching are grouped according to the primitives that are employed. Primitives are the basic components of an image representation. They constitute the language for expressing the correspondence problem. The primary quantities of an image are the intensity values for each pixel[1]. Methods that are based on raw intensities are called *intensity-based.* This notion is not meant to preclude color. The main characteristic of intensity-based matching is that no information extraction prior to the matching is performed. These methods aim at the utilization of the full amount of available information. The matching algorithm that is considered in this work is intensity-based. More remarks on this type of methods are given subsequently. A strategy contrary to intensity-based matching is followed by methods that limit the correspondence search to features that promise particular reliable and accurate matching. Features are groups of intensity values that are well distinguished from other parts of an images, for example edges and corners. The respective methods are called *feature-based.* Examples can be found in [17, 42]. The apparent advantage is the significant reduction of data size. The benefit causes a disadvantage though, because reducing the matching problem to the matching of features reduces also the density of the surface description. It is also possible to proceed upwards in the hierarchy of primitives and use the spatial relation between features as guideline for the matching process. This leads to the concept of *relational* matching.

1.1.1 Intensity-Based Matching Methods

The first step in intensity-based matching is to define a measure that indicates how similar two or more images or regions of images are in terms of intensities. The second step is to search for optimal similarity with respect to this measure. Some methods base their similarity measure on simple derived quantities such as the gradient of the intensity or on wavelet feature vectors [43, 44]. This is not contrary to the principle of intensity-based matching, as the term includes no prescription how to compare the intensities. The point is that no extraction of salient features or grouping of the intensity values is performed.

One simple variant is the correlation principle. The strategy is to find maxima of the correlation coefficient

$$\rho = \frac{< g - \mu_g, f - \mu_f >}{||g - \mu_g||\,||f - \mu_f||} \,. \tag{1.1.1}$$

g denotes a template array of intensities of moderate size (e.g. 17×17) taken from one image. f is a subset of the same size taken from a second image at the position that is to be checked for correspondence. μ_g and μ_f denote the respective mean values. The template is only superposed and shifted in the second image. Geometric distortions, such as foreshortening etc., are taken into account by a refined approach. The method proceeds patch by patch aiming at the localization of one point per step. The surrounding of each point is included within the correlation window, but the eventual result is a single pair of coordinates of the corresponding position in the second image. This is the characteristic of *local* methods, whereas *global* methods aim at retrieving the correspondence for extended patches.

[1] Pixel is an acronym for picture element.

Searching for maximal correlation of two image patches is very similar to minimizing the sum of squared intensity differences, or squared Euclidean distance

$$E = ||g - f||^2 \,. \tag{1.1.2}$$

Methods following this principle are referred to as *least-squares matching* methods. The close relation of correlation methods and least-squares methods stems from the fact that maximizing the inner product of two vectors of constant length is the very same as minimizing the Euclidean distance. Correlation receivers in digital communication are reasoned by exactly this principle [35, pp. 290]. Least-squares matching methods are easier extended to treat enlarged patches. If the template g covers a large part of the second image f, called the data image, shifting or affine distortion do not suffice to make it similar to the corresponding part of f. The necessity to apply other than affine transformations to the template may be viewed as the criterion for a patch to be considered "large". The methods are then called *global least-squares matching* methods. How the template is deformed to become similar to the data image is a matter of parameterization. The function that determines the spatial relationship is referred to as matching function. For example, parameters of the surface can be used directly to specify the relation between template and data image. An advantage of correlation methods is that they inherently perform a normalization of the image patches. Thereby radiometric differences are factored out automatically. Least-squares methods have to explicitly take the radiometry into account.

The idea of least-squares matching is widely used. The name itself is primarily popular in photogrammetric research. There least-squares matching is usually introduced in the framework of adjustment calculus [16, 50]. Examples of global least-squares matching can be found in [22, 23, 25, 61, 58]. The statistical background of adjustment calculus and the rich experience in this area has proved valuable for the solution of the matching problem. The application of least-squares image matching has so far been strongly linked with the Gauss-Newton method, the optimization method commonly used in adjustment calculus. It appears that this strategy is indeed well suited for image matching. It has however led to an identification of the term least-squares matching not only with the least-squares matching criterion, but also with this particular optimization method. For example, it is often stated that least-squares matching has a superb accuracy potential. The same is true for the correlation method and virtually any intensity-based method. However, the correlation method is often used together with a search strategy, which limits the accuracy to the granularity of the search. In the field of Computer Vision very similar and sometimes completely equivalent approaches to global least-squares matching are considered, though commonly under different names. Further, other optimization methods such as steepest descent [1, 2], relaxation schemes [18, 26, 37, 49], dynamic programming [5, 10], mean field annealing [51], and dynamic Monte Carlo methods [2] are used. As image matching is heavily investigated and contributions and ideas are added from all directions, it is important to strive for a separation of concerns. In particular it is important to keep matching criterion and optimization method logically separated.

The principle of intensity-based matching, namely to search for optimal correspondence in terms of a similarity measure, is very simple. However, even without any detailed knowledge, it is not difficult to conclude that the determination of the optimal correspondence is demanding. The problem appears to necessitate a search strategy like it is often used for correlation methods. Local optimization methods seem to have little chance to succeed, because once far away from the correct solution there is no hint in which direction to find the corresponding point. Some comments with respect to this topic are given in advance,

because the solution of the correspondence problem is tackled only after putting much effort into posing it. To keep the aim vivid, it is important to know how the problem is solved eventually.

The proposed strategy uses a local optimization method. These methods are in general more efficient than global optimizer. However, their application is possible only, if the initial guess of the correspondence is close enough to the correct solution. The region of attraction is bounded by half the wavelength of the highest frequency components of the images.[2] This corresponds to a maximal initial deviation of 1 pixel. This limit is shared by all local optimization methods. Real imagery is commonly lowpass dominated. The requirements are thus not that stringent, enlarging the convergence radius up to 1-3 pixel [22]. In general, the actual available information does however not match the required a priori knowledge. Hierarchical methods, which are very common in Computer Vision [8, 9, 27], provide a workaround. The accuracy of the initial guess has to suffice to start the process at the coarsest level of a pyramidal scheme. The result of each layer is used as initial configuration for the matching algorithm at the next finer layer.

1.2 The Proposed Method for Image Matching

The method proposed herein is intensity-based. It is a global method that is capable of processing large subregions of an image simultaneously. The formulation is based on a continuous model of the images, treating them as continuous rather than sampled functions. The major contribution of this work concerns the formulation of the similarity measure. The practical implementation is based on a least-squares measure though the theoretical concept admits other choices. As a novelty, the principle of forward modeling is applied strictly. Forward modeling means that, starting from a parameterized description of a model, measurable quantities are derived in a unidirectional way. The input-output dependency is used to infer model properties from measurements. Forward modeling is not new, but its application to image matching has been done without the necessary rigor or it has led to computationally very demanding formulations. The group of Wrobel has followed this idea for a long time [29, 58, 61, 62]. Their direct method indeed fulfills the requirements of forward modeling. The computational effort is however considerable. The new results uncover the impact of forward modeling on the matching problem. The eventual similarity measure looks quite similar to alternatives that are more common. The properties demonstrate however that the similarity measure is more in line with the physical image formation process. The computational burden is not increased.

The theoretical results have been used for the implementation of the measurement system for facial surgery. They show the applicability of the new similarity measure. The examples do however not reveal the superiority of the similarity measure with respect to other modeling alternatives. This is due to the fact that only a very simplified image-to-image matching constellation is used. Under these circumstances the capabilities of the new technique are not unfolded. Further, the treatment of occlusions, which are taken into account by the similarity measure in a natural way, requires a visibility analysis and a significantly increased programming effort. As the aim of this work is to setup solid fundamentals, topics, such as the proper discretization of the similarity measure and the theoretical analysis of the impact of occlusions, are assigned a higher priority. This does not mean that practical aspects are ignored. A separate chapter considers the modeling of the matching function and the topic

[2] More comments on this aspect can be found in [22].

of optimization. For the latter an algorithm is used, which has not been found in any other work. The method is closely related to the Gauss-Newton method, but resembles a relaxation scheme that is popular in the course of image matching.

Chapter 2 considers the foundations of the matching algorithm in detail. The geometric and radiometric model is explained and the underlying assumptions and simplifications are stated explicitly. The mathematical framework of the method is presented.

Chapter 3 introduces the similarity measure based on the forward modeling principle. Derivatives and related quantities, which are used for the optimization, are given. The basic concept is formulated without restricting to a particular geometric configuration. These aspects are considered in a separate section. As the primary formulation is a variational problem, the similarity measure and all related quantities have to be discretized. The new similarity measure necessitates a more detailed consideration of the discretization. The last section of this chapter concerns the modeling of occlusions. The current state of implementation does not support the treatment of occlusions. The theoretical properties of the similarity measure in the presence of occlusions are however interesting.

Chapter 4 provides the remaining components that are required to construct a matching algorithm. This comprises the modeling of the matching function and the topic of optimization.

In Chapter 5 a collection of examples is presented. The emphasis is on the measurement system for facial surgery. Practical aspects of the measurement setup, such as the design of a projection pattern for texture enhancement, are provided as well.

In Chapter 6 the results are summarized and the challenges for future developments are mentioned.

Chapter 2

Foundations of the Matching Algorithm

This chapter introduces the modeling concept, its underlying assumptions, and the general framework of the matching algorithm. It is a preparation for the discussion of a specific component of the model, the forward modeling similarity measure, in Chapter 3. The following considerations comprise two distinct aspects, the geometric model and the radiometric model, which together form a complete description of the imaging process. To cover a broad class of geometric configurations a rather generic concept of correspondence is introduced. It extends the commonly used set of notion, aiming at a precise definition of occluded and unmodeled regions. The terminology however covers situations, whose complexity exceeds the capabilities of the matching algorithm. Occlusions for example are part of analysis only, but not subject of practical treatment so far. With respect to radiometry, the basic concept and assumptions are listed. Finally, the general outline of the matching principle, which is posed as a variational problem, is given.

2.1 The Input Data

The input data for the matching problem are digital images. Generically an image is formed by a set of sites $\mathcal{S}$ and labels from a set of labels $\mathcal{L}$ corresponding to each site. This is precisely what is called a mapping or function in mathematics. Digital images generated by CCD based cameras or scanners are commonly defined on rectangular lattices. For monochrome images the labels are usually scalar values from the set $\mathcal{L} = \{0, \ldots, 2^b - 1\}$, also called greyvalues or image intensities,

$$f^d : [0, M-1] \times [0, N-1] \mapsto \mathcal{L} \,. \tag{2.1.1}$$

To have the images defined on a discrete domain is often appropriate for applications such as edge detection, filtering, region labeling etc. Performing Fourier analysis and convolution is possible without referring to a continuous reference space. In contrast, algorithms that need to interpolate between pixel positions inherently take into account that the discrete data values are samples of a continuous function. Formulating the matching problem with images defined on a continuous domain $\mathcal{F}$,

$$f : \mathcal{F} \mapsto \mathbb{R} \,, \tag{2.1.2}$$

avoids mixing the model up with considerations about interpolation kernels and grid spacing. If properly acquired, i.e. if the Shannon theorem holds, f^d represents f within a certain frequency range in a one-to-one manner.

2.2 The Geometric Model

2.2.1 The Notion of Correspondence

According to the above, an image is a function. However, the term image emphasizes a special property of these functions, being a "view" of something. To think of an image this way is not necessarily appropriate for all images in all possible applications, but for the case of stereo matching it is crucial that a notion of correspondence exists, a term closely related to the "view property". The term correspondence may refer to different though interrelated situations. One may talk about a point in an image corresponding to a point in an other image, to a point on the surface of an object, or to a point in the XY plane of a surface represented by heights $Z(X, Y)$. This enumeration already contains three domains or domain types that are used in image matching algorithms. Examples will refer to the following, non-comprehensive list of domains:

- the surface,
- the orthophoto domain,
- the images, and
- virtual images (e.g. the cyclopean view).

When referring to geometrical aspects, the term image is used for brevity instead of the more precise formulation image domain. All of the above listed domains may be considered image domains in a generic sense, a definition follows. Any point of the imaged surface can have a counterpart in any other of these domains. In order to be able to describe different constellations of matching algorithms in a common framework, a clarification of the term correspondence is given. The aim is to define the notion of correspondence independent of the image content and to rationalize that correspondence is inherently unique or unambiguous. This means that for one point in an image there is *at most* one corresponding point in an other image. There may be two points with the same local texture, so that it is impossible to distinguish which one is the correct, but there are never two corresponding points. The term correspondence covers essentially the geometric relationship between points. Conflicting situations to the uniqueness of correspondence are diffusely reflecting or semi-transparent objects. The term correspondence, as it is used herein, simply does not cover such situations. They are precluded within the context of this work. Notably, this concept of correspondence is not suitable for images defined on a discrete domain only, because in this case almost inevitable two object points will correspond to one image point.

Abstractly we define a relation between a point P in a point set $\mathcal{F}$ and point Q in $\mathcal{G}$ and write $P \rightleftharpoons Q$ if the two points correspond. The sets $\mathcal{F}$ or $\mathcal{G}$ may be any of the above listed domains. If there are many images of the same scene there may be many corresponding points, one in each image. For images of 3D objects the points on the object surface $\mathcal{S}$ have a distinctive role in that two points can correspond only if both correspond to the same object point. Nevertheless, for the pure notion of correspondence the relation between corresponding

points is symmetric,

$$P \rightleftharpoons Q \iff Q \rightleftharpoons P \,. \tag{2.2.1}$$

When dealing with multiple images $\mathcal{F}_i$, trivially for points $P_i \in \mathcal{F}_i$ transitivity is required,

$$P_i \rightleftharpoons P_j \wedge P_j \rightleftharpoons P_k \Longrightarrow P_j \rightleftharpoons P_k \,. \tag{2.2.2}$$

Correspondence clearly is defined reflexive. In fact, it is an equivalency relation on the set $\bigcup_i \mathcal{F}_i$, defining equivalence classes of points $[P]$.[1] The set of all correspondences captures the geometric relation between all domains.

Definition 1. The geometric connection C of domains $\mathcal{F}_i$ is given by an equivalency relation $P \rightleftharpoons Q$ of corresponding points $P, Q \in \bigcup_i \mathcal{F}_i$. C is defined as the set of all equivalence classes of corresponding points,

$$C := \left\{ [P] : P \in \bigcup_i \mathcal{F}_i \right\} \,. \tag{2.2.3}$$

Again the domains $\mathcal{F}_i$ may include any domains of interest. The term uniqueness of correspondence is, strictly speaking, a property of the image domains.

Definition 2. A domain $\mathcal{F}$ is an *image domain*, if the *back-projection* $\rho : \mathcal{F} \mapsto C, P \mapsto [P]$ is an injection.

There is no need that every class in C contains a member of every image, but by the required uniqueness there is at most one point of every image domain in an equivalency class. The set C is closely related to the surface $\mathcal{S}$, as the latter may be considered the origin of the correspondences. C is also equipped with a topology, inherited in a natural way from $\mathcal{S}$. Most equivalence classes $[P]$ will contain a surface point. The remaining classes will contain image points with no corresponding surface point. This is primarily a model artifact, and is not considered relevant for the moment. It may in fact be easier to think of $[P]$ as a label attached to every surface point, which contains all corresponding image points. The surface $\mathcal{S}$ obviously plays an outstanding role in the geometric model. For image-to-image matching constellations, it does however not appear in the model explicitly. The set of correspondences is defined without problem, not necessitating the definition of a reference image, or to make the surface an explicit part of the mode.

An example is depicted in Fig. 2.1. The surface is denoted by $\mathcal{S}$, the orthophoto domain by $\mathcal{F}_0$ and the images by $\mathcal{F}_1$, $\mathcal{F}_2$, and $\mathcal{F}_3$. Consider the region between $P'_{\mathcal{S}}$ and $P''_{\mathcal{S}}$ on the surface. With respect to images 0, 2, and 3 everything is well behaved. Examples of correspondence classes are $[P'_{\mathcal{S}}] = \{P'_{\mathcal{S}}, P'_0, P'_2, P'_3\}$ and $[P''_{\mathcal{S}}] = \{P''_{\mathcal{S}}, P''_0, P''_2, P''_3\}$. With respect to image 1 the considered region is *occluded*. In the case of occlusions the closure of the region, i.e. the region including its boundary, is considered (cf. [5]). Consequently P_1 is neither in $[P'_{\mathcal{S}}]$ nor in $[P''_{\mathcal{S}}]$. This decision has not much impact on the modeling anyway, since P_1 is an isolated point and thus has a Lebesgue measure of zero. The argument applies also to the 2D case, where the boundary of the occluded region is a curve. The imaging process of a digital camera is essentially an integrating process. This justifies that sets of vanishing measure are ignored. As will be shown later, the similarity measure is also an integral. Finitely many points or reasonable curves respectively do not contribute to the integral.

[1] The equivalency class [x] is the set of all elements that are equivalent to x. Denoting the equivalency relation by $x \equiv y$, it is given by $[x] := \{y \in X : x \equiv y\}$. One representative can be chosen to denote the whole set. [63, p. 936].

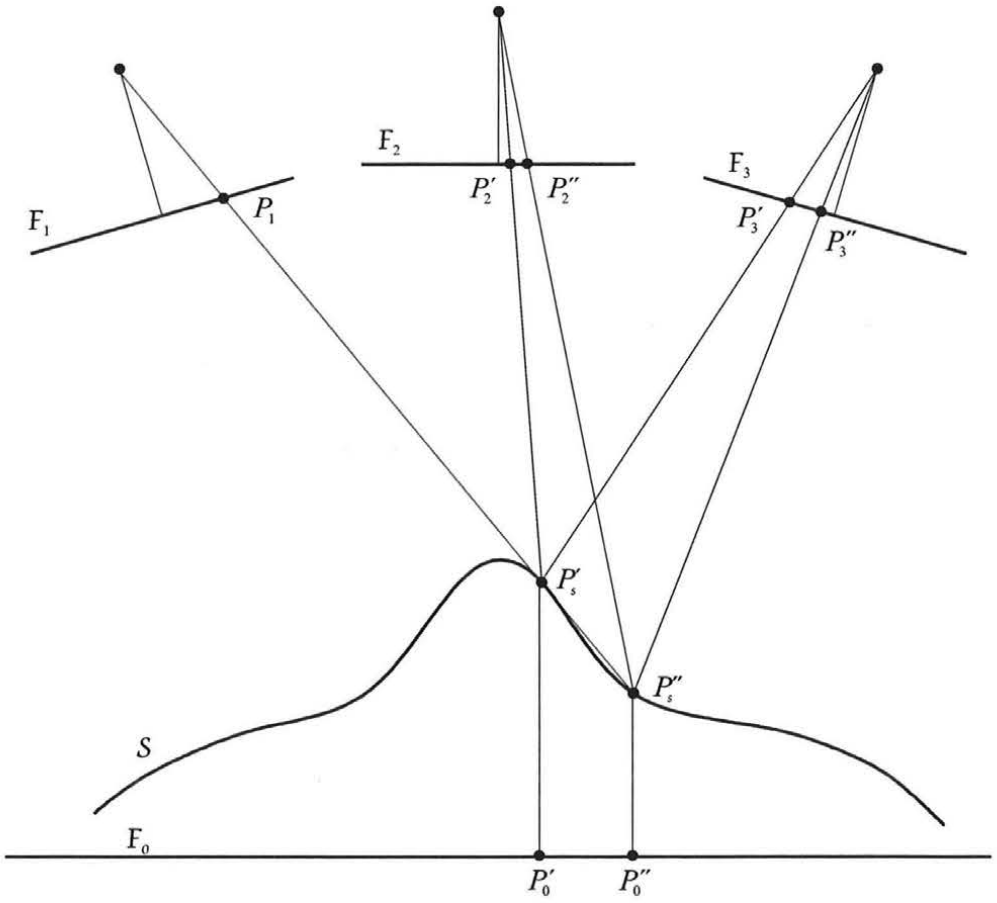

Figure 2.1: Example of an imaging configuration with multiple corresponding points.

Definition 3. The *occluded part* $O_{\mathcal{F}} \in C$ of a geometric connection C with respect to an image domain $\mathcal{F}$ is the closure of all points in C that have no counterpart in $\mathcal{F}$, $O_{\mathcal{F}} = \overline{C \backslash \rho(\mathcal{F})}$. The *visible part* $V_{\mathcal{F}} \in C$ is the complement of the occluded part, $V_{\mathcal{F}} = C \backslash O_{\mathcal{F}}$.

Using this definition, the *image-projection* is given by

$$\iota : V_{\mathcal{F}} \mapsto \mathcal{F}, [P] \mapsto [P] \cap \mathcal{F} \; . \tag{2.2.4}$$

As ι is the inverse function of the back-projection ρ, properly restricted to the visible regions, it is an injection, too.

For the reconstruction of the correspondences, a reference domain is chosen, which is always denoted by $\mathcal{G}$. It satisfies the requirements of Definition 2. Accordingly, for every point in $\mathcal{G}$ there is at most one corresponding point in each considered domain $\mathcal{F}_i$. Let $\mathcal{G}^{V_i} = \iota_{\mathcal{G}}(V_{\mathcal{G}} \cap V_i)$ be the visible part of $\mathcal{G}$, for which there is exactly one corresponding point in a particular $\mathcal{F}_i$.[2] The relation between corresponding points can thus be written by a function

$$m_i : \mathcal{G}^{V_i} \subseteq \mathcal{G} \mapsto \mathcal{F}_i, P \mapsto [P] \cap \mathcal{F} = \iota_i(\rho_{\mathcal{G}}(P)) \; . \tag{2.2.5}$$

The same arguments apply to the opposite direction. It is possible to restrict the range of m_i to a subset $\mathcal{F}_i^M \subseteq \mathcal{F}_i$ so that it is bijective. $\mathcal{F}_i^M = \iota_i(V_{\mathcal{G}} \cap V_i) = m_i(\mathcal{G}^{V_i})$ is the part that is modeled by the reference domain $\mathcal{G}$. A definition of the modeled part follows.

During the matching process the maps m_i are unknown and subject of iterative refinement. Commonly the m_i's are parameterized in order to incorporate geometrical constraints that are known a priori. For example in the case of image-to-image matching of normal case stereo images[3], the map is parameterized by the disparity to enforce the epipolar line constraint. For the case of multiple images, the maps are parameterized consistently, e.g. by a surface representation. The above notions allow considering image-to-image matching by disparity recovery, surface reconstruction from possible multiple images, and a variety of constellations in a common framework.

Together with the texture g, which will be explained in more detail, the maps m_i form the model (m_i, g) for the imaging process. An important issue is the selection of the reference domain $\mathcal{G}$. In principle, it only serves as the parameterization of the set of correspondences C, loosely speaking of the surface $\mathcal{S}$. Importantly, the proposed similarity measure is independent of the chosen reference domain (cf. Chapter 3). Shortcomings of the parameterization do however affect the performance. Direct modeling of the surface is the most rigorous approach. Nevertheless, simplified constellations are considered, e.g. it is quite common to choose one of the images as reference domain. Depending on the surface geometry, these approaches do not always perform a complete parameterization of C, leaving parts unmodeled. In that case $\mathcal{G}$ and C are not topologically equivalent. Consider the case of 1D matching as it is depicted in Fig. 2.2. The domain $\mathcal{G}$ is the reference domain and $\mathcal{F}$ is the domain of some image, it is therefore a given quantity. The interval $[Q', Q'']$ is an example of an occlusion with respect to $\mathcal{F}$. Similarly, the interval $[P', P'']$ is occluded with respect to domain $\mathcal{G}$. The situation is however not as symmetric as it might look. Recall the different roles, $\mathcal{G}$ is subject of model choice whereas $\mathcal{F}$ can not be changed. Mathematically the two situations are distinct

[2] Quantities related to $\mathcal{F}_i$ are denoted with the subscript i rather than the subscript $\mathcal{F}_i$. Quantities related to $\mathcal{G}$ are denoted with the subscript $\mathcal{G}$.

[3] Two images form a stereo pair in normal case arrangement if they were taken by cameras whose camera axes are perpendicular to the baseline. The baseline is the line that connects the projection centers of the cameras.

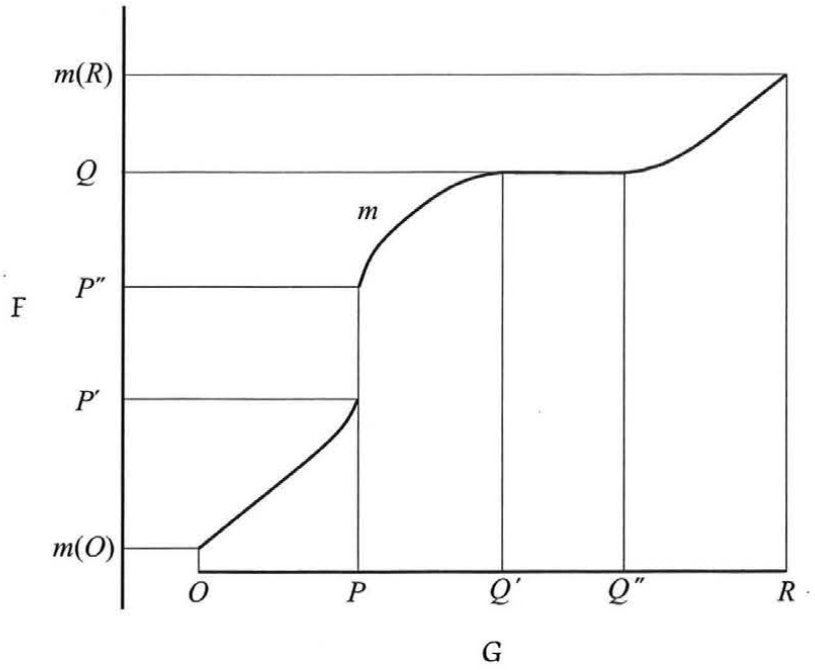

Figure 2.2: 1D matching configuration showing unmodeled and occluded regions.

too. The interval $[Q', Q'']$ has no corresponding image. Graphically this is depicted by a constant value of m over that interval. Though this is intuitive, it is misleading in that there is no mapping from $\mathcal{G}$ to $\mathcal{F}$ for that part. At point P the map m is discontinuous. The interval $[P', P'']$ is simply not captured by the model (m, g), hence this domain is said to be *unmodeled.*

Definition 4. With respect to the reference domain $\mathcal{G}$, the visible part of a geometric connection C is called *modeled part*, and the occluded part is called *unmodeled part.*

Trivially, anything outside the boundaries $m(O)$ and $m(R)$ is also unmodeled, but boundaries are inevitable, since for most objects it will not be possible to cover them completely. If the reference domain $\mathcal{G}$ is chosen such that it coincides with the surface $\mathcal{S}$ of the imaged 3D object and if the description of the surface permits enough flexibility, it is in principle possible to avoid unmodeled regions. Occlusions of course can not be avoided. The distinction between unmodeled and occluded regions is also of interest when considering the discretized formulations of the matching algorithm. Assume that the reference domain $\mathcal{G}$ is covered by an isotropic grid of sampling points (cf. Section 3.4.2). Regions in $\mathcal{F}$, where the corresponding points are distributed very sparsely, are essentially those that are unmodeled or "badly" modeled. This will affect the performance of the matching algorithm. For occluded regions no problems of that kind are encountered. In [5] also a formal concept of occlusion and visibility is introduced. The presented definitions are given with respect to an image-to-image matching constellation that uses a virtual cyclopean image (cf. Section 2.2.2) as reference domain. The notion of unmodeled regions is not part of the concept. By virtue of construction, unmodeled regions either are half-occluded or are not taken into account.

As mentioned, the uniqueness of correspondence grants that the maps m_i are bijective, if the image range is properly restricted. Ambiguities do not have to be feared, only occlusions have to be dealt with. Based on the bijectivity, the existence of the inverse mappings m_i^{-1}

is granted. The reversibility is of great importance for the conceptual development. For the computations, the inversion has to be avoided because of the involved complexity. In order to enable an efficient and well structured implementation, all computations are eventually done in the reference domain. The maps m_i are used to translate quantities of other domains, such as the images f_i, to the reference domain by $f_i \circ m_i$. However, the reference domain serves only as the parameterization of the geometric connection and thus has no intrinsic relevance. Obviously, some choices for the reference space are more natural than others. For example, assigning one image of a multi-image setup a special role, by using it as the reference image, is somewhat unsatisfying. Theoretically this choice is not worse or better than using the orthophoto domain for example. Both domains are topologically equivalent so that none has a superior capability with respect to modeled and unmodeled regions. The results of a matching algorithm should ideally be independent of the particular reference space.

The map m is either a mapping between 2D domains or 1D domains. The extension to the multi-dimensional case is possible, but is considered not relevant. For the most parts of this work no distinction is made between points and coordinates. This is hardly a restriction. When working with multiple resolutions of the images, it may be convenient to use different coordinates for the same point. This seems more an implementation detail, and does not justify an additional burden for the notation. The symbol $\mathbf{x}$ is reserved to denote a vector of coordinates, respectively a point in the reference domain $\mathcal{G}$. Boldface notation is used for vector-valued quantities. Corresponding to $\mathbf{x}$ the components are given by x_1 and x_2. Many considerations take only a single image $\mathcal{F}$ into account. $\mathbf{y}$ is used to denote a vector of coordinates in $\mathcal{F}$. If m is considered a mapping of vector-valued quantities it is denoted by $\mathbf{m}$.

2.2.2 Basic Matching Configurations

The above concept is very general. Nevertheless a couple of matching configurations are of primary importance. The proposed method is intended to be applicable for the reconstruction of object surfaces from images taken by photographic cameras. The typical case is that the imaging relation is given by a central perspective. Distortions of the ideal perspective and other cases than the central perspective are possible. If concrete examples are considered, the standard cases without distortions are used. The parameters of the orientation are assumed known as far as needed. No attempt is made to estimate any of these parameters.

For the 2D case, the map $\mathbf{m}$ is vector-valued and thus has two degrees of freedom at every point. For the case of surface reconstruction, m_1 and m_2 are commonly not considered to be free quantities. The map $\mathbf{m}$ is usually parameterized by a scalar function. In a generic setup this function is called matching function and is denoted by u. The important configurations are

- image-to-image matching,
- 2.5D surface reconstruction, and
- 3D surface reconstruction.

There are a couple of variants for image-to-image matching. Basically, one image serves as the reference domain $\mathcal{G}$ and the corresponding locations in a second image $\mathcal{F}$ are searched. If the imaging geometry is known, the search can be restricted to a line. Given a point $\mathbf{x} \in \mathcal{G}$, all points that possibly correspond to $\mathbf{x}$ can be found on the associated epipolar line in $\mathcal{F}$. For the normal case arrangement, the epipolar lines are parallel. Commonly, they are taken

to be parallel to the x_1 and y_1 direction of the image coordinate systems. The position along the epipolar line is determined by the disparity d. Thus, the map $\mathbf{m}$ is parameterized by the disparity function

$$\mathbf{m}(\mathbf{x}) = \begin{pmatrix} x_1 + d(\mathbf{x}) \\ x_2 \end{pmatrix} . \tag{2.2.6}$$

For other than normal case configurations, generalizations of the disparity can be used, but research in Computer Vision is mainly targeted at the normal case. In order to achieve symmetry, some authors introduce a fictitious cyclopean image, which is located symmetrically between the left and right camera of a normal case arrangement [5, 18, 26]. This domain serves as the reference domain $\mathcal{G}$. As it is a fictitious image domain, no intensity values are available. In such a case, the radiometric properties of the model have to be estimated from the available intensity values of the images. This will be reconsidered later. The mappings to the images are given by

$$\mathbf{m}_L(\mathbf{x}) = \begin{pmatrix} x_1 + \frac{d(\mathbf{x})}{2} \\ x_2 \end{pmatrix} \tag{2.2.7a}$$

and

$$\mathbf{m}_R(\mathbf{x}) = \begin{pmatrix} x_1 - \frac{d(\mathbf{x})}{2} \\ x_2 \end{pmatrix} . \tag{2.2.7b}$$

Another class of models directly seeks to retrieve the Z coordinate or height of the surface for every point of an (X, Y) plane [22, 23, 58]. In this context we call the (X, Y) plane the orthophoto domain. It serves as the reference domain $\mathcal{G}$. The maps $\mathbf{m}_i$ to the various images are consistently parameterized by the surface height $Z(X, Y)$ by means of the well-known equations [31, 32]. Again the texture is commonly unknown and has to be estimated.

The third mentioned constellation is that a full 3D representation of the surface is used for the modeling. This is an ambitious task and so far no implementation is known. It is important for the simple reason that it is obviously the most rigorous approach to the problem. It may be expected that it will be approached one day. It is for that reason that it makes sense to check the applicability of methods to this case.

When analyzing modeling properties such as the continuity of similarity measures, the behavior in the presence of occlusions, discretization schemes etc., the full variability and generality of the concept are a heavy burden. It is however in many cases possible to consider 1D constellations instead. For image-to-image matching the correspondence search occurs along a pair of epipolar lines, this is essentially a 1D problem. In the case of 2.5D surface reconstruction, all possibly matching candidates for a point $(X, Y) \in \mathcal{G}$ are located along the Vertical Line Locus (VLL). This line is the image of the line $X = \text{const}, Y = \text{const}$ and arbitrary Z values. In turn, for a fixed point $\mathbf{y} \in \mathcal{F}_i$, all possible corresponding points in the orthophoto domain are located along a line that runs through the nadir point of the i'th camera[4]. Again, this is a 1D correspondence problem. The situation in the 3D case is different, depending on how the surface is adapted to achieve good correspondence. It may however be assumed locally similar to the 2.5D case. This is not to say that matching in two dimensions is completely analogous to the one-dimensional case, but considering the one-dimensional problem often suffices to derive properties that hold for general cases.

[4]The nadir point lies in the (X, Y) plane. It is located vertically below the projection center of the camera.

2.3 The Radiometric Model

So far we have been concerned with the correspondences of images in a geometrical sense only. Of course, the correspondences are reflected by the relationship of the labels of corresponding points, which provides the bases of the possibility of image matching. The relationship between the labels is caused by the imaging process, which relates imaging labels to object properties. These properties are labels attached to the surface $\mathcal{S}$. For the matching algorithm, the labels are required for every point for which the correspondence is to be recovered. The object properties are hence formulated as a function

$$g : \mathcal{G} \mapsto L \, . \tag{2.3.1}$$

The domain $\mathcal{G}$ is the reference domain, as already used in Section 2.2.1. By the uniqueness of correspondence, there is no danger that this assigns two different labels to the same point of the surface.

Many constellations are imaginable, for example to match a SAR[5] image with a digital photograph. It would however be difficult, if not impossible, to relate the properties of matter within such differing frequency bands. Of interest are images taken by photographic devices. The content of an image, i.e. the value $f(\mathbf{y})$ for each point $\mathbf{y}$, is a measure for the radiation received by the camera from the direction associated with that point. What is relevant of a generic radiometric model depends on the considered frequency range of the radiation. For imaging devices working with visible light, the following aspects are relevant:

- the surface, reflecting incident light towards the camera,
- the atmosphere, attenuating the radiation and adding scattered radiation, and
- the imaging device, which translates incident radiation into image intensities.

A strict forward modeling is prohibited because of the complexity of the phenomena. Actual radiometric models for the matching problem use a mix of modeling considerations based on physical insight and empirical mechanisms that compensate for unmodeled aspects. From the viewpoint of correspondence recovery, it would be optimal, if the labels of corresponding points $\mathbf{y}_i \rightleftharpoons \mathbf{y}_j$ were identical, $f_i(\mathbf{y}_i) = f_j(\mathbf{y}_j)$, coinciding with the expectation that the color or brightness of a surface does not (significantly) vary with the viewing direction. Exact equality may surely not be assumed, though everyday experience reveals that object surfaces look rather similar if viewed from different viewing angles. Among the three listed mechanisms the reflectance properties of the surface appear to be the most critical. The impact of the atmosphere is negligible for close range applications and the sensor characteristics are amenable to calibration. Given a certain constellation of illumination, the amount of radiation reflected in direction of the camera by a small surface element is determined by the properties of the surface matter and the orientation of the surface element. Denoting the radiance (= power per area and tetrahedral angle, $\frac{\mathrm{W}}{\mathrm{m}^2\mathrm{sr}}$) incident upon the surface element from direction $\mathbf{e}_i$ by $L_i(\mathbf{e}_i)$, the radiance reflected in direction $\mathbf{e}_r$ is given by [33, p. 41]

$$L_r(\mathbf{e}_r) = \int_{2\pi} f(\mathbf{e}_r, \mathbf{e}_i, \mathbf{n}) L_i(\mathbf{e}_i) \, (-\mathbf{e}_i) \cdot \mathbf{n} \, d\Omega_i \, . \tag{2.3.2}$$

The function f is called the bi-directional reflectance distribution function. It depends on the material and on the structure of the surface, e.g. the surface roughness. Commonly

[5] Synthetic Aperture Radar

neither the material of the surface nor its reflectance properties are known in advance. For the formulation of the matching algorithm we do, like many others, resort to the model of Lambert reflection, i.e. uniform reflection in all directions. The radiometric properties of the surface are described by a brightness or color label g. It is commonly not calibrated to represent a physical property of matter in a quantifiable way. The nomenclature with respect to the brightness label g is not uniform. In [22, 23, 58] it is called object greyvalue or object intensity value. We will also call it texture, although this term is used with different meaning in Computer Vision applications. According to (2.3.1), g is an image conforming to Definition 2, hence there is not much difference between this image of brightness labels, or whatsoever, and any other image. The transfer functions T establish the relation between surface brightness labels and image brightness for corresponding points $\mathbf{y} \rightleftharpoons \mathbf{x}$

$$f^{\text{ideal}}(\mathbf{y}) = T\big(g(\mathbf{x})\big) \; . \tag{2.3.3}$$

The superscript "ideal" indicates that this quantity is not equal to the eventually obtained image f. Yet to be explained mechanisms cause the images f to deviate from f^{ideal}. For multiple images f_i, a separate transfer function has to be used for each image.

The transfer functions are bijective. This is used to invert the above relation and perform the radiometric adaptation on the images directly. T may be assumed spatially uniform or dependent on the position $\mathbf{x}$. It compensates for many effects, such as the camera transfer characteristic of the imaging device and the atmospheric influences. A spatial varying characteristic has to be used to reduce the influence of non-isotropic reflectance distribution functions. Simple methods, like the one that is used for the practical examples of this work, just perform a global histogram equalization of the images. In [18] a more elaborate treatment is proposed. Inspired by the physics of radiometry, a spatially varying multiplicative factor is used, which is adapted throughout the iterative matching process. In [58] parameters of a shading model, which includes also specular reflections, are recovered during the matching procedure. The transfer function is very often not mentioned in the literature, or a simple global adaptation is used [1, 5, 10, 37, 49, 51]. The further sections will omit T from explicit considerations.

The eventual measured image intensities f deviate from f^{ideal} by two important mechanisms. Firstly, any imaging device has only a limited spatial resolution. The intensity at point $\mathbf{y} \in \mathcal{F}$ is not caused by the contribution of a single point $\mathbf{x}$, but is an integral of the contributions of a certain neighborhood of $\mathbf{x}$. Mathematically this can be formulated by a convolution integral. In this specific context, it is called blurring operation. Including this operation in the model leads to interesting properties of the matching algorithm. However, no computationally feasible structure has been found so far. We know of no approach to image matching that includes the blurring operation as an integral part of the model. It is left aside within this work.

The second source that causes f to deviate from f^{ideal} is the stochastic component of the imaging process. The CCD-sensor of a camera or the scanning device will inevitably corrupt the data by some random components, like sensor noise or quantization noise. Considerations about the probability distribution of the underlying processes directly influence the similarity measure. The by far most popular modeling assumption is that the ideal image intensities are corrupted by additive white Gaussian noise,[6]

$$f(\mathbf{y}) = T\big(g(\mathbf{x})\big) + n_f \; . \tag{2.3.4}$$

[6]No notational distinction is made between stochastic functions and "ordinary" functions.

2.4 Formulation of the Matching Principle

2.4.1 Stochastic Modeling

The stochastic properties of the image data lead to a formulation of the matching principle in a natural way. For the treatment of uncertainties and stochastic influences, which are major topics in many Computer Vision applications, theoretical frameworks exist. Markov random fields and their representations provide a means to specify probabilities and probability distributions for quantities defined on discrete sets of sites [36, 60]. The probability density functions (PDF) of a Markov random field is usually specified via its Gibbsian form,

$$p(x) = \frac{1}{Z} \exp\bigl(-E(x)\bigr) \; . \tag{2.4.1}$$

$E(x)$ is called the energy of the configuration x. Z is a normalization constant. Solving for maximum probability thus is equivalent to minimizing the energy function $E(x)$.

As stated above, Markov random fields are usually defined on a finite domain. If continuous representations are used, this requirement is not fulfilled. From the textbooks [36, 60] it is not clear how to proceed in such a case. It is also characteristical that in [57] stochastic processes, unlike n-dimensional random variables, are defined via a family of probability distributions. Obviously, it is difficult to talk about probability distributions on non-finite dimensional spaces. It is however possible without difficulties to rely on the energy minimization paradigm, which applied to function spaces leads to variational problems. It is for this reason that we will talk about energy functionals rather than probability distributions. Once methods have to be implemented, the restriction to a finite dimensional problem is enforced. Though the model is formulated in a continuous setup, the methods from the Markov random field framework, such as sampling or simulated annealing, are applicable. Despite all theoretical troubles, some explanations, such as the following, refer to the concept of PDF. In these cases a restriction of the underlying configuration space to a finite dimensional subspace is understood, thus essentially $\mathbb{R}^n$ random variables are considered.

Let the model be given by the radiometric component g and a generic matching function u. The stochastic aspects of the imaging process are captured by the conditional PDF $p(f_i|u,g)$.[7] The corresponding energy term is called the *image energy* E^i_{img}. Another name is *similarity measure*, since it indicates the conformance of the image f_i with the template image g subject to the correspondence induced by u. The conditional probability density is definitely not a complete description of the model. It has to be complemented by the a priori model,

$$p(u,g) = p(g|u)p(u) \; . \tag{2.4.2}$$

The energy term associated with the a priori PDF $p(u)$ is called *deformation energy* E_{def}. Other notions, which are frequently used in the course of variational modeling, are *regularization term* or *stabilizing functional*. Using Bayes theorem, the a posteriori distribution of (u,g) is given by

$$p(u,g|f_1,\ldots,f_N) = \frac{p(g|u)p(u)\;\prod_{i=1}^{N} p(f_i|u,g)}{p(f_1,\ldots,f_N)} \; . \tag{2.4.3}$$

[7] It would be more exact to write $p_{\mathbf{f}_i|\mathbf{u},\mathbf{g}}(f_i|u,g)$ [57], which denotes the PDF of the random variable $\mathbf{f}_i$ at f_i conditioned on the random variables $\mathbf{u},\mathbf{g}$ that have the realizations u and g. It should however be clear what is meant.

The corresponding composite energy term is given by

$$E(u, g) = \sum_i E^i_{\text{img}}(f_i|u, g) + E_{\text{def}}(u) \; . \tag{2.4.4}$$

Within this work no a priori information for the texture g is used, thus the term $E(g|u)$ is omitted.

Performing maximum a posteriori (MAP) estimation is equivalent to the minimization of E with respect to g and u,

$$E(u^*, g^*) = \min! \; . \tag{2.4.5}$$

g^* and u^* denote the quantities for which the composite energy assumes its minimum. If g and u are functions on a continuous domain, this leads to a variational problem. The simultaneous optimization with respect to g and u is seldom performed. For image-to-image matching, one of the images may be taken as the (fixed) template image g. This is evidently not a rigorous approach. If the texture g is modeled as unknown quantity, the following principle can be employed. g^* is determined as the minimum of E for a particular u, which is kept fixed,

$$g^*(u) = \operatorname*{argmin}_g E(u, g) \; , \tag{2.4.6}$$

and u^* is found as the minimum of $E(u, g^*(u))$. The minimization with respect to g is often trivial. Many authors apply this principle implicitly. The above method is applied in [5, 10, 22, 58], though in different settings.

2.4.2 Variational Modeling

As already explained, for modeling purposes all functions, especially the images f_i, the texture g, and the maps m_i are assumed on a continuous domain. This significantly facilitates the treatment and avoids premature determination of the discretization scheme. On the other hand, this requires caution, as already became apparent in the previous section. Without further restrictions, there are infinitely many linear independent basis functions for spaces of functions, i.e. the dimension of the spaces is not finite. Many concepts familiar from problems in $\mathbb{R}^n$ can not be carried over to infinite dimensional spaces without further prerequisites. The continuous model is used only as a practical tool to treat a family of problems. The discretization will eventually lead back to spaces with a possible high, but finite number of dimensions. Of course, this does not sidestep the problem entirely. Neither is it the purpose to provide a primarily mathematical focussed treatment nor would this coincide with the author's competence. Therefore, a lack of mathematical rigor is accepted. Wherever it appears necessary and feasible comments are added.

Optimality criterions for functions lead to variational problems, i.e. optimization problems whose underlying parameter space is a space of functions.[8] Of interest are variational problems for functions on a two-dimensional domain. Let u be a scalar function on $\mathcal{G}$ and the functional be given by

$$E(u) = \int_{\mathcal{G}} L(x, y, u, u_x, u_y) \, \mathrm{d}x \, \mathrm{d}y \; . \tag{2.4.7}$$

[8] See [48, 63, 64] on the calculus of variations. Examples of the application of variational principles in Computer Vision can be found in [53, 55] especially for image matching in [1, 18, 26].

The variational problem is stated as

$$E(u^*) = \min! \,. \tag{2.4.8}$$

Commonly for variational problems the first variation is considered. The variation of a function is defined as the derivative with respect to a particular direction h,

$$\delta^n E(u;h) := \left.\frac{\partial^n E(u+th)}{\partial t^n}\right|_{t=0}, \tag{2.4.9}$$

wherein $t \in \mathbb{R}$. This effectively reduces the problem to the differentiation of a function in $\mathbb{R}$. A necessary condition for u^* to be a minimizing function for E is that the first variation vanishes with respect to any function or direction h,

$$\delta E(u^*;h) = 0, \ \forall h \,. \tag{2.4.10}$$

This leads to the Euler-Lagrange equation of the problem

$$\frac{\partial L_{u_x}}{\partial x} + \frac{\partial L_{u_y}}{\partial y} - L_u = 0 \,. \tag{2.4.11}$$

The Euler-Lagrange equation is a necessary, but not sufficient condition for some function u^* to be a minimizer of E. Boundary conditions can be prescribed and incorporated in the framework.

For image matching, the solution of the variational problem can be obtained numerically only. Some methods start from the discretized Euler-Lagrange equation, others apply optimization methods directly to the discretized functional. In all cases a problem in $\mathbb{R}^n$ is solved. It has been found more intuitive to formulate the variational problem analogous to the $\mathbb{R}^n$ problem. In particular not the variation is considered, but the derivative E'. The Euler-Lagrange equation is written in the form $E'(u^*) = 0$. The underlying concept is a generalized notion of differentiation extendible to functionals, the Fréchet derivative [3], [64, p. 463]. Let F be an operator

$$F : U(u_0) \subseteq X \mapsto Y, \tag{2.4.12}$$

which is defined on some neighborhood of u_0. X and Y are Banach spaces, i.e. linear spaces which are complete with respect to some norm. F has a Fréchet derivative at u_0 if and only if there is a continuous linear operator $F'(u_0) : X \mapsto Y$ such that

$$F(u_0 + h) - F(u_0) = F'(u_0)h + \epsilon(h)||h|| \tag{2.4.13}$$

for all $h \in X$, with $||h|| < r$ and $\epsilon(h) \to 0$ in Y for $h \to 0$.

Clearly, the Fréchet derivative is related with the variation (2.4.9). If $F'(u_0)$ exists then

$$\delta F(u_0;h) = F'(u_0)h \,. \tag{2.4.14}$$

Practically the variation is computed and then assumed equal to the derivative by (2.4.14). This requires existence of F'. The existence proofs are considered beyond the scope of what is presented herein and thus are omitted. The way we treat variational problems is readily recognized as a notational style. A high level mathematical concept is used to treat functionals and $\mathbb{R}^n$ problems in the same fashion, though without proof. The benefit is that the relationship between the continuous and the discrete problem is more easily seen. The

reader familiar with optimization in $\mathbb{R}^n$ will find the same methods applied to the discretized variational problem.

E' is a linear functional on the space U, the space of all possible solutions u, and its values are in $\mathbb{R}$. E' is therefore an element of the dual space U^*. For optimization purposes it is convenient to view E' as an element of U itself. If U is assumed a Hilbert space, then by the theorem of Riesz [64, p. 381] there is exactly one element $\nabla E \in U$ such that

$$E'h = \langle \nabla E, h \rangle \ , \quad \forall h \in U \ . \tag{2.4.15}$$

∇E is called the gradient. In this context the inner product is assumed to be given by

$$\langle \nabla E, h \rangle = \int_{\mathcal{G}} \nabla E \, h \, \mathrm{d}A \ . \tag{2.4.16}$$

The symbol $\mathrm{d}A$ denotes the differential element of area. It indicates that the integration extends over a two-dimensional domain. Assuming that (x, y) are components of a vector in a Cartesian coordinate system, $\mathrm{d}A = \mathrm{d}x \, \mathrm{d}y$.

For what is required herein, the distinction between derivative and gradient is not that critical. If the derivative is considered a function in U, for example to formulate recurrence formulas for optimization or simply for graphically illustrations, it is sounder to use the gradient ∇E.

Chapter 3

A Similarity Measure Based on Forward Modeling

A similarity measure, as the term is used herein, is an indication how well an image f conforms to a model $(\mathbf{m}, g)$. The name indicates that good conformance with the model commonly involves that the transformed image $f \circ \mathbf{m}$ is similar to the texture image g, or vice versa f is similar to $g \circ \mathbf{m}^{-1}$. g may be thought of as a template for the image f. The similarity is indicated by an energy value, which is low if the similarity is high. The notion energy or image energy is used because the quantity is the Gibbs energy of a random field describing the image formation process. We will use the terms image energy and similarity measure interchangeably. One has to be aware that this may be misleading in that high similarity is achieved for low energy values. The discrepancy stems from the fact that mathematical optimization usually focuses on function minimization. This also concerns optimal configurations of Gibbs random fields. Contrary, the term similarity measure expresses the objective in a positive way rather than stating what is considered undesirable and has to be minimized. Maybe due to this situation, no standard notions have come up so far.

First, the concept is introduced without restricting to a particular geometric setup, i.e. the map $\mathbf{m}$ is not parameterized and directly serves to establish the correspondence. Parameterizations that enforce particular geometric constraints are treated in Section 3.3. Only a single image f is considered. The resulting energy term is a basic component of the matching algorithm. It can be used to perform image-to-image matching using one image as the model or template for the other. In a multi-image constellation the conformance with the model is indicated by the sum of single image energies. It is thus easily constructed from the basic energy term. Details will be briefly mentioned.

First, the similarity measure based on the forward modeling principle is introduced. Derivatives and other quantities relevant for the optimization are given in a separate section. Afterwards particular geometric constellations and the discretization of the similarity measure are considered. The last section provides some analysis on the modeling in the presence of occlusions.

3.1 The Similarity Measure

For the sequel, the following setup is assumed. The model domain is denoted by $\mathcal{G}$ as usual, and the domain of image f by $\mathcal{F}$. The mapping $\mathbf{m}$ is defined as

$$\mathbf{m} : \mathcal{G} \mapsto \mathcal{F}^M \subseteq \mathcal{F},\ \mathbf{y} = \mathbf{m}(\mathbf{x})\ . \tag{3.1.1}$$

If necessary, coordinates in $\mathcal{G}$ are denoted by $\mathbf{x}$ and coordinates in $\mathcal{F}$ by $\mathbf{y}$. We either use the symbol $\mathcal{F}^M$ to denote the model part of $\mathcal{F}$ or write explicitly $\mathbf{m}(\mathcal{G})$. $\mathcal{G}$ is assumed entirely visible with respect to $\mathcal{F}$.

The formulation of a similarity measure is based on physical insight into the image formation process as well as on practical considerations. Two important assumptions are made. Firstly, the blurring effect is left untreated, thus the image intensity $f(\mathbf{y})$ at a particular point $\mathbf{y}$ depends solely on the surface texture $g(\mathbf{x})$ at the corresponding location $\mathbf{x} \rightleftharpoons \mathbf{y}$. Secondly, the stochastic components of the image f are assumed uncorrelated. $f(\mathbf{y})$ and $f(\tilde{\mathbf{y}})$ are thus mutually independent if $\mathbf{y} \neq \tilde{\mathbf{y}}$ and if conditioned on a particular model $(\mathbf{m}, g)$. First, the similarity measure is given for the case of Gaussian noise. Subsequently, the results are extended to cover general situations.

3.1.1 The Similarity Measure in the Case of Gaussian Noise

Based on the model (2.3.4), but disregarding the transfer function T, the PDF of $f(\mathbf{y})$ is given by a Gaussian distribution for a scalar random variable,

$$p(f(\mathbf{y})|g,\mathbf{m}) = p\big(f(\mathbf{y})|g(\mathbf{m}^{-1}(\mathbf{y}))\big) = \frac{1}{\sqrt{2\pi}\sigma_{n_f}} \exp\left(-\frac{1}{2\sigma_{n_f}^2}\left|f(\mathbf{y}) - g(\mathbf{m}^{-1}(\mathbf{y}))\right|^2\right). \tag{3.1.2}$$

The first equality is a consequence of neglecting the blurring operation of the image formation process. The quadratic argument expression for the exp function is readily identified as the contribution of a single point to the image energy. By the noise n_f is uncorrelated, summing over all modeled points leads to the image energy

$$E(\mathbf{m}) = \frac{1}{2\sigma_{n_f}^2} \int_{\mathcal{F}^M} |f - g \circ \mathbf{m}^{-1}|^2 \, \mathrm{d}A\,. \tag{3.1.3}$$

$\mathrm{d}A$ denotes integration over a two dimensional domain. Subscripts are used if two area measures $\mathrm{d}A$ of different domains might be confused. Otherwise the subscripts are omitted.

The imaging process takes place in $\mathcal{F}$ and hence the stochastic corruption of the image data happens in $\mathcal{F}$. Therefore a statistical measure for the probability of a configuration is inherently linked with this domain. The consideration leading to (3.1.3) are based on a model describing the image formation process. Starting from the quantities $(\mathbf{m}, g)$, which capture the properties of the surface, the image intensities f are derived. This is called forward modeling, as the deduction proceeds from the model to the data. For that reason we refer to the energy term (3.1.3) as *forward modeling similarity measure* or *forward modeling image energy*. It indicates the conformity of the data with the model. As both terms a rather long, short forms, such as forward similarity or forward energy, are used.

Many authors [1, 2, 22, 23, 49, 51] use a different type of image energy. It is a quadratic measure, too, but the contributions are calculated and summed in the reference domain $\mathcal{G}$,

$$E^r(\mathbf{m}) = \frac{1}{2\sigma_{n_g}^2} \int_{\mathcal{G}} |f \circ \mathbf{m} - g|^2 \, \mathrm{d}A\,. \tag{3.1.4}$$

Evidently this is a valid similarity measure in that low energy values indicate that f and g are similar at corresponding positions. The analogue to (2.3.4) for this type of modeling could be formulated by

$$g = f \circ \mathbf{m} + n_g\,. \tag{3.1.5}$$

Since (3.1.4) actually measures the conformity of the model with the data, it is herein called *reverse modeling similarity measure.*

The distinction between forward and reverse modeling takes the different roles of g and f into account. It is easy to confuse shortcomings of a particular model choice with shortcomings of the concept of forward and reverse modeling. In a simplified constellation for image-to-image matching, it may be appropriate to use one image as the template for the other. Forward and reverse similarity appear equally justified in such a situation. The reason is the arbitrary assignment of the roles of the two images. From such an example it is not possible to conclude that forward and reverse similarity are just two sides of the same thing. The principal difference between forward and reverse similarity is that the first is capable of measuring the similarity with respect to the metric of the image f. To see the point, it is helpful to consider multiple images f_i, each defined on a different image domain $\mathcal{F}_i$ of the same geometric connection C. The forward similarity compares image and model with respect to the individual metric of each image. Regions of an image contribute to the similarity measure according to their size in image space and not according to their extent in the model. For the reverse similarity, the comparison happens in reference space. Independent of the size a region of the model takes in an image, the weight is determined by the extent in the model domain. In the form (3.1.3) the forward similarity is computationally costly, an aspect, which has to be overcome. Before going into detail, generalizations of the above energy terms are introduced.

3.1.2 Generalized Formulation of the Similarity Measure

Inspired by (3.1.3), the global energy indicating the similarity between two images is based on a local measure

$$V : \mathcal{G} \times \mathcal{F} \mapsto \mathbb{R} \,. \tag{3.1.6}$$

It is defined on the space of all possible correspondences $\mathcal{G} \times \mathcal{F}$, indicating the fidelity of the correspondence between two points $\mathbf{x}$ and $\mathbf{y}$. Obviously this will involve the comparison of the values of f and g at the respective positions. The classical choice for the local measure, conforming with (3.1.3), is the *local least-squares measure*

$$V(\mathbf{x}, \mathbf{y}) = \frac{1}{2}|f(\mathbf{y}) - g(\mathbf{x})|^2 \,. \tag{3.1.7}$$

It is no problem to incorporate parameters for the radiometric correction into V or to add an adaptive weight to increase the robustness with respect to outliers. In this case V does not only depend on f and g but maybe on some other functions. An important property of V is that, by its definition, it takes only a single pair of points as argument, in contrast to the global measure E, which takes the mapping $\mathbf{m}$ as argument. Using the concept of a local energy measure appeared to be practical firstly for the compactness of notation and secondly to cover variations of the standard local measure (3.1.7). To keep things intuitive it is always justified to replace V by the more concrete local least-squares measure.

The global energy measure is the integral over the local contributions V, which like (3.1.3) is an integral over the domain $\mathcal{F}^M$. For a point $\mathbf{y} \in \mathcal{F}^M$ the local energy measure at $(\mathbf{m}^{-1}, \mathbf{id}_{\mathcal{F}})(\mathbf{y}) = (\mathbf{m}^{-1}(\mathbf{y}), \mathbf{y})$ has to be evaluated. $\mathbf{id}$ denotes the vector-valued identity mapping on the respective domain. For an overview of the mappings between the various domains see Fig. 3.1. The set $\mathcal{C}$ is defined by $\mathcal{C} = \{(\mathbf{x}, \mathbf{y}) : \mathbf{x} \in \mathcal{G}, \mathbf{y} = \mathbf{m}(\mathbf{x})\}$. It is a subset

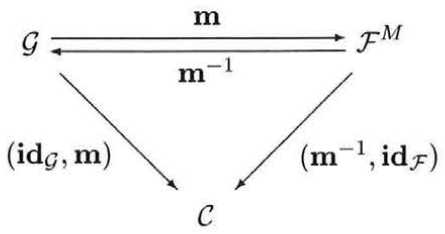

Figure 3.1: Mappings between the domains $\mathcal{G}$, $\mathcal{F}^M$ and $\mathcal{C}$

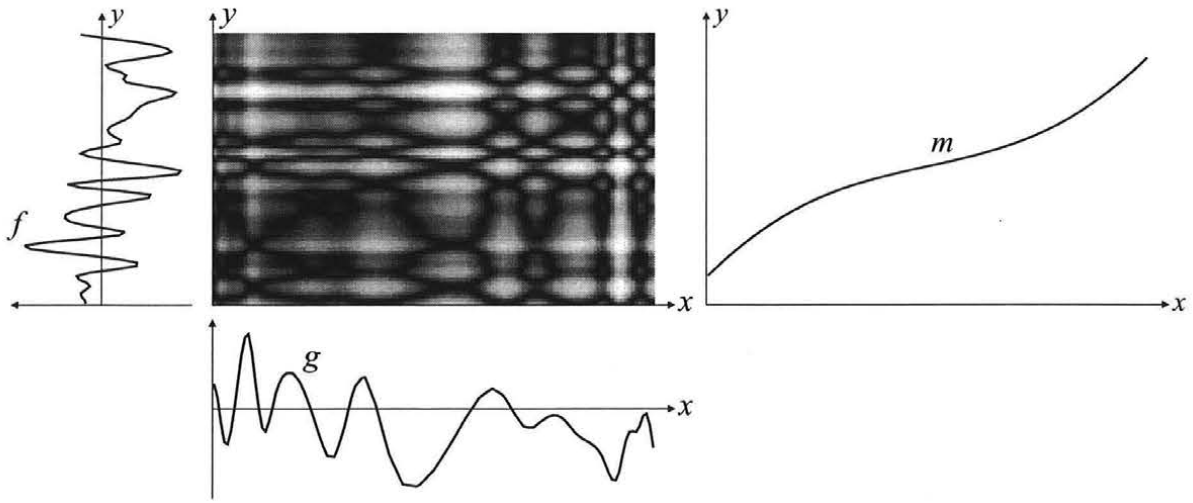

Figure 3.2: Graphical representation of the local energy measure for an artificially generated 1D constellation.

of the space of all possible correspondences. Apart of unmodeled regions, $\mathcal{C}$ is equivalent to the set C in (2.2.3). The generalization of (3.1.3) is given by

$$E(\mathbf{m}) = \int_{\mathcal{F}^M} V \circ (\mathbf{m}^{-1}, \mathbf{id}_\mathcal{F}) \, \mathrm{d}A \, . \tag{3.1.8}$$

The corresponding reverse modeling similarity measure is given by

$$E^r(\mathbf{m}) = \int_{\mathcal{G}} V \circ (\mathbf{id}_\mathcal{G}, \mathbf{m}) \, \mathrm{d}A \, . \tag{3.1.9}$$

Forward and reverse similarity are both based on the local measure. Obviously, the local measure does not uniquely define the global measure. This suggests a review of the role of the local measure. It is defined as a scalar function on the space $\mathcal{G} \times \mathcal{F}$. For 2D constellations, $\mathcal{G} \times \mathcal{F}$ is a domain in $\mathbb{R}^4$. To grasp the essence and keep things intuitive we will restrict to a 1D constellation. In this case V is a function of $\mathbb{R}^2$. It can be illustrated graphically. Fig. 3.2 shows an artificially generated 1D example. The ideal mapping m is shown on the right side. V is assumed to be given by the local least-squares measure (3.1.7). Its values on $\mathcal{G} \times \mathcal{F}$ are represented by an image, showing large values bright and low values dark. For the graphical representation the brightness has been adapted nonlinearly, otherwise the bright and dark spots would appear rather unbalanced. The ideal map m is seen as an elongated minimum of the same shape as m, indicating the optimal correspondence. Both

global measures are computed upon integrating over an intersection of the space $\mathcal{G} \times \mathcal{F}$. For the integral (3.1.8) V is evaluated at $\{(\mathbf{m}^{-1}(\mathbf{y}), \mathbf{y}) : \mathbf{y} \in \mathcal{F}^M\}$. This is exactly the set $\mathcal{C}$, which for a 2D constellation is a 2D face embedded in $\mathbb{R}^4$. In a similar way, the integral (3.1.9) involves all values of the local measure at $\mathcal{C}$, which in this case is parameterized over the reference domain $\mathcal{G}$. For 1D constellations the integrals are computed along lines embedded in $\mathbb{R}^2$. The reason that the forward and reverse energy generally yield different values is due to the different area measures $\mathrm{d}A_{\mathcal{F}}$ and $\mathrm{d}A_{\mathcal{G}}$. The integrals thus must not be interpreted as integrals of scalar density functions but as integrals of differential forms [14], in particular of 1- or 2-forms respectively. Using integration theory for differential forms the energy terms can be written as

$$E = \int_{\mathcal{C}} v \tag{3.1.10}$$

and

$$E^r = \int_{\mathcal{C}} v^r \, . \tag{3.1.11}$$

This way, the difference between forward and reverse modeling is reflected by the global and the local measures. We will not use this concept, as knowledge about it is not that widespread. What has to be kept in mind is that the local energy measure must not naively be interpreted as a density function. The graphical depiction of V suggests itself as a tool for understanding the performance of matching algorithms. Though the local energy measure is not a new concept, for it is closely related to Gibbs potentials, the interpretation as a function on $\mathbb{R}^2$ or higher dimension respectively, might provide a means to better understand the convergence properties of optimization methods. In particular, the region of attraction of local optimization algorithms is influenced by the distribution of minima and maxima of the local measure.

An immediately apparent demerit of the forward energy term (3.1.8) is that it involves the inverse map $\mathbf{m}^{-1}$. Utilizing the bijectivity of $\mathbf{m}$, variable substitution can be employed and (3.1.8) written as an integral over $\mathcal{G}$,

$$E(\mathbf{m}) = \int_{\mathcal{G}} V \circ (\mathbf{id}_{\mathcal{G}}, \mathbf{m}) \det \mathbf{m}' \, \mathrm{d}A \, . \tag{3.1.12}$$

Explicitly the Jacobi matrix $\mathbf{m}'$ is given by

$$\mathbf{m}' = \begin{pmatrix} \frac{\partial m_1}{\partial x_1} & \frac{\partial m_1}{\partial x_2} \\ \frac{\partial m_2}{\partial x_1} & \frac{\partial m_2}{\partial x_2} \end{pmatrix} \, . \tag{3.1.13}$$

Symbolically the transformation of the domain of integration is done by relating the area measures of $\mathcal{G}$ and $\mathcal{F}$ by

$$\mathrm{d}A_{\mathcal{F}} = \det \mathbf{m}' \, \mathrm{d}A_{\mathcal{G}} \, . \tag{3.1.14}$$

The determinant $\det \mathbf{m}'$ is commonly referred to as Jacobian. In geodesy it is known as the area distortion. The form (3.1.12) is called the *reference space formulation* of the forward energy, whilst (3.1.8) is the *image space formulation*. It follows the principle to perform all calculations with respect to the reference space. The consistency of model and underlying physics does not have to be sacrificed to achieve this computationally favorable structure.

Comparing (3.1.12) and (3.1.9) reveals that they differ by the weighting factor $\det \mathbf{m}'$. In (3.1.12) regions of $\mathcal{G}$ that are visible in $\mathcal{F}$ with bad spatial resolution receive a low weight and vice versa. This also grants that regions of an image contribute to the similarity measure according to their size in image space and not according to their extent in the model. Another way to see the desirable consequences of the weighting factor is by means of the multi-image case, which is considered in the next section. As a limiting case the weighting by the Jacobian $\det \mathbf{m}'$ leads to the modeling of occlusions. Mathematically the Jacobian may vanish at isolated points only, since otherwise the variable substitution is not admitted. This will also be the underlying assumption for the next sections. We will also require the inverse mapping of $\mathbf{m}$ to exist. The case of occlusions is treated separately in Section 3.5. Nevertheless, considering the limiting case that the Jacobian approaches zero, reveals much of the difference between forward and reverse modeling approach.

One of the important characteristics of the forward modeling similarity measure is its independence of the particular reference domain, which can be seen directly from (3.1.8). Changing the reference domain to $\tilde{\mathcal{G}}$, the integrand at position $\mathbf{y}$ is given by $\tilde{V}(\tilde{\mathbf{m}}^{-1}(\mathbf{y}), \mathbf{y})$. The value remains unchanged, only the coordinates of $\tilde{\mathbf{m}}^{-1}(\mathbf{y})$ are different. If $\tilde{\mathbf{m}}(\tilde{\mathcal{G}}) = \mathbf{m}(\mathcal{G})$, that is if the model scope remains the same, nothing changes. This is referred to as the invariance principle. It is a consequence of the fact that $\mathcal{G}$ is used only to parameterize the geometric connection C. In as far as the discretization is affected by the change of the reference domain, the numerical behavior will be obviously affected, too. Altogether, the forward modeling paradigm is more in line with the physical image formation.

3.1.3 Multi-Image Constellations

Estimating the texture g by means of (2.4.6) is easily done if the image energy terms are integrals of local contributions. We restrict to the case that the similarity measure of individual images is based on the local least-squares measure (3.1.7). All energy terms receive the same overall weighting. The plausibility of the weighting factor $\det \mathbf{m}'$ is easily seen from this example. The composite energy term is given by

$$E_{\text{multi}} = \frac{1}{2} \sum_i \int_{\mathcal{G}} |f_i \circ \mathbf{m}_i - g|^2 \det \mathbf{m}_i' \, \mathrm{d}A \,. \tag{3.1.15}$$

For fixed maps $\mathbf{m}_i$ the estimate according to (2.4.6) is easily evaluated to

$$\begin{aligned} g^* &= \frac{1}{w} \sum_i f_i \circ \mathbf{m}_i \det \mathbf{m}_i' \\ w &= \sum_i \det \mathbf{m}_i' \,. \end{aligned} \tag{3.1.16}$$

The applied weighting is very intuitive. It effects that the images that portray a particular surface area best, that is with best spatial resolution, dominate the reconstruction process. In the limiting case of occlusions, the contributions of the respective image parts are automatically eliminated.

The above principle may be used in different settings. For image-to-image matching, one image may serve as the template for the other. More in line with the imaging physics is to assume both images modeled by a common model. To that end, the above scheme can be applied using only two images. A simple example is the matching of two normal case images

using the cyclopean domain as reference domain (Section 2.2.2). The maps $\mathbf{m}_R$ and $\mathbf{m}_L$ are given by (2.2.7). Determining g^* according to (3.1.16) and inserting it into (3.1.15) leads to

$$E_{\text{multi}}(d) = \frac{1}{4}\int_{\mathcal{G}} |f_R \circ \mathbf{m}_R - f_L \circ \mathbf{m}_L|^2 \left(1 - \left(\frac{d'}{2}\right)^2\right) \mathrm{d}A\,. \tag{3.1.17}$$

If the derivative of the disparity d' assumes one of the values ± 2 the Jacobian of $\mathbf{m}_L$ or $\mathbf{m}_R$ vanishes, indicating an occlusion. The local contributions are again properly weighted.

The above matching scheme for two images works for normal case arrangements. For other constellations, there is no equivalent to the cyclopean domain. However, as mentioned repeatedly, the similarity measure is not affected by the choice of reference domain. The same principle as above can be employed for other reference domains, too. Consider two image f_0 and f_1 and let the reference domain $\mathcal{G}$ coincide with the domain $\mathcal{F}_0$. The composite energy term comprises the similarity measure for both images. A simplification is that, by virtue of construction, $\mathbf{m}_0$ is the identity mapping. It hence does not need to be taken into account explicitly. Writing $\mathbf{m}$ instead of $\mathbf{m}_1$, we have

$$E(\mathbf{m}, g) = \frac{1}{2}\int_{\mathcal{F}_0} |f_1 \circ \mathbf{m} - g|^2 \det(\mathbf{m}')\,\mathrm{d}A + \frac{1}{2}\int_{\mathcal{F}_0} |f_0 - g|^2\,\mathrm{d}A \tag{3.1.18}$$

The estimated texture evaluates to

$$g^*(\mathbf{m}) = \frac{f_0 + f_1 \circ \mathbf{m}\det(\mathbf{m}')}{1 + \det(\mathbf{m}')}\,, \tag{3.1.19}$$

and thus

$$E(\mathbf{m}) = E(\mathbf{m}, g^*(\mathbf{m})) = \frac{1}{2}\int_{\mathcal{F}_0} |f_1 \circ \mathbf{m} - f_0|^2 \frac{\det(\mathbf{m}')}{\det(\mathbf{m}') + 1}\,\mathrm{d}A\,. \tag{3.1.20}$$

We refer to this setup as the symmetric variant of image-to-image matching. Intuitively the symmetry is violated by taking one image as reference domain. Nevertheless, both images f_0 and f_1 are treated equally. Changing the domain of integration from $\mathcal{F}_0$ to $\mathcal{F}_1$ by variable substitution leads to an expression that is completely analogous to (3.1.20), only the indices are interchanged and $\mathbf{m}$ is replaced by $\mathbf{m}^{-1}$. The fact that $\mathbf{m}_0$ is the identity mapping involves a couple of simplifications, for example for the calculation of the derivative of the energy term. Altogether, the effort is increased only moderately in comparison with the basic constellation that uses one image as template, e.g. $g \equiv f_0$.

3.1.4 Remarks on the Continuity of the Similarity Measure

Only a few remarks concerning the continuity of the similarity measure are given. For practical purposes, continuity of the discretized similarity measure, as well as differentiability, is interesting as it influences applicability and performance of optimization algorithms. If the Finite-Sums approach (cf. Section 3.4.2) is used for the discretization, the conditions for continuity and differentiability are obtained straight forward. Prior to the discretization, that is for the similarity measure as it is presented in this section, the general conditions are not that easily shown. The following comments are actually rather limited, demonstrating only some principles.

One important point of continuity concerns the norm to be used for the map $\mathbf{m}$. Continuity is a concept linked to the topology of a space, which in the case of Banach spaces is

defined by some norm. In view of (3.1.12), it is clear that for $E(\mathbf{m})$ to be close to $E(\tilde{\mathbf{m}})$ it is not sufficient that $\mathbf{m}(\mathbf{x})$ is close to $\tilde{\mathbf{m}}(\mathbf{x})$ for all $\mathbf{x}$. It is required that the partial derivatives $\frac{\partial m_i}{\partial x_j}$ and $\frac{\partial \tilde{m}_i}{\partial x_j}$ are close, too. For infinite dimensional spaces of functions, these two requirements are not equivalent. More precisely, we assume $\mathbf{m} \in C^1(\mathcal{G}) \times C^1(\mathcal{G})$ and base our considerations on the following norm of $\mathbf{m}$

$$\|\mathbf{m}\|_{1,\infty} = \sum_{i \in 1,2} \left(\max_{\mathcal{G}} |m_i| + \max_{\mathcal{G}} \left| \frac{\partial m_i}{\partial x_1} \right| + \max_{\mathcal{G}} \left| \frac{\partial m_i}{\partial x_2} \right| \right) . \tag{3.1.21}$$

Whether less restrictive norms are sufficient, is not clear.

First, note that this norm ensures that if the Jacobian $\det \mathbf{m}'$ is positive, which has been one of the underlying assumptions so far, then this is valid for a certain neighborhood of the map, too. More precisely, there is a neighborhood $N_\delta(\mathbf{m}) = \{\tilde{\mathbf{m}} : \|\mathbf{m} - \tilde{\mathbf{m}}\| < \delta, \delta > 0\}$, such that for $\tilde{\mathbf{m}} \in N_\delta(\mathbf{m})$ it follows that $|\det \mathbf{m}' - \det \tilde{\mathbf{m}}'| < \frac{c}{2}$. This is simply a consequence of the continuity of the determinant and the fact that the above defined norm ensures uniform convergence of all first order partial derivatives of $\mathbf{m}$. Consequently, if $\det \mathbf{m}' > c$ then $\det \tilde{\mathbf{m}}' > \frac{c}{2}$. This result is important, as it grants the existence of a neighborhood of $\mathbf{m}$ for which the energy term is defined without complications.

Continuity of the similarity measure is established very easily if V is assumed continuous and bounded. Consider the difference of the energy for two mappings $\mathbf{m}$ and $\tilde{\mathbf{m}}$,

$$|E(\mathbf{m}) - E(\tilde{\mathbf{m}})| = \left| \int_{\mathcal{G}} V \circ (\mathbf{id}_{\mathcal{G}}, \mathbf{m}) \det \mathbf{m}' \, \mathrm{d}A - \int_{\mathcal{G}} V \circ (\mathbf{id}_{\mathcal{G}}, \tilde{\mathbf{m}}) \det \tilde{\mathbf{m}}' \, \mathrm{d}A \right| . \tag{3.1.22}$$

Trivially,

$$|E(\mathbf{m}) - E(\tilde{\mathbf{m}})| \leq \int_{\mathcal{G}} |V \circ (\mathbf{id}_{\mathcal{G}}, \mathbf{m}) \det \mathbf{m}' - V \circ (\mathbf{id}_{\mathcal{G}}, \tilde{\mathbf{m}}) \det \tilde{\mathbf{m}}'| \, \mathrm{d}A . \tag{3.1.23}$$

As the norm of $\mathbf{m}$ enforces uniform convergence of $\mathbf{m}$ including all first order partial derivatives, the integrand, which is continuous by assumption, converges uniformly to zero for $\delta \mapsto 0$. The domain $\mathcal{G}$ may be assumed bounded, causing also no complications.

Requiring V to be continuous matches the expectations. The condition is however not generic. Let V be given by the local least-squares measure. Assuming f and g piecewise constant still leads to a continuous similarity measure. This is supported by practical tests and theoretical considerations as well. The principle is very similar to a well-known property of convolution: Convolving two piecewise constant functions leads to a continuous result. A mathematical simple argument to establish the general conditions of continuity for the forward similarity has not been found. Further investigations have not been considered worthwhile, as the are of limited value for the practical implementation. In their generality they can not be used in the discrete context. The condition that V be continuous is relevant for the discretized problem, too.

3.1.5 Related Approaches

The first and so far only application of the forward modeling principle to image matching is due to the group of Wrobel [29, 58, 61, 62]. The formulation is based on the concept of nonlinear adjustment calculus following physical arguments as far as feasible. The set of equation and the thereof-derived optimality principle corresponds to the image space formulation (3.1.8). The resulting algorithm is called the direct method. The indirect method targeted at

a formulation in the domain of $\mathcal{G}$ is not equivalent to the reference space formulation (3.1.12). The reason is that the concept treats the images as a discrete set of measurements rather than continuous functions. The image data as acquired by scanners or CCD-cameras are indeed a discrete set of measurements. The continuous approach however proved to be very valuable, as it supports the use of variable substitution and partial integration, which help to uncover the structure of the problem.

The matching algorithms utilizing the reverse energy or close relatives are numerous [1, 2, 22, 23, 49, 51]. Since a couple of algorithms uses symmetric similarity measures the forward or reverse modeling property is not easily identified. In [5, 18, 26] two normal case stereo images are matched using a cyclopean image and a energy term of the following type

$$\frac{1}{4}\int_{\mathcal{G}} |f_R \circ \mathbf{m}_R - f_L \circ \mathbf{m}_L|^2 \,\mathrm{d}A \,. \tag{3.1.24}$$

The definitions of $\mathbf{m}_R$ and $\mathbf{m}_L$ are given by (2.2.7). The measure lacks the characteristics of forward modeling, which for the particular setup leads to (3.1.17).

3.2 Derivatives of the Similarity Measure

Derivatives are needed for optimization purposes. In particular the first derivative E' is needed by many algorithms to determine and find the optimum $\mathbf{m}^*$ for which E assumes its minimum. The other quantity that is required for the practical implementation is the Gauss-Newton term E^{GN}. For quadratic error measures the Gauss-Newton term is an approximation of the second derivative (cf. Section 4.2.3). The name Gauss-Newton term is however not a standardized notion. In adjustment calculus E^{GN} is referred to as the matrix of the normal equation.

In this section only derivatives with respect to $\mathbf{m}$ are considered. If the map is parameterized by some matching function u, the chain rule has to be applied to get the derivative with respect to u (cf. Section 3.3). The texture g is assumed to be known or to be estimated according to (2.4.6). This has no consequence for the results. Note that

$$\frac{\mathrm{d}E(g^*(u),u)}{\mathrm{d}u} = \frac{\partial E(g^*(u),u)}{\partial g}\frac{\partial g^*(u)}{\partial u} + \frac{\partial E(g^*(u),u)}{\partial u} \,. \tag{3.2.1}$$

By (2.4.6), the partial derivative with respect to g vanishes, presuming differentiability with respect to g. The first derivative as well as the Gauss-Newton term are based solely on first order approximations, thus whether g is known or estimated has no influence.

The reference space formulation suggests that the forward energy can be treated as a weighted least-squares problem. It is indeed possible to ignore the dependency of the weighting factor $\det \mathbf{m}'$ on the mapping $\mathbf{m}$ when calculating derivatives. Practical matching results obtained by this method have been found satisfactory. However, not the original problem is solved but some neighboring problem, because for most optimization algorithms the first derivative is used not only to find but also to determine the minimum. The principle is similar to the iterated reweighted least-squares problem for robust estimation (e.g. [66, p. 70]), which however, in its particular setup, does not change the problem. This line is not pursued. The required results are obtained easily from Section 3.2.4 upon incorporating the weighting factor $\det \mathbf{m}'$. The exact derivative of the forward energy is not easy to compute. It is given in the next section. Again the dual role of forward and reverse modeling similarity measure is revealed.

3.2.1 First Derivative of the Similarity Measure

The calculation of the first derivative can be done using either of the two formulations of the forward energy (3.1.8) and (3.1.12). In this section the image space formulation (3.1.8) is taken as the starting point. This facilitates the geometric understanding. In Section B.1 an alternative calculation starting from the reference space formulation (3.1.12) is given. It has two advantages: Firstly, it is more mathematical. In particular, the necessary steps are the same as needed to establish the Euler-Lagrange equation (2.4.11). Secondly, the numerical implementation is based on the reference space formulation, too. Differentiating the discretized energy term yields completely analogous results. The calculation however faces obstacles, because partial integration and variable substitution can not be carried over to the discrete case adequately. Having the corresponding results from the continuous case at hand facilitates the treatment.

The first step to obtain the derivative of the energy term is to replace the map $\mathbf{m}$ by a slightly altered map

$$\tilde{\mathbf{m}} = \mathbf{m} + t\boldsymbol{\mu} \,. \tag{3.2.2}$$

This is inserted into one of the expressions for the global measure. Differentiating with respect to t yields the first variation of the image energy term. More precisely, the variation is calculated by

$$\delta E(\mathbf{m}; \boldsymbol{\mu}) = \lim_{t \to 0} \frac{1}{t} \big(E(\mathbf{m} + t\boldsymbol{\mu}) - E(\mathbf{m}) \big) \,. \tag{3.2.3}$$

The result is linear with respect to $\boldsymbol{\mu}$. It is taken as the derivative of the image energy (without proof), $E'(\mathbf{m})\boldsymbol{\mu} = \delta E(\mathbf{m}; \boldsymbol{\mu})$.

Inserting the modified map (3.2.2) into (3.1.8) yields

$$E(\tilde{\mathbf{m}}) = \int_{\tilde{\mathbf{m}}(\mathcal{G})} V \circ (\tilde{\mathbf{m}}^{-1}, \mathbf{id}_{\mathcal{F}}) \, \mathrm{d}A \,. \tag{3.2.4}$$

The modification of the map does not only change the single point correspondences, but also results in a modified image domain $\tilde{\mathbf{m}}(\mathcal{G})$. Parts of $\mathcal{F}$ that contribute to $E(\mathbf{m})$ are no longer taken into account by $E(\tilde{\mathbf{m}})$, whilst other regions provide additionally contributions. This shift of the boundary gives rise to a boundary term. The situation is depicted in Fig. 3.3.

To apply (3.2.3), we calculate the difference

$$E(\tilde{\mathbf{m}}) - E(\mathbf{m}) = \int_{\tilde{\mathbf{m}}(\mathcal{G})} V \circ (\tilde{\mathbf{m}}^{-1}, \mathbf{id}_{\mathcal{F}}) \, \mathrm{d}A - \int_{\mathbf{m}(\mathcal{G})} V \circ (\mathbf{m}^{-1}, \mathbf{id}_{\mathcal{F}}) \, \mathrm{d}A \,. \tag{3.2.5}$$

This expression is rearranged to facilitate the separation of boundary and interior terms,

$$\begin{aligned} E(\tilde{\mathbf{m}}) - E(\mathbf{m}) &= \\ &= \int_{\tilde{\mathbf{m}}(\mathcal{G})} V \circ (\tilde{\mathbf{m}}^{-1}, \mathbf{id}_{\mathcal{F}}) - V \circ (\mathbf{m}^{-1}, \mathbf{id}_{\mathcal{F}}) \, \mathrm{d}A && \text{(3.2.6a)} \\ &+ \int_{\tilde{\mathbf{m}}(\mathcal{G}) \setminus \mathbf{m}(\mathcal{G})} V \circ (\mathbf{m}^{-1}, \mathbf{id}_{\mathcal{F}}) \, \mathrm{d}A - \int_{\mathbf{m}(\mathcal{G}) \setminus \tilde{\mathbf{m}}(\mathcal{G})} V \circ (\mathbf{m}^{-1}, \mathbf{id}_{\mathcal{F}}) \, \mathrm{d}A \,. && \text{(3.2.6b)} \end{aligned}$$

(3.2.6a) captures the changes in the interior of $\mathbf{m}(\mathcal{G})$. The fact that the integration extends over $\tilde{\mathbf{m}}(\mathcal{G})$ shall not be reason of confusion. If $t \to 0$, it follows that $\tilde{\mathbf{m}}(\mathcal{G}) \to \mathbf{m}(\mathcal{G})$.

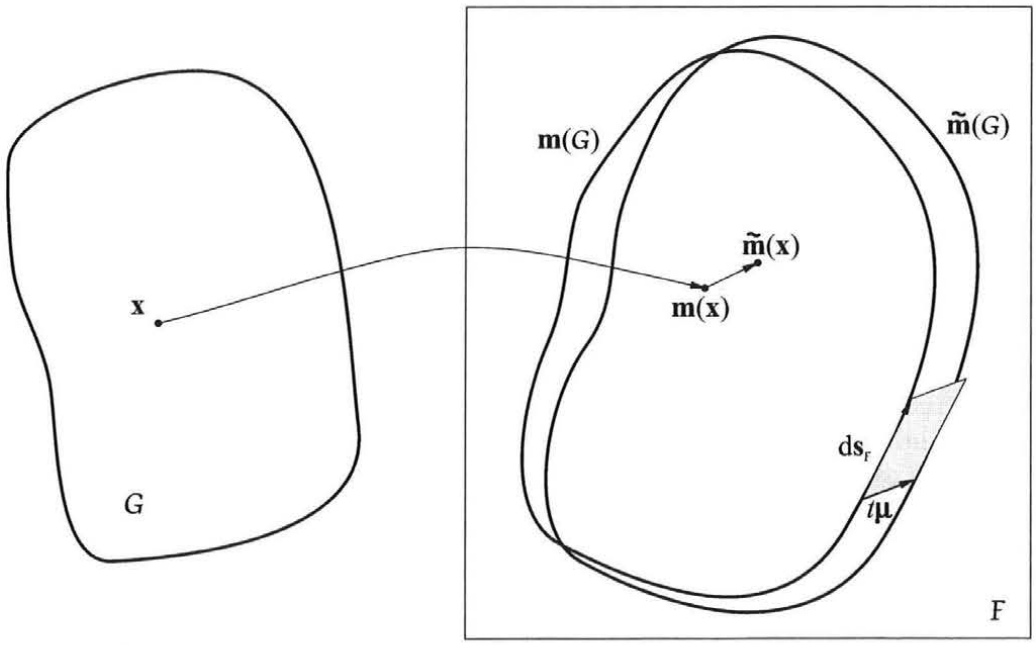

Figure 3.3: The image of domain $\mathcal{G}$ under the mapping $\tilde{\mathbf{m}} = \mathbf{m} + t\boldsymbol{\mu}$.

Furthermore, V may be assumed bounded, thus

$$\lim_{t\to 0}\frac{1}{t}\int_{\tilde{\mathbf{m}}(\mathcal{G})} V\circ(\tilde{\mathbf{m}}^{-1},\mathbf{id}_{\mathcal{F}}) - V\circ(\mathbf{m}^{-1},\mathbf{id}_{\mathcal{F}})\,\mathrm{d}A = \int_{\mathbf{m}(\mathcal{G})}\lim_{t\to 0}\frac{1}{t}\big(V\circ(\tilde{\mathbf{m}}^{-1},\mathbf{id}_{\mathcal{F}}) - V\circ(\mathbf{m}^{-1},\mathbf{id}_{\mathcal{F}})\big)\,\mathrm{d}A \quad (3.2.7)$$

Consequently, the differentiation has to take only the integrand into account. It is however not performed straight forward, as the integrand depends on the inverse map. For every point $\mathbf{y}\in\mathbf{m}(\mathcal{G})$, the change of its corresponding partner in $\mathcal{G}$ caused by the change of the map is needed. Based on simple geometric considerations, a first order approximation of the inverse map $\tilde{\mathbf{m}}^{-1}$ is given by

$$\tilde{\mathbf{m}}^{-1}(\mathbf{y}) \simeq \mathbf{x} - t\big(\mathbf{m}'(\mathbf{x})\big)^{-1}\boldsymbol{\mu}(\mathbf{x})\,. \quad (3.2.8)$$

Plugging this into (3.2.7) leads to

$$-\int_{\mathbf{m}(\mathcal{G})} V_{\mathbf{x}}^{T}\mathbf{m}'^{-1}\boldsymbol{\mu}\,\mathrm{d}A\,. \quad (3.2.9)$$

$V_{\mathbf{x}}$ denotes the column vector $\left(\frac{\partial V}{\partial x_1},\frac{\partial V}{\partial x_2}\right)^T$. We have tacitly not included the proper transformations to the domain $\mathbf{m}(\mathcal{G})$ as they significantly clutter the formula.

(3.2.6b) accounts for the boundary term. The differential element of area $\mathrm{d}A_{\mathcal{F}}$ for the boundary regions $\tilde{\mathbf{m}}(\mathcal{G})\backslash\mathbf{m}(\mathcal{G})$ and $\mathbf{m}(\mathcal{G})\backslash\tilde{\mathbf{m}}(\mathcal{G})$ can be approximated by $\mathrm{d}A_{\mathcal{F}} = t\boldsymbol{\mu}\times\mathrm{ds}_{\mathcal{F}}$, as illustrated in Fig. 3.3. The sign of this expression is correct for regions that are in $\tilde{\mathbf{m}}(\mathcal{G})$ but not in $\mathbf{m}(\mathcal{G})$, which therefore have to be counted positively, as well as for the reverse situation. Computing $\lim_{t\to 0}\frac{1}{t}(\dots)$ for the boundary term (3.2.6b) leads to

$$\int_{\partial\mathbf{m}(\mathcal{G})} V\boldsymbol{\mu}\times\mathrm{ds}\,. \quad (3.2.10)$$

Again a shorthand notation is employed.

(3.2.9) and (3.2.10) are not in a computationally favorable form, since they require integration in the image domain $\mathcal{F}$, whilst we are striving to have all computations done in the model domain $\mathcal{G}$. Accordingly, the next step is to perform variable substitution to change the domain of integration from $\mathbf{m}(\mathcal{G})$ to $\mathcal{G}$. This is the very same as has been done to derive (3.1.12) from (3.1.8). First, note that the inverse Jacobi matrix can be written as[1]

$$\mathbf{m}'^{-1} = \frac{1}{\det \mathbf{m}'} \mathbf{m}'^{\mathrm{ad}} . \tag{3.2.11}$$

Consequently, the Jacobian $\det \mathbf{m}'$ cancels when performing variable substitution for (3.2.9). Explicitly, $\mathbf{m}'^{\mathrm{ad}}$ is given by

$$\mathbf{m}'^{\mathrm{ad}} = \begin{pmatrix} \frac{\partial m_2}{\partial x_2} & -\frac{\partial m_1}{\partial x_2} \\ -\frac{\partial m_2}{\partial x_1} & \frac{\partial m_1}{\partial x_1} \end{pmatrix} . \tag{3.2.12}$$

Variable substitution for the boundary term is performed by means of the relation

$$\mathrm{ds}_{\mathcal{F}} = \mathbf{m}' \, \mathrm{ds}_{\mathcal{G}} . \tag{3.2.13}$$

Further $\boldsymbol{\mu} \times (\mathbf{m}' \, \mathrm{ds}_{\mathcal{G}}) = (\mathbf{m}'^{\mathrm{ad}} \boldsymbol{\mu}) \times \mathrm{ds}_{\mathcal{G}}$, so that stacking all the results together leads to

$$E'(\mathbf{m})\boldsymbol{\mu} = \int_{\partial \mathcal{G}} V \circ (\mathbf{id}_{\mathcal{G}}, \mathbf{m}) \, (\mathbf{m}'^{\mathrm{ad}} \boldsymbol{\mu}) \times \mathrm{ds} - \int_{\mathcal{G}} V_{\mathbf{x}}^T \circ (\mathbf{id}_{\mathcal{G}}, \mathbf{m}) \, \mathbf{m}'^{\mathrm{ad}} \boldsymbol{\mu} \, \mathrm{d}A . \tag{3.2.14}$$

A graphical illustration of the derivative is given for the 1D case (Section 3.2.3).

3.2.2 The Gauss-Newton Term

The assumption that errors of various kinds obey a Gaussian distribution is very common. The corresponding estimation problems lead to minimization problems of quadratic expressions, commonly referred to as least-squares problems. These problems have an advantageous structure with respect to optimization, causing them to be favored even in cases in which the theoretical justification is not stringent. In the particular case the problem is of least-squares type if the local measure V is a quadratic expression,

$$V(\mathbf{x}, \mathbf{y}) = \frac{1}{2} |N(\mathbf{x}, \mathbf{y})|^2 . \tag{3.2.15}$$

The considerations are based on the image space formulation (3.1.8), because with respect to domain $\mathbf{m}(\mathcal{G})$ the energy is a pure quadratic expression. Explicitly it can be written as

$$E(\mathbf{m}) = \frac{1}{2} \int_{\mathbf{m}(\mathcal{G})} |N \circ (\mathbf{m}^{-1}, \mathbf{id}_{\mathcal{F}})|^2 \, \mathrm{d}A \tag{3.2.16}$$

or using the L^2 norm

$$E(\mathbf{m}) = \frac{1}{2} \| N \circ (\mathbf{m}^{-1}, \mathbf{id}_{\mathcal{F}}) \|^2_{2, \mathbf{m}(\mathcal{G})} . \tag{3.2.17}$$

[1] According to [13, p. 508], the matrix elements of A^{ad} at the i'th row and j'th column are the algebraic complements for the i'th column and j'th row (sic!) of the determinant of A. A^{ad} is called (German) "adjungiert" which is ambiguous because this term is usually used for the adjoint matrix. In [63, p. 632] it is called the adjunct matrix. We will not assign it any specific name.

Based on this structure, it is possible to apply the Gauss-Newton algorithm for the optimization, the standard algorithm in adjustment calculus. The procedure is to linearize the inner expression $N \circ (\mathbf{m}^{-1}, \mathbf{id}_{\mathcal{F}})$ with respect to the parameter, $\mathbf{m}$ in the particular case, and use the resulting quadratic problem to approximate the original problem (see also Section 4.2.3). Assuming $\tilde{\mathbf{m}} = \mathbf{m} + \boldsymbol{\mu}$, this leads to

$$E(\tilde{\mathbf{m}}) \simeq E(\mathbf{m}) + E'(\mathbf{m})\boldsymbol{\mu} + \frac{1}{2}E^{\mathrm{GN}}(\mathbf{m})\boldsymbol{\mu}\boldsymbol{\mu} \,. \tag{3.2.18}$$

E^{GN} is a bilinear operator that approximates the second derivative E'' (cf. Section 4.2.3). We employ the same notation as used in [64, pp. 464] for the second derivative and write the arguments of the bilinear form without brackets.

The forward similarity is however not a standard situation. In particular it is unclear how to treat the boundary. There is no simple recipe how to obtain the Gauss-Newton term in this situation. The conclusion of the following arguments is that the boundaries are ignored. The linearization of $N \circ (\mathbf{m}^{-1}, \mathbf{id}_{\mathcal{F}})$ is performed according to

$$N \circ (\tilde{\mathbf{m}}^{-1}, \mathbf{id}_{\mathcal{F}}) \simeq N \circ (\mathbf{m}^{-1}, \mathbf{id}_{\mathcal{F}}) + \begin{cases} N_{\mathbf{x}}^T \mathbf{m}'^{-1} \boldsymbol{\mu} & \mathbf{m}(\mathcal{G}) \cap \tilde{\mathbf{m}}(\mathcal{G}) \\ 0 & \text{else} \end{cases} \,. \tag{3.2.19}$$

In the interior region $\mathbf{m}(\mathcal{G}) \cap \tilde{\mathbf{m}}(\mathcal{G})$ the ordinary linearization is applied. Adding the linear term in the boundary region leads to expressions of higher than quadratic order. This is taken as the reason to omit them in the boundary region. (3.2.19) is inserted into (3.2.16),

$$\begin{aligned} E(\tilde{\mathbf{m}}) \simeq\ & E(\mathbf{m}) \\ & + \int_{\mathbf{m}(\mathcal{G}) \cap \tilde{\mathbf{m}}(\mathcal{G})} N N_{\mathbf{x}}^T \mathbf{m}'^{-1} \boldsymbol{\mu} \, \mathrm{d}A && \text{(3.2.20a)} \\ & + \frac{1}{2} \int_{\tilde{\mathbf{m}}(\mathcal{G}) \setminus \mathbf{m}(\mathcal{G})} |N|^2 \, \mathrm{d}A - \frac{1}{2} \int_{\mathbf{m}(\mathcal{G}) \setminus \tilde{\mathbf{m}}(\mathcal{G})} |N|^2 \, \mathrm{d}A && \text{(3.2.20b)} \\ & + \frac{1}{2} \int_{\mathbf{m}(\mathcal{G}) \cap \tilde{\mathbf{m}}(\mathcal{G})} |N_{\mathbf{x}}^T \mathbf{m}'^{-1} \boldsymbol{\mu}|^2 \, \mathrm{d}A \,. && \text{(3.2.20c)} \end{aligned}$$

The approximation is to be valid for small increments $\boldsymbol{\mu}$. If $\|\boldsymbol{\mu}\| \to 0$ for example for the norm (3.1.21) or if $\boldsymbol{\mu}$ is replaced by $t\boldsymbol{\mu}$ and $t \to 0$ then $\tilde{\mathbf{m}}(\mathcal{G}) \to \mathbf{m}(\mathcal{G})$. Accordingly, for (3.2.20a) and (3.2.20c) the domain of integration is replaced by $\mathbf{m}(\mathcal{G})$. The boundary term (3.2.20b) is treated like in Section 3.2.1. By means of variable substitution, (3.2.20a) and (3.2.20b) lead to the interior and boundary terms of the first derivative, but specialized to the local measure (3.2.15). (3.2.20c) leads to the Gauss-Newton term

$$E^{\mathrm{GN}}(\mathbf{m})\boldsymbol{\mu}\boldsymbol{\mu} = \int_{\mathcal{G}} |N_{\mathbf{x}}^T \circ (\mathbf{id}_{\mathcal{G}}, \mathbf{m}) \, \mathbf{m}'^{\mathrm{ad}} \boldsymbol{\mu}|^2 \frac{1}{\det \mathbf{m}'} \, \mathrm{d}A \,. \tag{3.2.21}$$

The above arguments are not compelling. The essential consequence is that no boundary term is obtained for the Gauss-Newton term. The situation is not entirely satisfactory. The Gauss-Newton algorithm relies on both quantities E' and E^{GN}. Since the boundary term of the derivative has no analogue in the quadratic term, the computed iterative update is unbalanced. Practical experience shows that the performance near the boundary is sometimes not optimal and that omitting the boundary term of E' improves convergence. The method though lacks theoretical justification, as it concerns not only the optimization method but changes also the solution $\mathbf{m}^*$ for which $E'(\mathbf{m}^*) = 0$.

A point to note about the Gauss-Newton term (3.2.21) is the inverse Jacobian in the integrand. In the case of small Jacobian $\det \mathbf{m}'$, i.e. in the neighborhood of an occlusion, the quadratic term given by the Gauss-Newton method gets excessively large. As will be shown in Section 3.5, the forward energy is not differentiable if the Jacobian vanishes, i.e. at the point of transition from visible to occluded. Clearly, if the linear approximation fails, because of lacking differentiability, there is little hope that a quadratic term is meaningful. A possibility to cross validate (3.2.21) is based on the fact that, for vanishing local measure $N \circ (\mathbf{id}_{\mathcal{G}}, \mathbf{m}) = 0$, the Gauss-Newton term should be equal to the second derivative. This is indeed the case. The details are however omitted.

Yet another thing to note about (3.2.21) is that it is a "diagonal" bilinear operator. This is due to the fact that the values of f are assumed independent, if conditioned on the model $(\mathbf{m}, g)$, and that the blurring effect has been neglected. The discretized form of E^{GN} is a diagonal matrix. The diagonal form is however destroyed once the similarity term is combined with a stabilizing functional or if the matching function is composed of basis functions.

3.2.3 Specialization to 1D Constellations

The 1D case provides a means for easy to implement testing routines and facilitates analysis. Therefor the results of the preceding section are given for this special case. The reference domain is assumed an interval $[x_0, x_1]$, thus the energy is given by

$$E(m) = \int_{x_0}^{x_1} V \circ (\mathrm{id}_{\mathcal{F}}, m) m' \, dx \; . \tag{3.2.22}$$

The derivative evaluates to

$$E'(m)\mu = V \circ (\mathrm{id}_{\mathcal{F}}, m)\mu \Big|_{x_0}^{x_1} - \int_{x_0}^{x_1} V_x \circ (\mathrm{id}_{\mathcal{F}}, m)\mu \, dx \; . \tag{3.2.23}$$

and the Gauss-Newton is given by

$$E^{\mathrm{GN}}(m)\mu\mu = \int_{x_0}^{x_1} |N_x \circ (\mathrm{id}_{\mathcal{F}}, m)\mu|^2 \frac{1}{m'} \, \mathrm{d}x \; . \tag{3.2.24}$$

For the 1D case a graphically illustration of the derivative is easily possible. The following example (Fig. 3.4) is, as always, based on the local least-squares measure. The template g and the data f are assumed Gaussian functions. We adopt the viewpoint of optimization and consider the mapping $m^{(k)}$ as the state of the k'th iteration step. The best match, i.e. minimal image energy, is achieved for a mapping m^* that perfectly aligns template and data, $f \circ m^* = g$. To improve the similarity $f \circ m^{(k+1)}$ should be closer to g than $f \circ m^{(k)}$, in particular it should be situated left of $f \circ m^{(k)}$. This requires $m^{(k+1)}(x) > m^{(k)}(x)$, $\forall x$. Gradient descent like optimization method use the negative gradient to compute reasonable directions for minimizing the image energy. The gradient is easily derived from (3.2.23),

$$\nabla E(m) = -V(x_0, m(x_0))\delta(. - x_0) - V_x \circ (\mathrm{id}_{\mathcal{F}}, m) + V(x_1, m(x_1))\delta(. - x_1) \; . \tag{3.2.25}$$

The Dirac functions $\delta(. - x_0)$ and $\delta(. - x_1)$ account for the boundary terms. For the simple example of Fig. 3.4 the interval is assumed to extend to infinity, thus no boundary terms are

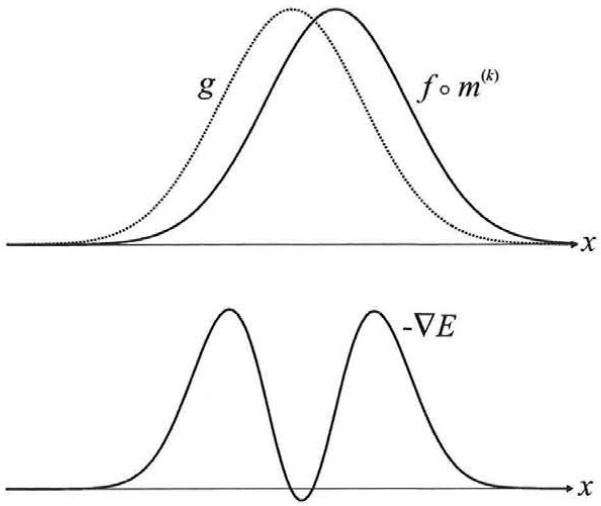

Figure 3.4: The negative gradient $-\nabla E$ for a 1D constellation.

included. Together with the particular local measure this leads to

$$\nabla E(m) = (f \circ m - g)g' \,. \tag{3.2.26}$$

The negative gradient is depicted in the lower part of Fig. 3.4. Clearly $-\nabla E$ points dominantly to the correct direction. The magnitude depends on the slope at the respective position. Regions with steep slopes, the "edges", are not only used for feature-based matching, but also provide the main clue for intensity-based matching algorithms. Outside the central region, where g and $f \circ m^{(k)}$ decay to zero, no matching information is available and the gradient decays to zero, too. A small part, near the maxima of the two functions, is noted where $-\nabla E$ points to the "wrong" direction. It is clear that the image energy can not be used alone to perform successful matching. Its gradient is nevertheless the driving force for a large class of matching algorithms. Additional modeling assumptions (cf. Section 4.1) are used to prevent each value $m(x)$ from proceeding independently of its neighbor values. These mechanisms grant that m keeps in a reasonable shape throughout an iterative optimization procedure.

3.2.4 Derivatives of the Reverse Modeling Similarity Measure

Calculating the derivative of the forward energy requires some intermediate steps. For the reverse energy (3.1.9) the situation is much simpler because the integrand depends on $\mathbf{m}$ only and contains no derivatives of $\mathbf{m}$. The first derivative is given by

$$E^{r\prime}(\mathbf{m})\boldsymbol{\mu} = \int_{\mathcal{G}} V_{\mathbf{y}}^T \circ (\mathbf{id}_{\mathcal{G}}, \mathbf{m})\, \boldsymbol{\mu}\, \mathrm{d}A \,. \tag{3.2.27}$$

The Gauss-Newton term evaluates to

$$E^{r\mathrm{GN}}(\mathbf{m})\boldsymbol{\mu}\boldsymbol{\mu} = \int_{\mathcal{G}} |N_{\mathbf{y}}^T \circ (\mathbf{id}_{\mathcal{G}}, \mathbf{m})\, \boldsymbol{\mu}|^2 \, \mathrm{d}A \,. \tag{3.2.28}$$

Comparing these results with those obtained for the forward similarity reveals the following differences. (3.2.14) includes a boundary term due to the fact that the modeled part of the image changes. This happens because the reference domain, which parameterizes the

part of the model that is to be recovered, is kept fixed. The relevant part of $\mathcal{F}$, the modeled part $\mathcal{F}^M$, is not known a priori. For the reverse similarity the roles of $\mathcal{F}$ and $\mathcal{G}$ are reverted. The "data" domain $\mathcal{G}$ is fixed and no boundary term results. The derivative $V_{\mathbf{x}}^T \mathbf{m}'^{\mathrm{ad}} \boldsymbol{\mu}$ in (3.2.14) translates to $V_{\mathbf{y}}^T \boldsymbol{\mu}$ in (3.2.27). The relation of the Gauss-Newton terms (3.2.21) and (3.2.28) is similar. For the case of the reverse similarity measure the differentiation of the local measure, V or N respectively, is performed in the image domain, with respect to $\mathbf{y}$, for the forward similarity the differentiation is done in the reference domain, with respect to $\mathbf{x}$. Note that the expression for the forward similarity can not be of the form $V_{\mathbf{x}}^T \boldsymbol{\mu}$. The terms $V_{\mathbf{x}}$ and $\boldsymbol{\mu}$ do not match because $\mathbf{x}$ and $\boldsymbol{\mu}$ are quantities of different domains specified with respect to different coordinate systems. The matrix $\mathbf{m}'^{\mathrm{ad}}$ performs the necessary translation.

3.3 Examples of Particular Geometric Constellations

So far the similarity measure has been considered without restricting to a particular geometry. For every point $\mathbf{x} \in \mathcal{G}$, both coordinates of the map, $m_1(\mathbf{x})$ and $m_2(\mathbf{x})$, have been considered as free quantities. In the course of surface reconstruction, this flexibility is not needed. Usually $\mathbf{m}$ does not serve as the matching parameter itself, but depends on some scalar-valued function. The parameterization of the map $\mathbf{m}$ provides a means to incorporate knowledge about the geometry, the orientation, of the imaging constellation. In a generic setting the matching function will be denoted by u. It extends the pair (x_1, x_2) to the triple $(x_1, x_2, u(\mathbf{x}))$, which encodes the position of the corresponding point in the 3D object space. The geometric relationships considered in this section are always local, i.e. $\mathbf{m}(\mathbf{x}) := \mathbf{m}(\mathbf{x}, u(\mathbf{x}))$. In that case it is still easy to use the results of the previous section. Much the same way as when differentiating composite functions according to the chain rule, the increments $\boldsymbol{\mu}$ in the expressions for E' and E^{GN} depend on the increment of u denoted by v, $\boldsymbol{\mu}(\mathbf{x}) \simeq \frac{\partial \mathbf{m}(\mathbf{x})}{\partial u(\mathbf{x})} v(\mathbf{x})$. Consequently, when considering particular matching configurations not $\mathbf{m}'^{\mathrm{ad}}$ is needed but $\mathbf{m}'^{\mathrm{ad}} \frac{\partial \mathbf{m}}{\partial u}$. Care has to be taken, since for the general case this applies to the first derivative E' only. When translating higher order derivatives E'' etc., and the dependency of $\mathbf{m}$ on u is not linear, mixed terms will appear. From the current perspective, the use of higher order terms has minor relevancy. It is therefore not considered separately. The Gauss-Newton term E^{GN} is based solely on first order derivatives and the principle can be applied unaltered. A straightforward calculation of $\mathbf{m}'^{\mathrm{ad}} \frac{\partial \mathbf{m}}{\partial u}$ is rather laborious for some cases. Taking the special structure into account simplifies the matter significantly,

$$\begin{aligned} \mathbf{m}'^{\mathrm{ad}} \frac{\partial \mathbf{m}}{\partial u} &= \begin{pmatrix} \frac{\partial m_2}{\partial x_2} + \frac{\partial m_2}{\partial u}\frac{\partial u}{\partial x_2} & -\frac{\partial m_1}{\partial x_2} - \frac{\partial m_1}{\partial u}\frac{\partial u}{\partial x_2} \\ -\frac{\partial m_2}{\partial x_1} - \frac{\partial m_2}{\partial u}\frac{\partial u}{\partial x_1} & \frac{\partial m_1}{\partial x_1} + \frac{\partial m_1}{\partial u}\frac{\partial u}{\partial x_1} \end{pmatrix} \begin{pmatrix} \frac{\partial m_1}{\partial u} \\ \frac{\partial m_2}{\partial u} \end{pmatrix} \\ &= \begin{pmatrix} -\begin{vmatrix} \frac{\partial m_1}{\partial x_2} & \frac{\partial m_1}{\partial u} \\ \frac{\partial m_2}{\partial x_2} & \frac{\partial m_2}{\partial u} \end{vmatrix} & \begin{vmatrix} \frac{\partial m_1}{\partial x_1} & \frac{\partial m_1}{\partial u} \\ \frac{\partial m_2}{\partial x_1} & \frac{\partial m_2}{\partial u} \end{vmatrix} \end{pmatrix}^T \end{aligned} \tag{3.3.1}$$

Most importantly, the quantity depends on the "location" $(\mathbf{x}, u(\mathbf{x}))$ only. All terms containing derivatives of u cancel independent of the type of mapping, as long as the map is parameterized by a scalar function. As can be seen from (3.2.14), the components of the vector (3.3.1) enter the formulation of the similarity measure as coefficients of partial derivatives ∂_{x_1} and ∂_{x_2} respectively. Hence, the above vector is essentially used to perform derivation in a particular direction. For the common case of perspective imagery, the search for corresponding points can be confined to epipolar lines, nadir lines, or vertical locus lines. $\mathbf{m}'^{\mathrm{ad}} \frac{\partial \mathbf{m}}{\partial u}$ is a vector tangential to these lines, thus the derivation is performed in the direction of these lines. Note

however that the derivative is taken in the reference domain rather than the image domain. We denote the components of (3.3.1) by $-p_1$ and $-p_2$. The minus signs are inserted to be compatible with (3.3.4), it is however a matter of subjective taste how to include them. The symbol $\tilde{\mathrm{d}}$ is used to denote the derivative

$$\tilde{\mathrm{d}}_u g = g_{\mathbf{x}}^T \mathbf{m}'^{\mathrm{ad}} \frac{\partial \mathbf{m}}{\partial u} = -p_1 g_{x_1} - p_2 g_{x_2} \,. \tag{3.3.2}$$

The other quantity frequently used in the course of this work is the Jacobian or area distortion $\det \mathbf{m}'$. The following identity reveals the natural connection between the Jacobian and the derivative $\tilde{\mathrm{d}}_u$,

$$\det \mathbf{m}' = -p_1 u_{x_1} - p_2 u_{x_2} + \begin{vmatrix} \frac{\partial m_1}{\partial x_1} & \frac{\partial m_1}{\partial x_2} \\ \frac{\partial m_2}{\partial x_1} & \frac{\partial m_2}{\partial x_2} \end{vmatrix} \,. \tag{3.3.3}$$

If the Jacobian and the derivative $\tilde{\mathrm{d}}_u$ are expressed in a compatible way, computational efficiency can be gained by factoring out common parts. The connection can also be employed to find suitable expressions for (p_1, p_2). It is much more likely to find formulas for the local area distortion $\det \mathbf{m}'$ of a particular mapping $\mathbf{m}$ in the literature than for the coefficients (p_1, p_2). The coefficients p_i can be deduced directly from any formulation of the area distortion. To fully uncover the regularity of (3.3.3), it will be written in yet another way,

$$\det \mathbf{m}' = \begin{vmatrix} \frac{\partial m_1}{\partial x_1} & \frac{\partial m_1}{\partial x_2} & \frac{\partial m_1}{\partial u} \\ \frac{\partial m_2}{\partial x_1} & \frac{\partial m_2}{\partial x_2} & \frac{\partial m_2}{\partial u} \\ -\frac{\partial u}{\partial x_1} & -\frac{\partial u}{\partial x_2} & 1 \end{vmatrix} \,. \tag{3.3.4}$$

It appears that p_1 and p_2 are the algebraic complements of the third row of the above matrix. The third component of this triple of algebraic complements is the determinant in (3.3.3). The pair (p_1, p_2) may thus be extended to the triple (p_1, p_2, p_3) with p_3 given in (3.3.3).

Using a scalar matching function u in combination with the reverse modeling similarity measure follows the same principles, i.e. $\boldsymbol{\mu}$ is again replaced by $\frac{\partial \mathbf{m}}{\partial u} v$. The results are obtained easily. The equivalent to the reference space derivative $\tilde{\mathrm{d}}_u$ is simply

$$\frac{\partial f \circ \mathbf{m}}{\partial u} = f_{\mathbf{y}}^T \frac{\partial \mathbf{m}}{\partial u} = \frac{\partial m_1}{\partial u} f_{y_1} + \frac{\partial m_2}{\partial u} f_{y_2} \,. \tag{3.3.5}$$

Obviously this derivative is evaluated in image space. The extra effort that has to be paid using the forward or reverse modeling depends on the type of parameterization. For the DLT configuration (cf. Section 3.3.3) the difference concerns the paper work, but not the computational complexity.

In the sequel the following sample configurations are considered,

- image-to-image mapping for normal case constellations, using the *ordinary disparity* as matching function,
- image-to-image mapping for general constellations, using the *generalized disparity* as matching function, and
- orthophoto-to-image mapping, using the *height* $Z(X, Y)$ as matching function.

The purpose of this section is primarily to serve as reference for the implementation. This is necessary mainly because the quantity $\mathbf{m}'^{\mathrm{ad}}$ and combinations including it are unusual and most likely will be hard to find in the literature. As a side effect, the particular simple matching by disparity may be viewed as an example facilitating the understanding of the method.

A remark on notation: It is common practice to denote coordinates in object space by (X, Y, Z). Similarly for image coordinates some choices are more usual then others. Despite of that, we stay with the convention to denote the coordinates of the target domain by $\mathbf{y}$ and the coordinates of the source domain by $\mathbf{x}$. When comparing different schemes in succession, that choice has been found to be more transparent.

3.3.1 Normal Case Image-to-Image Mapping by Disparity

The normal case arrangement is surely the most often used configuration for matching algorithms (e.g. [5, 10, 18, 19, 26, 37, 51]). It can be handled easily due to the fact that the epipolar lines are parallel. Given an image point $\mathbf{x}$ the corresponding point $\mathbf{y} = \mathbf{m}(\mathbf{x})$ in the other image is determined by the disparity d,

$$\mathbf{m}(\mathbf{x}) = \begin{pmatrix} x_1 + d(\mathbf{x}) \\ x_2 \end{pmatrix} . \tag{3.3.6}$$

We follow the common practice to have the x_1- and y_1-axis parallel to the epipolar lines. The structural properties mentioned above are not necessary, as the results are obtained without complication anyway. The Jacobi matrix is given by

$$\mathbf{m}' = \begin{pmatrix} 1 + d_{x_1} & d_{x_2} \\ 0 & 1 \end{pmatrix} . \tag{3.3.7}$$

Noting that $\boldsymbol{\mu} = (\delta, 0)^T$, because there is no increment of the second coordinate, we have

$$\mathbf{m}'^{\mathrm{ad}} \frac{\partial \mathbf{m}}{\partial d} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} . \tag{3.3.8}$$

The Jacobian is almost trivially given by

$$\det \mathbf{m}' = 1 + d_{x_1} . \tag{3.3.9}$$

This corresponds to a coefficient vector $\mathbf{p} = (-1, 0, 1)$.

Plugging the results into the first derivative of the similarity measure (3.2.14) results in

$$E'(d)\delta = \int_{\partial \mathcal{G}} V \delta \, dx_2 - \int_{\mathcal{G}} V_{x_1} \delta \, dA . \tag{3.3.10}$$

Combining this with the local least-squares measure yields

$$E'(d)\delta = \frac{1}{2} \int_{\partial \mathcal{G}} |f \circ m - g|^2 \delta \, dx_2 + \int_{\mathcal{G}} (f \circ m - g) g_{x_1} \delta \, dA . \tag{3.3.11}$$

The close relation to the 1D case (cf. Section 3.2.3) is obvious.

3.3.2 Image-to-Image Mapping by Generalized Disparity

The normal case constellation is particularly simple for the application of image matching algorithms. Unfortunately, real imaging configurations seldom conform to the rather stringent requirements. Having only two images, it is possible to artificially generate a pair of normal case images. Nevertheless, such preprocessing steps are not satisfactory though sometimes inevitable. Another possibility is to use a generalized concept of disparity [39]. We assume that the two cameras C and C' are specified by the matrices of intrinsic camera parameters $\mathbf{C}$ and $\mathbf{C}'$, as defined by

$$\mathbf{C} := \begin{pmatrix} 1 & 0 & -h_1 \\ 0 & 1 & -h_2 \\ 0 & 0 & -c \end{pmatrix} . \tag{3.3.12}$$

(h_1, h_2) are the coordinates of the cardinal point and c is the camera constant. The rotation matrices $\mathbf{R}$ and $\mathbf{R}'$ are given with respect to some reference coordinate system, which may coincide with one of the camera coordinates systems to cover the case of pure relative orientation. A vector $\mathbf{n}$ has to be specified that indicates the direction in which the depth z relative to the projection center of C is to be measured. The homogeneous coordinates $\mathbf{x}_h$ and $\mathbf{y}_h$ of corresponding image points are related by [39]

$$\mathbf{y}_h \propto (\mathbf{R}'\mathbf{C}')^{-1}(\mathbf{I} - \frac{1}{z(\mathbf{x})}\mathbf{b}\mathbf{n}^T)\mathbf{R}\mathbf{C}\mathbf{x}_h . \tag{3.3.13}$$

$\mathbf{b}$ denotes the basis vector from camera C to camera C' and $\mathbf{I}$ is the 3×3 identity matrix. The symbol $\propto$ indicates proportionality by a factor. We define the generalized disparity d to be inversely proportional to the depth z. The proportionality factor may be chosen according to convenience. Setting $d = \frac{c|\mathbf{b}|}{z}$ is conform with the ordinary disparity. (3.3.13) can be rewritten in condensed form

$$\mathbf{y}_h \propto \mathbf{H}(\mathbf{x})\mathbf{x}_h = (\mathbf{V} - d(\mathbf{x})\mathbf{K})\mathbf{x}_h . \tag{3.3.14}$$

$\mathbf{H}(\mathbf{x})$, $\mathbf{V}$ and $\mathbf{K}$ are each 3×3 matrices. From (3.3.13), it follows that $\mathbf{K}$ has rank one. For $d \to \infty$ or equivalently depth $z = 0$, the dominant term is $-d\mathbf{K}\mathbf{x}_h$. Due to $\text{rank}(\mathbf{K}) = 1$ and since homogeneous coordinates may be scaled arbitrarily, all values $-d\mathbf{K}\mathbf{x}_h$ correspond to the same point. This point is the epipolar point of camera C'. $\mathbf{y}_h \propto \mathbf{V}\mathbf{x}_h$ is obtained if $z \to \infty$ or $d = 0$. The corresponding point $\mathbf{y}$ is the vanishing point of the ray associated with $\mathbf{x}$. Using the abbreviations

$$\begin{pmatrix} v_1(\mathbf{x}) \\ v_2(\mathbf{x}) \\ v_3(\mathbf{x}) \end{pmatrix} = \mathbf{V} \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix} = \begin{pmatrix} V_{11}x_1 + V_{12}x_2 + V_{13} \\ V_{21}x_1 + V_{22}x_2 + V_{23} \\ V_{31}x_1 + V_{32}x_2 + V_{33} \end{pmatrix} , \tag{3.3.15}$$

$k_i(\mathbf{x})$ defined analogously, and $h_i = v_i - dk_i$ the map $\mathbf{m}$ is equal to

$$\mathbf{m}(\mathbf{x}) = \begin{pmatrix} \frac{h_1(\mathbf{x})}{h_3(\mathbf{x})} \\ \frac{h_2(\mathbf{x})}{h_3(\mathbf{x})} \end{pmatrix} = \begin{pmatrix} \frac{v_1(\mathbf{x}) - d(\mathbf{x})k_1(\mathbf{x})}{v_3(\mathbf{x}) - d(\mathbf{x})k_3(\mathbf{x})} \\ \frac{v_2(\mathbf{x}) - d(\mathbf{x})k_2(\mathbf{x})}{v_3(\mathbf{x}) - d(\mathbf{x})k_3(\mathbf{x})} \end{pmatrix} . \tag{3.3.16}$$

The coefficient vector $\mathbf{p}$ is evaluated to

$$\mathbf{p} = \frac{1}{h_3^3} \begin{pmatrix} \begin{vmatrix} k_1 & H_{12} & h_1 \\ k_2 & H_{22} & h_2 \\ k_3 & H_{32} & h_3 \end{vmatrix} \begin{vmatrix} H_{11} & k_1 & h_1 \\ H_{21} & k_2 & h_2 \\ H_{31} & k_3 & h_3 \end{vmatrix} \begin{vmatrix} H_{11} & H_{12} & h_1 \\ H_{21} & H_{22} & h_2 \\ H_{31} & H_{32} & h_3 \end{vmatrix} \end{pmatrix} . \tag{3.3.17}$$

$h_i(\mathbf{x})$ and $k_i(\mathbf{x})$ are a byproduct of the evaluation of $\mathbf{m}(\mathbf{x})$. The chosen notation emphasizes the regularity of the expressions. Computational efficiency is increased if $h_3(\mathbf{x})$ and $k_3(\mathbf{x})$ are factored out and if certain sub-determinants are calculated in advance. Recall that $\frac{k_1(\mathbf{x})}{k_3(\mathbf{x})} = \text{const}$ and $\frac{k_2(\mathbf{x})}{k_3(\mathbf{x})} = \text{const}\ \forall \mathbf{x}$. The case that the depth $z(\mathbf{x})$ is chosen as the matching parameter can be handle very similarly, because the structure of (3.3.14) and (3.3.16) is retained if d is replaced by $\frac{1}{z}$.

Despite all attempts to calculate (3.3.17) efficiently, a large number of additions and multiplications remains to be performed. A significant reduction of the computational burden is achieved if the vector $\mathbf{n}$ coincides with the axis of the reference camera C. The advantage stems from the fact that in this case, the matrix $\mathbf{K}$ contains nonzero elements in the third column only. Consequently, the map $\mathbf{m}$ can be written in a DLT style,

$$\mathbf{m}(\mathbf{x}) = \begin{pmatrix} \frac{V_{11}x_1+V_{12}x_2-K_{13}d(\mathbf{x})+V_{13}}{V_{31}x_1+V_{32}x_2-K_{33}d(\mathbf{x})+V_{33}} \\ \frac{V_{21}x_1+V_{22}x_2-K_{23}d(\mathbf{x})+V_{23}}{V_{31}x_1+V_{32}x_2-K_{33}d(\mathbf{x})+V_{33}} \end{pmatrix} . \tag{3.3.18}$$

This type of map is considered in the next section for the case of orthophoto-to-image mapping. The results directly apply to the mentioned special case. It is interesting to note that using a generalized concept of disparity leads to equations familiar from object-space-to-image mapping. Naively one would expect that this, if possible at all, could be achieved upon utilizing the depth z as the third coordinate. As it seems, using depth as matching parameter an analogue to (3.3.18) can not be found.

3.3.3 Orthophoto-to-Image Mapping by DLT

The case that the model domain is the XY-plane of some 3D object coordinate system is rather important, especially when considering multi-image matching. The geometric relationships are well known and available from many textbooks [31, 32]. We may directly start by writing the imaging equations in a DLT style,

$$\mathbf{y}_h \propto \mathbf{A}\mathbf{X}_h . \tag{3.3.19}$$

The matrix $\mathbf{A}$ has 3×4 entries. $\mathbf{X}_h$ is a 4 element vector, $\mathbf{X}_h \propto (x_1, x_2, Z(\mathbf{x}), 1)$. Explicitly the map reads

$$\mathbf{m}(\mathbf{x}) = \begin{pmatrix} \frac{A_{11}x_1+A_{12}x_2+A_{13}Z(\mathbf{x})+A_{14}}{A_{31}x_1+A_{32}x_2+A_{33}Z(\mathbf{x})+A_{34}} \\ \frac{A_{21}x_1+A_{22}x_2+A_{23}Z(\mathbf{x})+A_{24}}{A_{31}x_1+A_{32}x_2+A_{33}Z(\mathbf{x})+A_{34}} \end{pmatrix} . \tag{3.3.20}$$

The coefficient vector $\mathbf{p}$ evaluates to

$$\mathbf{p} = \frac{1}{a_3^3} \begin{pmatrix} \begin{vmatrix} a_1 & A_{12} & A_{13} \\ a_2 & A_{22} & A_{23} \\ a_3 & A_{32} & A_{33} \end{vmatrix} & \begin{vmatrix} A_{11} & a_1 & A_{13} \\ A_{21} & a_2 & A_{23} \\ A_{31} & a_3 & A_{33} \end{vmatrix} & \begin{vmatrix} A_{11} & A_{12} & a_1 \\ A_{21} & A_{22} & a_2 \\ A_{31} & A_{32} & a_3 \end{vmatrix} \end{pmatrix} . \tag{3.3.21}$$

The a_i are defined similar to v_i and k_i in the previous section, $a_i = A_{i1}x_1 + A_{i2}x_2 + A_{i3}Z(\mathbf{x}) + A_{i4}$. For the implementation, it is advantageous to evaluate certain sub-determinants in advance. Let α_i be the sub-determinant of $\mathbf{A}$, calculated by omitting the i'th column and additionally including a factor $(-1)^i$, e.g. $\alpha_1 = (-1)^1 | \ A_{*,2} \ \ A_{*,3} \ \ A_{*,4} \ |$, then (3.3.21) can be written as

$$\mathbf{p} = \frac{1}{a_3^3} \begin{pmatrix} x_1\alpha_4 - \alpha_1 & x_2\alpha_4 - \alpha_2 & Z\alpha_4 - \alpha_3 \end{pmatrix} . \tag{3.3.22}$$

The α_i are characteristic quantities for the map $\mathbf{m}$ that can be precomputed. A computational effort like for the evaluation of (3.3.22) has to be expected for any comparable method. It seems unlikely that it can be reduced any further.

3.4 Discretization of the Similarity Measure and its Derivatives

The variational problem of finding an optimal matching function, be it the mapping $\mathbf{m}$ between two images, the disparity d, the surface height $Z(X,Y)$ etc., is a problem with infinitely many degrees of freedom. Solving it on a computer requires a restriction to finite dimensional approximations, with the aim of converting the problem into an optimization problem in $\mathbb{R}^n$. Two basic choices to come up with a finite dimensional model are considered,

- replacing integrals by sums and derivatives by differences utilizing only samples of the underlying functions (*Finite-Sums* approach), or
- using finite dimensional approximations for the functions and performing integration and differentiation analytically(*Finite-Elements, Ritz* approach).

These steps are carried out at the finest modeling level, essentially treating the problem at pixel level. This might be misunderstood as being contrary to other approaches. In [22, 23, 58] the matching function is the surface height $Z(X,Y)$. It is composed of quadratic bilinear Finite-Elements. These elements cover a couple of surface texture cells, which are defined on a finer grid. A typical order of magnitude for the number of contained cells is 8×8. The support of this surface model elements can not be compared with the granularity of the discretization. The formulation of the mentioned approaches starts from a finite number of points. Hence the problem of discretization never occurs. For the presented continuous formulation, the problem to be solved first is the conversion to a problem in $\mathbb{R}^n$ with some possible very large n. Issues are the interrelation between the discretized and the continuous formulation and the resulting error of approximation. Afterwards, as part of a strategic decision, the problem dimension can be reduced further by using $p < n$ basis functions in $\mathbb{R}^n$.

The formulations obtained by the two discretization methods differ in the required style of implementation and in the mathematical properties. The Finite-Elements method uses parameterized approximations of functions to perform integration and differentiation analytically. As it turns out, this approach has to be discarded for 2D constellations since it requires an extreme computational effort. The Finite-Sums approach assumes only samples of functions to be given, utilizing approximation rules for carrying out numerical quadrature and differentiation. These rules are themselves based on assumptions about the functions in that many rules in numerical analysis are based on considerations about polynomials. The characteristic of this method is nevertheless completely different. One consequence of the Finite-Sums approach is that, if the image energy E and its derivative E' are discretized independently, the discrete formulation of E' is not necessarily the derivative of the discrete formulation of E. From principles, this situation can not be encountered if the Finite-Elements approach is used. Of course, the discretized similarity measure is amenable to differentiation. In case of the forward modeling similarity measure, the result however lacks structure. The section on the Finite-Sums approach will establish the relationship between the discretized energy E its derivative and the approximation of E'. First, the Finite-Elements or Ritz approach is considered.

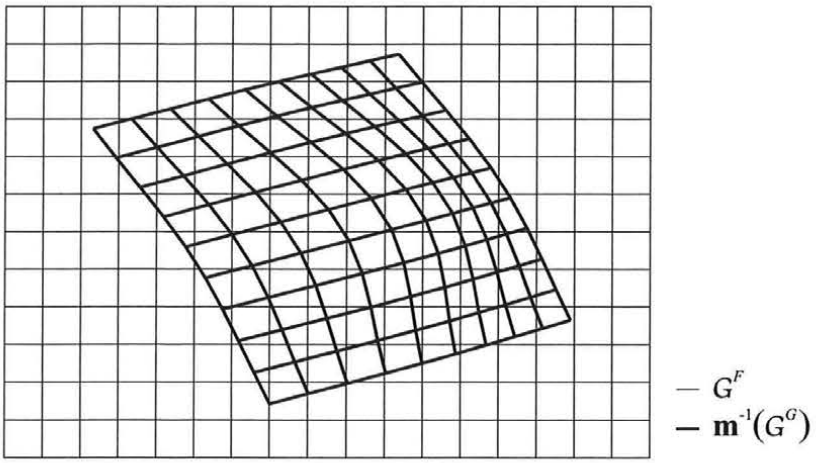

Figure 3.5: Cell structure for an superposition of an irregular gird over an regular grid.

3.4.1 Finite-Elements, Ritz Approach

The Ritz method [64, p. 411] is applicable to variational problems and to differential equations that can be converted to variational formulations. The problem

$$F(u) = \min! \quad u \in M \,, \tag{3.4.1}$$

with $M \subseteq X$ and X an infinite dimensional Banach space, is solved for a finite dimensional subspace X_n,

$$F(u_n) = \min! \quad u_n \in X_n \cup M \,. \tag{3.4.2}$$

Choosing a basis $e_1, \ldots, e_n$ in X_n and defining $u_n = \sum_{i=1}^{n} c_i e_i$ yields an optimization problem in $\mathbb{R}^n$ for the coefficients $c_1, \ldots, c_n$. The Finite-Elements method may be interpreted as Ritz method with special narrow supported functions e_i.[2]

The tractability of this approach relies on the fact that the functional (3.4.2) or the derivative with respect to the coefficients

$$\frac{\partial F(u_n)}{\partial c_i} = F'(u_n) e_i \tag{3.4.3}$$

is computable. For the matching problem this is not easily accomplished. Just consider the simple case that all functions are interpolated from samples at regular grid locations with some narrow supported interpolation kernel. For example, let the two functions f and g be constructed by bilinear interpolation from samples on grids $\mathcal{G}^F$ and $\mathcal{G}^G$ respectively. Assume that $\mathbf{m}$ is specified by its values on grid $\mathcal{G}^G$ also using bilinear interpolation between the grid points. $\mathbf{m}$ will map the quadratic grid cells of $\mathcal{G}^G$ to quadrangles in $\mathcal{F}$. Overlaying the irregular grid with the regular grid $\mathcal{G}^F$ results in a very complicate structure of grid cells, see Fig. 3.5. Some grid cells will contain multiple cells of the other grid, other cells will overlap only partially. The expressions for the functions f and g have local validity only and thus every sub region bounded by vertices of $\mathcal{G}^F$ and $\mathcal{G}^G$ has to be analyzed and treated separately. Consequently, the integration involves considerable effort.

The situation appears different for the 1D case, Fig. 3.6. Overlaying the irregular 1D

[2]The usage of Finite-Elements is not limited to the Ritz-Method. They are also used with the Galerkin-Method [63, pp. 1169], which is applicable to differential equations for which the exact variational formulation is not known. A weighted residual formulation is used instead.

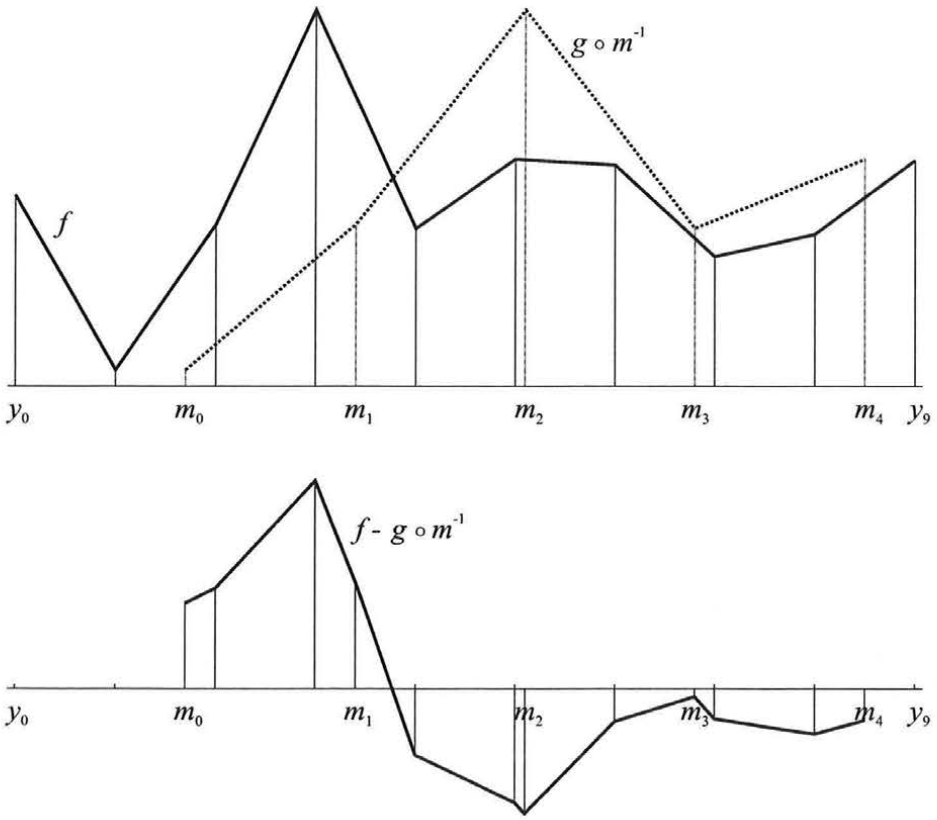

Figure 3.6: A 1D matching configuration using piecewise linear Finite-Elements for f, g, and m.

grid of points m_i over the regular, i.e. equidistant, 1D grid of points y_i yields a succession of intervals with varying length. This structure can be analyzed with reasonable effort. Considering only samples within the model scope $[m_0, m_4]$, the series of interval boundaries is given by $(m_0, y_2, y_3, m_1, \ldots, y_8, m_4)$ for the example. A piecewise linear interpolation is assumed for f, g, and m. Fig. 3.6 also depicts the difference $f - g \circ m^{-1}$. It is easily seen that every interval has to be treated separately. Interestingly, for approaches of Ritz type the image space formulation of the similarity measure plays an important role. The map $m(x)$ is not used to "pull back" the function f by $f \circ m$, but to "push forward" the function g to the domain $\mathcal{F}$ by $g \circ m^{-1}$.[3] It is also possible to use other than piecewise linear interpolation kernels, e.g. with increased smoothness, for any of the involved functions. The effort however rises significantly. For the map m, using anything but piecewise linear interpolation is not recommendable. Recall that the scope of the current considerations is the pixel level. It is doubtful if it makes sense to allow distortions of individual pixels other than linear.

According to the above, the usability of Finite-Elements for the discretization of the similarity measure is limited. The method is applicable for the 1D case and thus can be used for matching normal case stereo images by plugging several 1D constellations in parallel. The advantage with respect to the Finite-Sums approach is that the integrals are computed exactly. For the Finite-Sums approach only a finite set of sampling points of the reference domain $\mathcal{G}$ is taken into account. This bears the danger that, if the m' is very large, the function $f \circ m$ is sampled with insufficient density. In the discrete case a region of the image domain $\mathcal{F}$ can not only be unmodeled, but also more or less badly modeled. For the Finite-Elements approach, a large Jacobian will cause a single basis element to cover a large part of $\mathcal{F}$. This reduces the flexibility of the model, but contrary to the Finite-Sums approach all values of f are taken into account and no undersampling can happen. When working with Finite-Elements, the approximation only concerns the representation of the functions, leading to a better predictability of the performance. It thus can be used for comparison with other approaches.

3.4.2 The Finite-Sums Approach

Under the term *Finite-Sums* approach a method is considered that replaces the integral for the global similarity measure by a sum over a finite set of points employing numerical quadrature. Derivatives are approximated by finite differences. Both steps follow standard numerical methods [24, 59, 63]. Commonly the performance of these methods is studied with respect to the difference between the exact value and the numerical result. We are however not interested in the precise value of the energy functional. The critical point is that the computed energy $E(\mathbf{m})$ remains a good measure for the global similarity of the intensity values and thus favors the corresponding mappings $\mathbf{m}$.

The setup for the formulation is the following. It is assumed that it is possible to compute values of f at any desired location. Practically, f will be given by samples on a grid and an appropriate interpolation kernel has to be chosen. Interesting aspects of the functional description of f are the spectral properties, the number of continuous derivatives, and the effort of computing single function values $f(\mathbf{y})$. Contrary to the Ritz approach the analytical integration is avoided and accordingly the details of the shape of the function f do not affect the formulation. The values of $\mathbf{m}$ are assumed to be given on a grid $\mathcal{G}_h$ with isotropic spacing

[3] On the notions pull back and push forward for tensor fields see [64, pp. 596].

h. For the 2D case it is given by

$$\mathcal{G}_h = \left\{ \mathbf{x}_{k,l} = \begin{pmatrix} o_1 + kh \\ o_2 + lh \end{pmatrix} : (k,l) \in I \right\} , \tag{3.4.4}$$

with the set of indices

$$I = \{(k,l) : 0 \leq k \leq M, 0 \leq k \leq N, k,l \in \mathbb{Z}\} . \tag{3.4.5}$$

The specialization to 1D is obvious. If it is necessary to use single coordinates of a grid point $\mathbf{x}_{k,l}$, superscripts denote the coordinate indices $x^1_{k,l}$, $x^2_{k,l}$ to avoid confusion with the grid index. To reduce the number of subscripts, vector-valued indices $\mathbf{i} \in I$ are preferably used. Some formulas however necessitate the use of the standard double subscripting. The node values are denoted by

$$\mathbf{m_k} = \mathbf{m}(\mathbf{x_k}) . \tag{3.4.6}$$

The vector of grid samples is denoted by $\mathbf{m}_I$. Assuming only isotropic spacing does not limit the generality of the considerations, but the grid setup itself is a restriction for some cases. If the model $(\mathbf{m}, g)$ is to represent a 3D object with full flexibility, it might be appropriate to have it defined on a triangulated set of points. No practical experience in that direction is available in the course of this work, hence the restriction to the grid setup.

The values for the function g can be specified either on the same grid $\mathcal{G}_h$ or on the dual grid

$$\mathcal{G}^*_h = \left\{ \mathbf{x}_{k,l} = \begin{pmatrix} o_1 + kh \\ o_2 + lh \end{pmatrix} : (k,l) \in I^* \right\} , \tag{3.4.7}$$

with the set of indices I^* given by

$$I^* = \left\{ \left(k + \frac{1}{2}, l + \frac{1}{2} \right) : 0 \leq k \leq M-1, 0 \leq k \leq N-1, k,l \in \mathbb{Z} \right\} . \tag{3.4.8}$$

The dual grid nodes correspond to the centers of the square faces formed by the edges of the primary grid $\mathcal{G}_h$. According to whether one or two grids are used, we talk about single grid discretization or double grid discretization. Using two grids is advantageous when approximating derivatives or related quantities like the Jacobian $\det \mathbf{m}'$. For example consider the estimation of the first derivative from samples at a 1D grid. The simplest method is to use the symmetric difference

$$\begin{aligned} g'(x_k) &= \frac{1}{2h}(g_{k+1} - g_{k-1}) + R \\ |R| &\leq \frac{h^2}{6}|g'''(x_k + \vartheta)|, \ -h < \vartheta < h . \end{aligned} \tag{3.4.9}$$

The error of this approximation depends on the grid spacing h, but using finer grids is not really an option because that would increase the size of the data to be processed, which is large anyway. Using two mutually displaced grids provides an improvement, as most derivatives are needed at the respective other grid. The approximation of the derivative between grid positions, i.e. at the dual grid, is done by

$$\begin{aligned} g'(x_{k+\frac{1}{2}}) &= \frac{1}{h}(g_{k+1} - g_k) + R \\ |R| &\leq \frac{h^2}{24}|g'''(x_k + \vartheta)|, \ 0 < \vartheta < h \end{aligned} \tag{3.4.10}$$

The error of the approximation tends to be smaller by a factor of $\frac{1}{4}$ than for the symmetric difference. Another viewpoint to gain insight on this, is by means of Fourier analysis. Both difference operators are linear and shift invariant (LSI) and can be characterized in terms of Fourier spectra,

$$\frac{1}{2}(g_{k+1} - g_{k-1}) \circ\!\!-\!\!\bullet\, i\sin(\theta_1)G(i\theta) \tag{3.4.11}$$

$$g_{k+1} - g_k \circ\!\!-\!\!\bullet\, 2ie^{i\frac{\theta}{2}} \sin\left(\frac{\theta}{2}\right) G(i\theta)\ . \tag{3.4.12}$$

$G(i\theta)$ denotes the spectrum of g. Differentiation corresponds to a multiplication of the spectrum by $i\theta$. For the symmetric difference this is approximated by $i\sin(\theta)$. This is acceptable for $\theta < \pi/2$. For higher frequencies the sine function drops of and therefore performs rather badly for these components. The forward difference does not exhibit such a behavior though obviously the performance for higher frequencies is not perfect.

When working with two grids, $\mathcal{G}_h$ is referred to as the primary or geometry grid and $\mathcal{G}_h^*$ is called the dual or texture grid. Duality though is a concept that requires some kind of symmetry. Strictly speaking, it is misleading to talk about a primary grid and a dual grid, because the primary grid is the dual of the dual grid. The grids form a pair. In order to completely reveal the duality of the grids in a topological sense an artificial point, which is connected to all boundary points, has to be added to the grid $\mathcal{G}_h^*$. It forms the closure of the grid. Practically, we will nevertheless use the terms primary grid and dual grid. It should be clear what is meant. Working with two grids comes close to working with only one grid, but with the grid spacing reduced to $h/2$. Indeed the values of g and $\mathbf{m}$ are needed at the respective other grid. They have to be interpolated, which when working with one grid only is not necessary. The better performance is thus paid for with a higher number of operations. When using the disparity d for matching, it might also be advantageous to consider a grid for the texture that is shifted in one direction only. This is a little bit to be specific to be considered separately. It is not difficult to grasp the characteristic of such a setup from the presented cases. It can be generated from switching multiple 1D constellations in parallel.

The considerations so far have been focused on the reference domain $\mathcal{G}$. The texture grid and the geometry grid are rigidly linked and the performance of difference and interpolation operations is easily predictable. The distribution of sampling points with respect to the image domain $\mathcal{F}$ depends on the samples $\mathbf{m_k}$. There is no guarantee that the sampling density suffices to retain the essential characteristics of f. This is the already mentioned problem of badly modeled regions. Commonly the grid spacing h is chosen, such that on average grid cells in $\mathcal{G}$ are mapped to approximately the size of grid cells in $\mathcal{F}$ (cf. recommendations in [22, 23, 58]). As low frequencies dominate the spectrum of real imagery, the choice of h is not entirely critical.

The results of the discretization are denoted with the symbol h attached, indicating the dependency on the grid spacing h. The discretized form of E' is denoted by E'^h to make clear that the discretization is performed on E' and to distinguish it from $(E^h)'$. As already mentioned interchanging the order of operations yields different results in case of the forward similarity. For the following, we will assume only scalar-valued matching functions u. The vector of samples u_I is an element of the $(M+1) \times (N+1)$ dimensional vector space $\mathbb{R}^I$. The derivative of the energy term E'^h is a linear form on that vector space. We write

$$E'^h v_I = \sum_{\mathbf{i}\in I} E'^h_{\mathbf{i}} v_{\mathbf{i}}\ . \tag{3.4.13}$$

Similarly the bilinear form $E^{\mathrm{GN}h}$ is determined by its coordinates $E^{\mathrm{GN}h}_{\mathbf{ij}}$.

The results for the 1D case are given at first, simply for the reason that some properties of the method become rather lengthy when formulated for the 2D case. For the reverse modeling similarity measure only the 2D case is briefly considered.

The Finite-Sums Approach for the 1D Case

The grid setup for the 1D case is the very same as in the 2D case, hence parts of the notation which are already obvious are skipped.

Single Grid Discretization The discretization of (3.2.22) requires two steps: the approximation of the derivative and the approximation of the integral. The complexity of the formulas, especially considering the extension to the 2D case, necessitates the introduction of some notation. The derivative m' is approximated by the symmetric difference in the interior of I and the singles sided differences at the boundaries. Using the notation[4]

$$\partial m_i = \begin{cases} m_1 - m_0 & i = 0 \\ \frac{1}{2}(m_{i+1} - m_{i-1}) & 1 \le i \le M-1 \\ m_M - m_{M-1} & i = M \end{cases} , \tag{3.4.14}$$

we can write

$$m'(x_k) \approx \frac{1}{h}\partial m_k \qquad k \in I . \tag{3.4.15}$$

The approximations used throughout this section are all based on first order neighborhoods. Thus always the outermost grid positions receive a special treatment. The set of boundary indices is denoted by $\partial I = \{0, M\}$.

If m and g are given on the same grid, it will obviously be easiest to compute samples of the local measure V at (x_k, m_k). Accordingly the integral expression (3.2.22) is approximated by

$$E^h(m_I) = \sum_{k=0}^{M} a_k V(x_k, m_k)\partial m_k . \tag{3.4.16}$$

The weights a_k are determined by the trapezoidal rule for numerical quadrature,

$$a_k = \begin{cases} 1 & I \backslash \partial I \\ \frac{1}{2} & \partial I \end{cases} . \tag{3.4.17}$$

To some extent the weighting is a subjective matter, corresponding to the decision where to place the boundaries of the integration domain. If the outermost positions x_0 and x_M of the grid are assumed the interval end points, the weighting according to (3.4.17) follows. In principle uniform weighting is equally justified. The underlying assumption is that the sample positions are located in the middle of pixels and that the integration area extends over all pixels. The associated interval of integration is $[x_0 - \frac{h}{2}, x_M + \frac{h}{2}]$ extending effectively

[4]The notation follows [63, p. 1167]. We do however define the symmetric difference as an operator defined for sequences, thus the factor $\frac{1}{h}$ is taken into account separately. No adequate standard notation has been found. Other possibilities found are $\mu\delta$ [24, p. 225] or $\overline{\delta}$ [63, p. 1123]. As it is a fundamental operation, it has been found more convenient and more intuitive to use a single symbol instead of a composite symbol.

a width of half a pixel beyond the outermost grid locations. It has to be emphasized that the first choice is judicious with respect to consistency with the double grid discretization and with respect to simplicity of the results. Performing the following considerations using uniform weighting immediately reveals the assertion.

Importantly, the grid spacing h does not appear in (3.4.16), as it would be expected from the trapezoidal rule. The factor h cancels with the factor $\frac{1}{h}$ from the numerical approximation of m'. This is by no accident. The invariance property of the forward similarity implies that it is invariant with respect to the coordinate system used for the domain $\mathcal{G}$ and thus independent of the unit of length. Fixing the grid and choosing a new unit of length (e.g. centimeters instead of pixels) does not change anything. This must not be confused with the following. For a fixed unit of length reducing the grid spacing h increases the number of grid cells that cover a fixed area. Consequently, (3.4.16) becomes a better approximation of the integral value. It is for this reason that the grid spacing h is not entirely omitted from the formulation.

Differentiating (3.4.16) with respect to an interior node value m_i yields

$$\frac{\partial E^h}{\partial m_i} = V_y(x_i, m_i)\frac{m_{i+1} - m_{i-1}}{2} - \frac{1}{2}(V(x_{i+1}, m_{i+1}) - V(x_{i-1}, m_{i-1})) \,. \tag{3.4.18}$$

This looks somewhat different from the expression in the continuous case (3.2.23), it is an 1D analogue of (B.1.3). In the discrete case we do not have partial integration at hand to proceed exactly as in the continuous case. The method to establish the relation to (3.2.23) is to expand $\frac{\partial E^h}{\partial m_i}$ into a power series of the grid spacing h. Defining the function

$$\varphi(x, h) = V(x + h, m(x + h)) - V_y\,(x, m(x))\,(m(x + h) - m(x)) \tag{3.4.19}$$

we can rewrite (3.4.18) in the form

$$\frac{\partial E^h}{\partial m_i} = -\frac{1}{2}(\varphi(x_i, h) - \varphi(x_i, -h)) \,. \tag{3.4.20}$$

This expression is antisymmetric in h, hence if the Taylor series of $\varphi(x, h)$ is inserted all terms of even order, importantly the quadratic terms, cancel. The first derivative of the discretized energy functional is approximated by omitting all terms of higher than linear order. For that purpose, the first derivative of the helper function φ is needed,

$$\left.\frac{\partial \varphi}{\partial h}\right|_{h=0} = V_x \,. \tag{3.4.21}$$

Consequently, the derivative of E , including also the boundary elements, is given by

$$\frac{\partial E^h}{\partial m_i} = -a_i h V_x(x_i, m_i) + \begin{cases} -V(x_0, m_0) & + O(h^2) & i = 0 \\ & + O(h^3) & I \backslash \partial I \\ +V(x_M, m_M) & + O(h^2) & i = M \end{cases} \tag{3.4.22}$$

The explicit part is equal to $E_i'^h$, as it would have been obtained by a direct discretization of (3.2.23). The error introduced by terminating the Taylor series is hidden behind the Landau symbol O. This is the discrepancy between the exact derivative of the discretized similarity measure and $E_i'^h$. The error of approximating E' by E'^h is an related though distinguished

quantity. The difference between $E_i'^h$ and $\frac{\partial E^h}{\partial m_i}$ can be roughly estimated upon using the third derivative of φ at position $h = 0$, which evaluates to

$$\left.\frac{\partial^3 \varphi}{\partial h^3}\right|_{h=0} = V_{xxx} + 3V_{xxy}m' + 3(V_{xyy}m'^2 + V_{xy}m'') + V_{yyy}m'^3 + 3V_{yy}m'm'' \,. \tag{3.4.23}$$

Ideally, one would like to consider $\frac{\partial^3 \varphi(x,h)}{\partial h^3}$ and utilize the mean value theorem, yielding an exact error bound. The corresponding expression firstly is rather lengthy and secondly it is rather cumbersome to draw any conclusions from it directly. (3.4.23) is primarily of value for the comparison with the double grid discretization. The estimate for the error of approximation is given by

$$\left| E_i'^h - \frac{\partial E^h}{\partial m_i} \right| \lesssim \frac{h^3}{6} \left| \frac{\partial^3 \varphi}{\partial h^3} \right|_{h=0} \tag{3.4.24}$$

The prerequisite is that $\varphi(x, h)$ is three times continuously differentiable with respect to h. This condition will hardly be fulfilled in practice. For the case of the local least-squares measure (3.1.7), this would imply that $f \in C^3$. Practically, f is often constructed using piecewise linear interpolation (or bilinear interpolation for the 2D case) clearly violating the condition. As already mentioned, (3.4.24) may be used for comparison purposes, provided one is aware of its limitations.

Remark. Improving the asymptotic behavior of the error of approximation (3.4.24) can be done by using a higher order approximation of m',

$$m'(x_k) = h \sum_j d_{k-j} m_k + O(h^{p+1}) \,. \tag{3.4.25}$$

Special boundary treatment is necessary for the numerical differentiation as well as for the quadrature formula. The error in (3.4.22) will be of order $O(h^{p+1})$. Such approximations of derivatives can be constructed based on Lagrange polynomial interpolation for each point and $p/2$ neighbors to either side [24]. The classical symmetric difference has one neighbor on either side, corresponding to $p = 2$. The filter taps are given by $d_{-j} = [-1, 0, 1]/2$. For $p = 4$ the taps are $d_{-j} = [1, -8, 0, 8, -1]/12$. It has to be noted that the improvement concerns the asymptotic behavior only, providing merely theoretical insight. It is not clear whether this is of practical value.

The inexact pairing of E^h and E'^h has been found important for the performance of optimization methods. Some algorithms try to solve for $E(\mathbf{m}^*) = \min!$ without using gradient information, others try to solve for vanishing derivative $E'(\mathbf{m}^*) = 0$ without ever computing the objective function E itself. Methods using both quantities may be affected by the discrepancy between $E^{h\prime}$ and E'^h. An example is the Powell-Wolfe rule (Section 4.2.2). This justifies the lengthy treatment of that aspect. The discretization of the Gauss-Newton (3.2.24) term is not that critical. In optimization it is, like the Hesse matrix, used to improve the search direction, aiming at rapid convergence [30, 45]. Shortcomings will thus affect only the convergence rate, but not the result. For the Gauss-Newton term the same order of approximation and the same weighting is applied. The discretization of a bilinear operator yields a bilinear operator in $\mathbb{R}^I$, i.e. a matrix. In the particular case the matrix has nonzero entries only along the diagonal. The diagonal elements are given by

$$E_{i,i}^{\mathrm{GN}h} = a_i \frac{|hN_x(x_i, m_i)|^2}{\partial m_i} \,. \tag{3.4.26}$$

The diagonal property is a consequence of the fine granularity of the discretization. The method mentioned in Section 4.1.1 provides a reduction of the parameter space. Thereby a coupling between the samples of m is introduced, causing off-diagonal elements in the matrix.

For the evaluation of (3.4.22) a local similarity measure has to be provided. A complete list of results for the local least-squares measure (3.1.7) is given in Section A.0.1. So far the considerations have been focussed on the influence of the grid layout on the numerical integration and the approximation of m'. (3.4.22) contains the derivative of the local measure V_x, which again necessitates numerical differentiation. Using the local least-squares measure the required quantity is

$$-hV_x(x_k, m_k) = hg'(x_k)(f(m_k) - g(x_k)) \;, \tag{3.4.27}$$

thus the derivative of the texture g has to be approximated. In particular, $hg'(x_k) \approx \partial g_k$, is chosen. The grid arrangement again plays a crucial role. According to the introductory section, the symmetric difference performs rather badly for highly oscillatory components. Since the first derivative is the driving force for a broad class of optimization algorithms, this might diminish the performance. Using higher order differentiation schemes will mitigate this aspect only partially. In the extreme case, if $g_k = (-1)^k$, no approximation of whatever order will enable a successful matching. As experience shows, real imagery is low frequency dominated ensuring the applicability of the single grid discretization despite of theoretical drawbacks.

Double Grid Discretization If the node values of g are given on the dual grid the approximation of (3.2.22) is advantageously based on evaluations of V at the dual grid positions. The derivative m' at the dual grid positions is approximated utilizing the central difference operator [24, p. 225]

$$\delta m_k = m_{k+\frac{1}{2}} - m_{k-\frac{1}{2}} \qquad k \in I^* \;, \tag{3.4.28}$$

leading to

$$m'(x_k) \approx \frac{1}{h}\,\delta m_k \qquad k \in I^* \;. \tag{3.4.29}$$

Evaluating V at (x_k, m_k) for $k \in I^*$ requires m_k to be interpolated from neighborhood values, as already mentioned in the introductory section. Defining

$$\overline{m}_k = \frac{1}{2}(m_{k+\frac{1}{2}} + m_{k-\frac{1}{2}}) \qquad k \in I^* \;. \tag{3.4.30}$$

we can write

$$m(x_k) \approx \overline{m}_k \qquad k \in I^* \;. \tag{3.4.31}$$

None of the above requires a special boundary treatment because the grid locations of the dual grid are located in the interior of the domain covered by the primary grid. Equipped with these definitions, the discretized form of the energy functional is given by

$$E^h(m_I) = \sum_{k \in I^*} V\,(x_k, \overline{m}_k)\,\delta m_k \;. \tag{3.4.32}$$

This corresponds to the midpoint rule of numerical quadrature. The derivative with respect to the node values is easily computed. For an interior index $i \in I \backslash \partial I$

$$\begin{aligned}\frac{\partial E^h}{\partial m_i} = &\frac{1}{2} V_y \left(x_{i+\frac{1}{2}}, \frac{m_{i+1} + m_i}{2} \right) (m_{i+1} - m_i) - V \left(x_{i+\frac{1}{2}}, \frac{m_{i+1} + m_i}{2} \right) \\ &+ \frac{1}{2} V_y \left(x_{i-\frac{1}{2}}, \frac{m_i + m_{i-1}}{2} \right) (m_i - m_{i-1}) + V \left(x_{i-\frac{1}{2}}, \frac{m_i + m_{i-1}}{2} \right)\end{aligned} \tag{3.4.33}$$

We follow a similar procedure like in the previous case. By means of the function

$$\begin{aligned}\varphi \left(x, \frac{h}{2} \right) = &V \left(x + \frac{h}{2}, \frac{m(x+h) + m(x)}{2} \right) \\ &- \frac{1}{2} V_y \left(x + \frac{h}{2}, \frac{m(x+h) + m(x)}{2} \right) (m(x+h) - m(x)) \,,\end{aligned} \tag{3.4.34}$$

(3.4.33) can be rewritten in the form

$$\frac{\partial E^h}{\partial m_i} = - \left(\varphi \left(x_i, \frac{h}{2} \right) - \varphi \left(x_i, -\frac{h}{2} \right) \right) \,. \tag{3.4.35}$$

Expanding this into a power series of h and omitting all terms of order higher than linear again leads to (3.4.22). The difference between the single grid and the double grid discretization is hidden behind the Landau symbol O. The error of approximation is of course different. The error estimate is again based on the third derivative of φ at $h = 0$

$$\left. \frac{\partial^3 \varphi}{\partial h^3} \right|_{h=0} = V_{xxx} - 3V_{xxy} m' - 2V_{yyy} m'^3 - 6V_{yy} m' m'' \,, \tag{3.4.36}$$

leading to

$$\left| E_i'^h - \frac{\partial E^h}{\partial m_i} \right| \lesssim \frac{h^3}{24} \left| \frac{\partial^3 \varphi}{\partial h^3} \right|_{h=0} \,. \tag{3.4.37}$$

Comparing (3.4.37) and (3.4.24) and inserting (3.4.36) and (3.4.23) respectively reveals that for the double grid discretization, not surprisingly, the error tends to be smaller by a factor of $\frac{1}{4}$.

The style of grid arrangement also affects the computation of the derivative of the local similarity measure. The list of results as obtained for the local least-squares measure (3.1.7) is given in (A.0.2). We will give some explanatory remarks in the following. To compute (3.4.22) in the double grid setup the same quantity (3.4.27) is needed. A difference is that $g(x_k)$ for $k \in I$ is not directly available, but has to be interpolated from neighborhood values. Advantageously, the approximation of the derivative $g_x(x_k)$ is needed between the samples of the dual grid, i.e. at the primary grid locations, which is possible with higher accuracy. The operations are quite the same as defined in (3.4.28) and (3.4.30). This time they are applied to calculate quantities at the primary grid from samples at the dual grid. A special treatment of the boundary is necessary. For the interpolation operation there is essentially one choice,

$$\overline{g}_k = \begin{cases} g_{\frac{1}{2}} & i = 0 \\ \frac{1}{2}(g_{i+\frac{1}{2}} + g_{i-\frac{1}{2}}) & I \backslash \partial I \\ g_{M-\frac{1}{2}} & i = M \end{cases} \,. \tag{3.4.38}$$

For the difference operator the choice is not that obvious. The following observation serves as guideline. In case of the local least-squares measure the considered double grid discretization has an interesting equivalent within the Finite-Elements method. The derivative (A.0.2b) is exactly the same as would have been obtained with the Finite-Elements method assuming m piecewise linear and g to be piecewise constant. Equivalence, including the equivalent treatment of boundary nodes, is reached in using the following definition

$$\delta g_k = \begin{cases} 0 & i = 0 \\ \frac{1}{2}(g_{i+\frac{1}{2}} - g_{i-\frac{1}{2}}) & I \backslash \partial I \\ 0 & i = M \end{cases} . \tag{3.4.39}$$

As already has been stressed in the introductory section, approximating the derivative between grid nodes can be done with higher accuracy, circumventing some demerits of what is encountered in the single grid approach. For the double grid setup the limiting case $g_k = (-1)^k$ is no problem. Very high frequency components are nevertheless delicate with respect to the convergence of the matching algorithm.

The Finite-Sums Approach for the 2D Case

The principles of the discretization have already been given in the previous section. They apply to the 2D case as well. All considerations from the 1D case can be carried over. The notational effort is however in some cases more than doubled. Pretty much is obvious from the preceding and does not have to be repeated. Some remarks are however added. Only the case that $\mathbf{m}$ is parameterized by a scalar function u is considered. This captures all that is needed herein. It is understood that the samples of the map are given by $\mathbf{m_k} = \mathbf{m}(\mathbf{x_k}, u_\mathbf{k})$ $\mathbf{k} \in I$. The collected results for the local least-squares measure can be found in Appendix A.

Single Grid Discretization The definitions introduced for the 1D case are used extensively in the 2D case because the number of indices otherwise covers the essential results. a_k^i denotes the weighting according to (3.4.17) for the i'th coordinate direction. The trapezoidal rule for the 2D case uses the tensor product weighting $a_{k,l} = a_k^1 a_l^2$. Consequently, the corners are weighted by $\frac{1}{4}$, the edges by $\frac{1}{2}$ and the interior by 1. The integral (3.1.12) is approximated by the sum

$$E^h(u_I) = h^2 \sum_{\mathbf{k} \in I} a_\mathbf{k} V(\mathbf{x_k}, \mathbf{m_k}) A_\mathbf{k} . \tag{3.4.40}$$

The quantity $A_\mathbf{k}$ denotes the approximation of the Jacobian $\det \mathbf{m}'(\mathbf{x_k})$. Again the difference operator ∂ is used, extended by an index to denote the coordinate direction. $A_\mathbf{k}$ is thus given by

$$A_\mathbf{k} = -\frac{1}{h} p_1(\mathbf{x_k}, u_\mathbf{k}) \partial_1 u_\mathbf{k} - \frac{1}{h} p_2(\mathbf{x_k}, u_\mathbf{k}) \partial_2 u_\mathbf{k} + p_3(\mathbf{x_k}, u_\mathbf{k}) . \tag{3.4.41}$$

In contrast to the 1D case, the independence with respect to the unit of length of h is not clearly apparent. Changing the unit of length, for example to ensure that $h = 1$, will not only affect the value of h, but also the components of p. Effectively $A_\mathbf{k}$ is proportional to $\frac{1}{h^2}$. The invariance property of the similarity measure is thus not violated. The discretization of

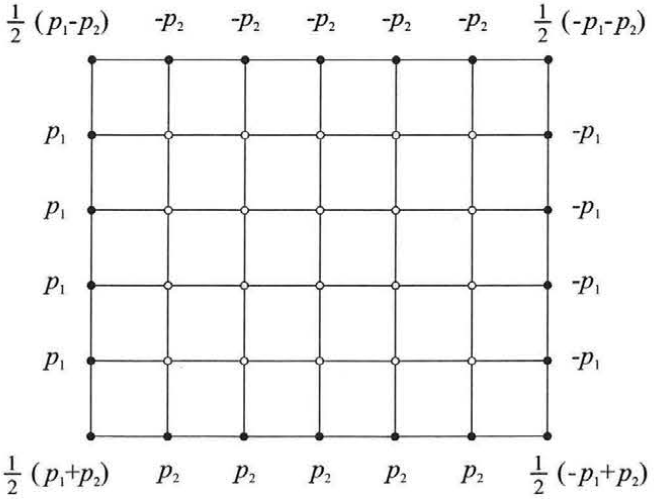

Figure 3.7: The weighting factors for the boundary term according to (3.4.42).

the first derivative of the forward similarity leads to

$$\begin{aligned} E'^{h}_{i,j} &= -a_{i,j} h^2 \, \tilde{\mathrm{d}}_u V \\ &+ \begin{cases} +a_j^2 h p_1 V & i = 1 \\ -a_j^2 h p_1 V & i = M \end{cases} \\ &+ \begin{cases} +a_i^1 h p_2 V & j = 1 \\ -a_i^1 h p_2 V & j = N \end{cases} . \end{aligned} \tag{3.4.42}$$

The relation to $\frac{\partial E^h}{\partial u_{i,j}}$ follows a similar line as in the 1D case. The weights of the boundary term are depicted in Fig. 3.7. The values have to be multiplied by $hV(\mathbf{x}_k, \mathbf{m}_k)$. The boundary term takes plenty of room in (3.4.42). A compact notation is possible using the normal vector $\mathbf{n}$ at the boundary. It is defined by

$$\begin{aligned} n^1_{k,l} &= \begin{cases} -a_l^2 & k = 0 \\ +a_l^2 & k = M \end{cases} \\ n^2_{k,l} &= \begin{cases} -a_k^1 & l = 0 \\ +a_k^1 & l = N \end{cases} . \end{aligned} \tag{3.4.43}$$

By means of this definition, the boundary term is simply given by $-h\mathbf{n} \cdot \mathbf{p}V$.

Double Grid Discretization It is pretty obvious how the discretization of the particular terms is performed for the double grid setup. An implementation detail concerns the computation of differences. Assume values given on one of the grids and an approximation of the derivative at the respective dual grid to be required. The difference operator that is employed is a generalization of (3.4.28),

$$\delta_1 \, u_{k,l} = \frac{1}{2}(u_{k+\frac{1}{2},l+\frac{1}{2}} - u_{k-\frac{1}{2},l+\frac{1}{2}} + u_{k+\frac{1}{2},l-\frac{1}{2}} - u_{k-\frac{1}{2},l-\frac{1}{2}}) \tag{3.4.44}$$

and similar for the second coordinate direction. It is always used in combination with the derivative $\tilde{d}_u$. An example is the approximation of the Jacobian $\det \mathbf{m}'$. Explicitly it is given by

$$\begin{aligned} A_{k,l} = &-\frac{1}{h}p_1(\mathbf{x}_{k,l}, u_{k,l})(u_{k+\frac{1}{2},l+\frac{1}{2}} - u_{k-\frac{1}{2},l+\frac{1}{2}} + u_{k+\frac{1}{2},l-\frac{1}{2}} - u_{k-\frac{1}{2},l-\frac{1}{2}}) \\ &-\frac{1}{h}p_2(\mathbf{x}_{k,l}, u_{k,l})(u_{k+\frac{1}{2},l+\frac{1}{2}} + u_{k-\frac{1}{2},l+\frac{1}{2}} - u_{k+\frac{1}{2},l-\frac{1}{2}} - u_{k-\frac{1}{2},l-\frac{1}{2}}) \\ &+p_3(\mathbf{x}_{k,l}, u_{k,l}) \ . \end{aligned} \tag{3.4.45}$$

Expressions of that type appear more than once. It is possible to reduce the number of required algebraic operations for some constellations. Each term $u_{k\pm\frac{1}{2},l\pm\frac{1}{2}}$ enters twice. Depending on how the p_i are calculated, a rearrangement of the expression can reduce the computational effort. What effectively can be done is that a second coordinate system, which is rotated by 45°, is used. The $\tilde{p}_i$ for the rotated coordinate system have to be provided. The derivatives of u with respect to the coordinate directions of the rotated system are the derivatives in the diagonal directions of the grid. They require less operations than (3.4.44). The Jacobian hence can also be evaluated by

$$\begin{aligned} A_{k,l} = &-\frac{1}{h}\tilde{p}_1(\mathbf{x}_{k,l}, u_{k,l})(u_{k+\frac{1}{2},l+\frac{1}{2}} - u_{k-\frac{1}{2},l-\frac{1}{2}}) \\ &-\frac{1}{h}\tilde{p}_2(\mathbf{x}_{k,l}, u_{k,l})(u_{k-\frac{1}{2},l+\frac{1}{2}} - u_{k+\frac{1}{2},l-\frac{1}{2}}) \\ &+\tilde{p}_3(\mathbf{x}_{k,l}, u_{k,l}) \ . \end{aligned} \tag{3.4.46}$$

This principle has been used for the practical implementations of the DLT configuration. The benefit of that strategy however depends on the circumstances.

Discretization of the Reverse Modeling Similarity Measure

Lastly, we add some comments on the discretization of the reverse modeling similarity measure, which is performed straight forward. Importantly, a double grid discretization does not make sense, as no derivatives in the reference space are required. Applying the same weighting as in the single grid approach leads to

$$E^h(u_I) = \sum_{\mathbf{k}\in I} a_{\mathbf{k}} h^2 V(\mathbf{x_k}, \mathbf{m_k}) \ . \tag{3.4.47}$$

The weighting is applied for consistency reasons. The results of this section are not affected by the type of weighting. The derivative is easily evaluated to

$$\frac{\partial E^h}{\partial u_{\mathbf{i}}} = a_{\mathbf{i}} h^2 V_{\mathbf{y}}(\mathbf{x_i}, \mathbf{m_i}) \frac{\partial \mathbf{m_i}}{\partial u_{\mathbf{i}}} \ . \tag{3.4.48}$$

This is equal to the direct discretization of (3.2.27), but specialized to a scalar matching function. The Gauss-Newton term is given by

$$E^{\mathrm{GN}h}_{\mathbf{ii}} = a_{\mathbf{i}} h^2 |N_{\mathbf{y}}(\mathbf{x_i}, \mathbf{m_i}) \frac{\partial \mathbf{m_i}}{\partial u_{\mathbf{i}}}|^2 \ . \tag{3.4.49}$$

All derivatives are computed in the image domain $\mathcal{F}$. They are carried out analytically, because the image function f is not specified by samples only, but by some analytical function representation.

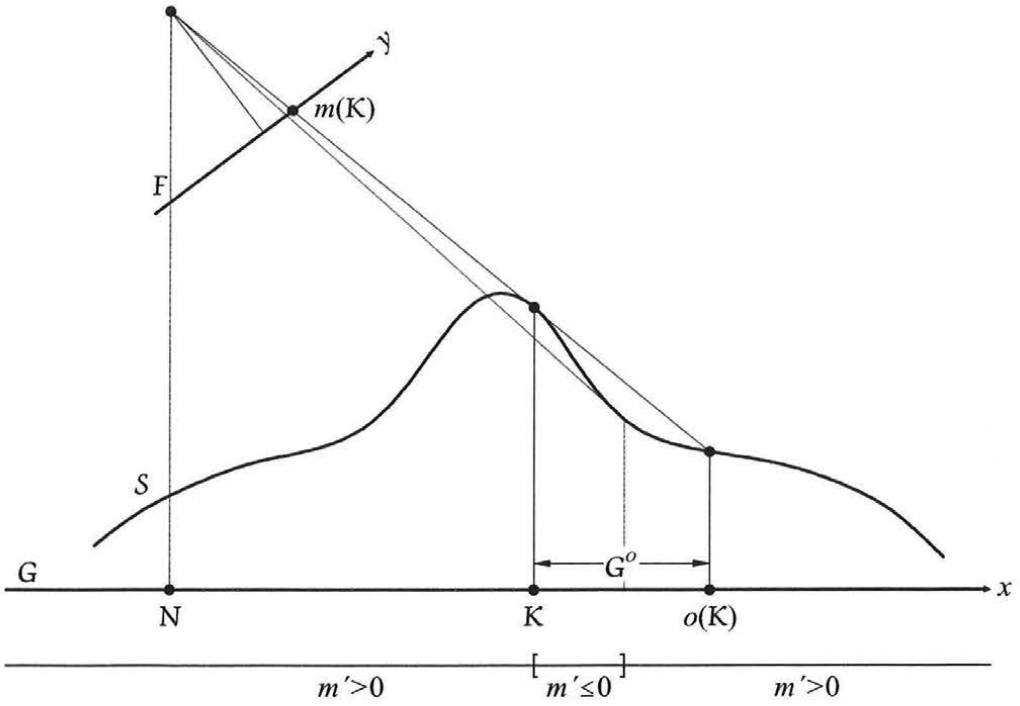

Figure 3.8: Example of an occlusion for a 1D orthophoto-to-image matching configuration.

The results specialized to the local least-squares measure can be found in Appendix A. For this particular setup the discrete formulation is equivalent to the nonlinear adjustment problem

$$f(u_{\mathbf{k}}) = g_{\mathbf{k}} + v_{\mathbf{k}} \,, \tag{3.4.50}$$

with the unknowns $u_{\mathbf{k}}$ and the residuals $v_{\mathbf{k}}$ (cf. [22, 23]).

3.5 Modeling in the Presence of Occlusions

Occlusions are a delicate matter in the course of image matching. One must not expect an in depth treatment, because too little practical experience is available until now. On the other hand, the forward modeling similarity measure covers the case of occlusions in a natural way so that it seems worthwhile to extend the results to that case.

Occlusions arise if along an imaging ray multiple surface elements are located. A 1D sketch of an orthophoto-to-image matching configuration involving occlusions is depicted in Fig. 3.8. The occluded surface regions can be divided in two classes. Surface elements not visible may have a surface normal vector pointing away from the camera, $\det \mathbf{m}' \leq 0$. For 1D constellations the condition is given by $m' \leq 0$. These regions are easily found to be invisible. An example is the backside of a sphere with respect to some viewing direction. The other case is that the surface element is oriented towards the camera, $\det \mathbf{m}' > 0$, but that there is something in front. Both types of regions can easily be seen in Fig. 3.8. The boundary $\partial \mathcal{G}^O$ of the occluded region can be divided in two classes, too. Parts of the boundary are locations where the Jacobian $\det \mathbf{m}'$ changes its sign. These points form the *contour line* $\mathcal{K}$ of an object. From the viewpoint of the camera, the remainder of $\partial \mathcal{G}^O$ is located behind those contour points. They constitute what herein is called the *shadow line*. The function

that associates to every point of a contour line its corresponding point of the shadow line is defined by

$$\mathbf{o} : \mathcal{K} \mapsto \mathbf{o}(\mathcal{K}) \subset \partial\mathcal{G}^O \,. \tag{3.5.1}$$

We write $\mathbf{o}(\mathcal{K})$ for the shadow line. With respect to domain $\mathcal{F}$ the images of $\mathcal{K}$ and $\mathbf{o}(\mathcal{K})$ coincide. The boundary of the occluded part can be written as

$$-\partial\mathcal{G}^O = \mathcal{K} - \mathbf{o}(\mathcal{K}) \,. \tag{3.5.2}$$

The minus signs are included because it is convenient to have $\mathcal{K}$ consistently oriented with the boundary of the reference domain $\partial\mathcal{G}$. The case that parts of the boundary $\partial\mathcal{G}$ are occluded is not considered. There is nothing special about that situation. It just makes the notation a little bit uncomfortable. The boundary of the visible part is given by

$$\partial\mathcal{G}^V = \partial\mathcal{G} + \mathcal{K} - \mathbf{o}(\mathcal{K}) \,. \tag{3.5.3}$$

For simplicity reasons the following considerations will be argued, at least partially, based on the 1D case. Consider the constellation depicted in Fig. 3.8 and some point x of the reference domain located to the right of the nadir point $\mathcal{N}$. To determine whether it is visible or occluded, it suffices to consider the surface points, respectively the map m, to the left of x. For the particular situation a point x is visible if and only if $m(x) > m(\xi)$ for all $\xi < x$. For 1D constellations it is always simple to determine the occlusions. Depending on the circumstances, either all points $\xi < x$ or all points $\xi > x$ have to be considered, but this has no influence on the principle. For the examples the first case is assumed. The point where m assumes its maximum to the left of an occluded region is the contour point. As will become apparent from the following, the case that the contour point (or the contour line for the 2D case) is not clearly defined, most importantly if $m' = 0$ for a whole interval, is a rather troublesome situation.

3.5.1 The Similarity Measure in the Presence of Occlusions

The formulation of the similarity measure strongly depends on the definition of the map $\mathbf{m}$ as the mapping from $\mathbf{x} \in \mathcal{G}$ to the corresponding point $\mathbf{y} \in \mathcal{F}$. The map $\mathbf{m}$ can be defined only for the visible part $\mathcal{G}^V$ (cf. (2.2.5)). For a point in the occluded part $\mathcal{G}^O$, there is no corresponding point in the image domain. So far we have assumed that the reference domain is entirely visible. Obviously, if occlusions come into play, this has to be dropped. The image space formulation (3.1.8) can be used unaltered. However, the inversion of the map $\mathbf{m}^{-1}$ has to take the occlusions into account. For the reference space formulation (3.1.12), the integral has to be restricted to the visible part,

$$E(\mathbf{m}) = \int_{\mathcal{G}^V} V \circ (\mathbf{id}_{\mathcal{G}}, \mathbf{m}) \det \mathbf{m}' \, \mathrm{d}A \,. \tag{3.5.4}$$

If the occluded regions are known a priori, no problem arises. In the course of optimization, the visible part of the reference domain is, in connection with the map $\mathbf{m}$, subject of iterative refinement. This complicates the matter significantly, as shown in Section 3.5.2, where the derivative of the image energy in the presence of occlusions is considered.

An alternative, but equivalent way to treat occlusions is to mask out occluded regions by some visibility indicator. The method is supported by the fact that the similarity measure

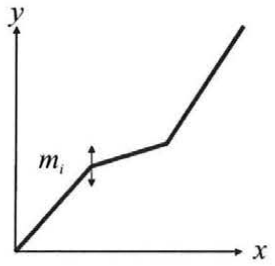

Figure 3.9: No occlusion.

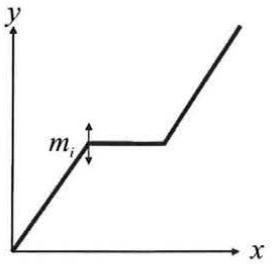

Figure 3.10: Transition from visible to occluded.

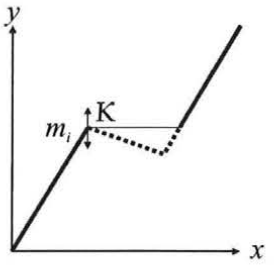

Figure 3.11: Occlusion, $\mathcal{K}$ is a regular contour point.

(3.1.12) already includes, as a limiting case, the situation that parts of the model are not visible, if the Jacobian $\det \mathbf{m}' \to 0$. We write

$$E(\mathbf{m}) = \int_{\mathcal{G}} V \circ (\mathbf{id}_{\mathcal{G}}, \mathbf{m}) \, \mathrm{vis}_{\mathbf{m}} \, \mathrm{d}A \; . \tag{3.5.5}$$

The visibility indicator is given by

$$\mathrm{vis}_{\mathbf{m}}(\mathbf{x}) = \begin{cases} \det \mathbf{m}'(\mathbf{x}) & \forall \mathbf{x} \in \mathcal{G}^V \\ 0 & \forall \mathbf{x} \in \mathcal{G}^O \end{cases} \; . \tag{3.5.6}$$

This definition ensures that $\mathrm{vis}_{\mathbf{m}}(\mathbf{x}) \geq 0$. The extension is by no means sophisticated and follows neatly from the considered concept. The problem that the map $\mathbf{m}$ is not defined for occluded regions is not solved by the introduction of the visibility indicator. However, based on (3.5.5) the occluded areas do not contribute to the global similarity measure. We are free to define $\mathbf{m}$ in these regions according to convenience. In [5] a normal case configuration is considered. The disparity and the thereof-derived map are explicitly defined for occluded regions,

$$\tilde{\mathbf{m}}(\mathbf{x}) = \max_{\substack{\xi_1 \leq x_1 \\ \xi_2 = x_2}} \mathbf{m}(\boldsymbol{\xi}) \; . \tag{3.5.7}$$

The definition ensures that the new map $\tilde{\mathbf{m}}$ has vanishing Jacobian $\det \tilde{\mathbf{m}}'$ in the occluded regions. Such a concept may be used for the easier mathematical treatment of the problem. Usually there exists some natural definition of $\mathbf{m}$ in the occluded regions.

The visibility indicator is defined to be a non-negative, but otherwise continuous quantity. It is neither bounded to the interval $[0, 1]$ nor a simple indicator variable with values in $\{0, 1\}$, like in [37, 51]. Some approaches [5, 10] omit the use of an indicator variable. However, any model that takes occlusions into account inevitable has to distinguish between points being in the visible part $\mathcal{G}^V$ and points being in the occluded part $\mathcal{G}^O$. Whether this is modeled by an indicator variable or not, it is always a binary decision. The concept of visibility, as introduced herein, grants a smooth transition between occlusion and visibility. Consider the dependency of E on the map $\mathbf{m}$. The critical point with respect to continuity is that an entire part of $\mathcal{G}$ disappears simultaneously. This eliminates the contributions to the integral of the whole part. As will be shown, it is of less impact if the area of the occluded part is increased or reduced continuously. The effect is explained by means of the three 1D situations depicted in Fig. 3.9-3.11. m is assumed to be parameterized by the nodes of a piecewise linear function. Starting from the situation depicted in Fig. 3.9 and shifting the node m_i in the positive y direction leads to the transition state Fig. 3.10. Shifting the node further causes an occlusion

Fig. 3.11. The graph of m is plotted as dotted line within the occluded part. The contour point $\mathcal{K}$ coincides with the node. At the transition state Fig. 3.10, a binary indicator variable abruptly switches off the contributions of the entire interval, which leads to a discontinuity of the similarity measure. (3.5.6) ensures that this does not happen. Starting from the situation depicted in Fig. 3.9 and shifting the node m_i in the positive y direction, the interval receives a lower and lower weight m' ($\widehat{=} \det \mathbf{m}'$) the closer m_i gets to the transition point. At the transition the Jacobian vanishes completely, $m' = 0$. Proceeding to case Fig. 3.11 the weight remains zero by (3.5.6), because the interval is now occluded. The forward energy is thus continuous in the presence of occlusions. A discontinuity of the first derivative E' is encountered however (cf. Section 3.5.2). It will not be mentioned further, but evidently this concerns the impact of occlusions only. The influence of the local measure V on continuity and differentiability is not considered explicitly.

The symbol $\mathrm{vis}_{\mathbf{m}}$ encodes a complex phenomena and it shall not give reason to overlook the theoretical and practical complications. First, note that the visibility is inherently connected to the imaging geometry. Evaluating the Jacobian $\det \mathbf{m}'$ of a map is always possible, provided the map is differentiable. Distinguishing between visible and occluded points is possible only based on detailed knowledge about the imaging geometry. In particular the relative or absolute orientation is needed. For applications such as the one presented in [1], where X-ray images of a hand, taken at different times, are compared, the concept of visibility, as used herein, can not be carried over. The second important point is that the visibility indicator depends on the whole map $\mathbf{m}$. The knowledge of $\mathbf{m}$ for some part of $\mathcal{G}$ is not sufficient to determine whether some point of that part is visible or occluded. In contrast $\det \mathbf{m}'$ can be evaluated at every position where $\mathbf{m}$ can be differentiated. This has to be kept in mind for the mathematical treatment. Last, the computational complexity of the visibility analysis should not be forgotten. For the case of matching normal case stereo images, determining the occluded regions is almost trivial. If the model is a true 3D model, the visibility analysis is rather time consuming.

3.5.2 First Derivative of the Similarity Measure in the Presence of Occlusions

Using derivatives of the similarity measure in the presence of occlusions has to be done very carefully. Consider the map $\mathbf{m}$ composed of a finite number of basis functions (cf. Section 4.1.1). For that case, theoretical considerations and practical experiments reveal that the derivative of the image energy term is piecewise continuous only. If the map $\mathbf{m}$ is assumed an element of an infinite dimensional space, discontinuities that are more severe have to be expected. So far, the results have not been used for the practical implementation and thus they have to be viewed with criticism.

For the differentiation, the effect of small changes of the map $\mathbf{m}$ on the forward modeling similarity measure is considered. The easiest and most intuitive method uses the image domain formulation (3.1.8). This is done by looking at the resulting changes from the viewpoint of the image domain $\mathcal{F}$, as it has been done in Section 3.2.1. The explanations favor intuition over mathematical rigor. Applying a small change $\boldsymbol{\mu}$ to the map $\mathbf{m}$ will, aside of the already considered effects, shift the image of the contour line $\mathbf{m}(\mathcal{K})$ to a new location. The new image of the contour line is shifted for two reasons. Firstly, modifying the map $\mathbf{m}$ will change the location of the contour line in $\mathcal{G}$ to a new position $\tilde{\mathcal{K}}$. Secondly, the map itself changes, hence the new image is given by $(\mathbf{m} + \boldsymbol{\mu})(\tilde{\mathcal{K}})$. We base our considerations on the *regularity*

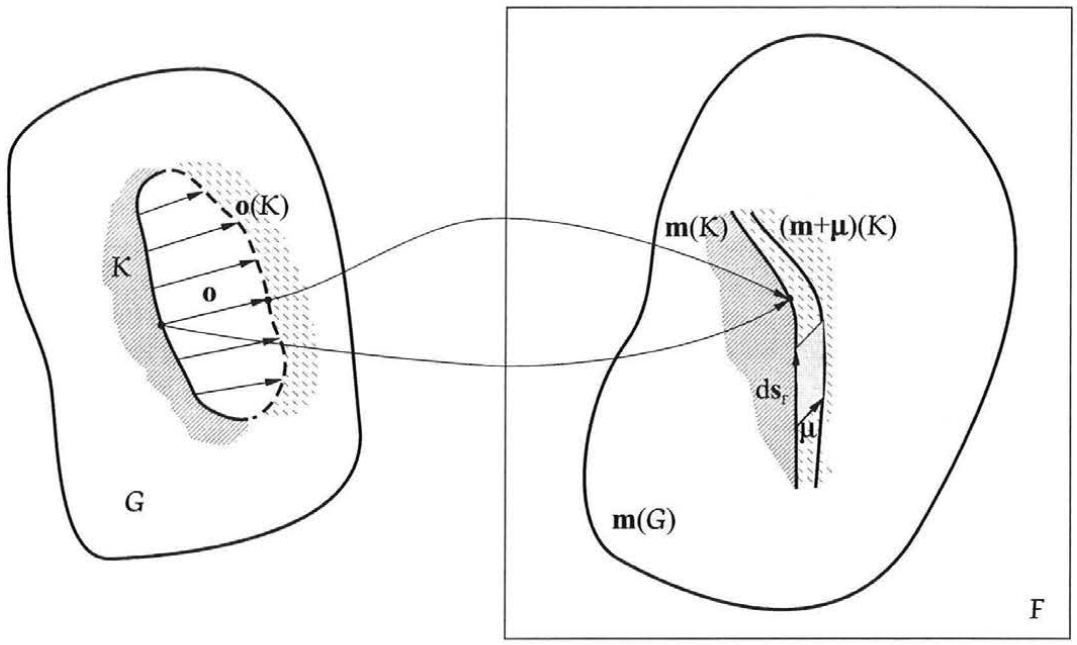

Figure 3.12: The image of domain $\mathcal{G}$ under the mapping $\mathbf{m}+\boldsymbol{\mu}$, in the presence of occlusions.

assumption that

$$(\mathbf{m}+\boldsymbol{\mu})(\tilde{\mathcal{K}}) \simeq (\mathbf{m}+\boldsymbol{\mu})(\mathcal{K}) \tag{3.5.8}$$

holds up to first order. An illustration is depicted in Fig. 3.12. The regularity assumption does neither imply that $\mathcal{K} = \tilde{\mathcal{K}}$ nor that $\mathcal{K} \simeq \tilde{\mathcal{K}}$ up to first order. Nevertheless, $\tilde{\mathcal{K}}$ is not plotted in Fig. 3.12, as it is not needed in the following.

The regularity assumption is not valid under all circumstances, but it is justified for some relevant cases. Consider again Fig. 3.9-3.11. Fig. 3.11 shows a situation involving an occlusion. The contour point $\mathcal{K}$ coincides with the node. m_i remains the contour point also if it is shifted a small amount in either direction. Thus (3.5.8) holds strictly. The critical situation is the transition between the two cases, Fig. 3.10. Moving the node downwards leads to the situation depicted in Fig. 3.9, which involves no occlusion. Moving the point upwards generates an occlusion. At the transition the regularity assumption is violated.

The regularity assumption also holds if $\mathbf{m}$ and $\boldsymbol{\mu}$ are continuously differentiable in a neighborhood of $\mathcal{K}$ and if the contour is a *regular contour*. We again argue based on the 1D case. The contour point $m'(x_K) = 0$ is said to be regular if for some neighborhood N of x_K, $|m'(x)| \geq \epsilon|x - x_K|$ for some $\epsilon > 0$. This condition expresses the fact that the solution of $m'(x) = 0$ is well defined and thus not very sensible to perturbations. Denoting by η the supremum $\eta = \sup_N \mu'$, the following holds

$$|m'(x) + \mu'(x)| \geq \epsilon|x - x_K| - \eta\,, \quad \forall x \in N\,. \tag{3.5.9}$$

From this follows that the shifted contour point $\tilde{x}_K$, satisfying $m'(\tilde{x}_K) + \mu'(\tilde{x}_K) = 0$, is in the interval $\tilde{x}_K \in [x_K - \frac{\eta}{\epsilon}, x_K + \frac{\eta}{\epsilon}]$. Consequently, if the derivative of μ' converges uniformly to m', $\tilde{x}_K$ converges to x_K with the same rate. Now, since $m'(x) = 0$ by assumption

$$m(\tilde{x}^K) + \mu(\tilde{x}^K) = m(x^K) + \mu(x^K) + O(\eta^2)\,, \tag{3.5.10}$$

thus (3.5.8) holds. We do not claim that these considerations cover all details. Nevertheless, the regularity of a contour is an important property.

Based on the regularity assumption, the situations is much the same as encountered for the boundary terms in Section 3.2.1. Recall that the shift of the image of the boundary $\mathbf{m}(\partial\mathcal{G})$ results in a boundary term of the first derivative. The boundary $\partial\mathcal{G}$ is a curve. By the properties of integration, values at non-pathological curves do not contribute to an area integral. The values of V at $\partial\mathcal{G}$ do however contribute to $E'\boldsymbol{\mu}$ by a curve integral. They receive a special weight according to whether $\boldsymbol{\mu}$ causes the image of the reference domain $\mathbf{m}(\mathcal{G})$ to grow or shrink locally. Shifting a contour line has the same effect. However, as the contour line is shifted, the corresponding parts of the shadow line are shifted, too. From the viewpoint of the image domain $\mathcal{F}$ the contour line and the shadow line are the same line, as can be seen from Fig. 3.12. A shift of $\mathbf{m}(\mathcal{K})$ does not cause the extent of the modeled part to be changed. Loosely speaking, what is taken away on one side of $\mathbf{m}(\mathcal{K})$ is added on the other side.

A simple example shall illustrate the situation. Consider looking at a book with one eye only. The eye takes the role of the imaging device (image domain $\mathcal{F}$!). At the inner side of the contour of the book appears the texture of the book, at the outer side the texture of the background, maybe a table. Recall that the contour is consistently oriented so that it is possible to talk about the inner and outer side. If the shape or position of the book is changed a little bit, it suffices to shift the book a little bit, then, depending on the direction of the movement of the contour, there will be locally more texture of the book and less texture of the table or vice versa.

The boundary term of E', which is calculated in the reference domain $\mathcal{G}$, has to take the values of the local measure at the contour line $\mathcal{K}$ and at the shadow line $\mathbf{o}(\mathcal{K})$, i.e. at the entire boundary of the occluded part $\partial\mathcal{G}^O$ (cf. Fig. 3.12), into account. The position of $\mathcal{K}$ in the image $\mathcal{F}$ is determined by the values of the map $\mathbf{m}$ at $\mathcal{K}$. Consequently, the derivative E' comprises an additional curve integral over the contour line $\mathcal{K}$. The further details are omitted, as the argumentation is completely the same as in Section 3.2.1 leading to

$$\begin{aligned} E'(\mathbf{m})\boldsymbol{\mu} = &\int_{\partial\mathcal{G}} V \circ (\mathbf{id}_{\mathcal{G}}, \mathbf{m})\, (\mathbf{m}'^{\mathrm{ad}}\boldsymbol{\mu}) \times \mathrm{d}\mathbf{s} - \int_{\mathcal{G}^V} V_{\mathbf{x}}^T \circ (\mathbf{id}_{\mathcal{G}}, \mathbf{m})\, \mathbf{m}'^{\mathrm{ad}}\boldsymbol{\mu}\, \mathrm{d}A \\ &+ \int_{\mathcal{K}} \big(V \circ (\mathbf{id}_{\mathcal{G}}, \mathbf{m}) - V \circ (\mathbf{o}, \mathbf{m})\big)\, (\mathbf{m}'^{\mathrm{ad}}\boldsymbol{\mu}) \times \mathrm{d}\mathbf{s}\,. \end{aligned} \tag{3.5.11}$$

The integral over $\mathcal{K}$ also includes the values of the local measure V at $(\mathbf{o}(\mathcal{K}), \mathbf{m}(\mathcal{K}))$, i.e. at the shadow line. It otherwise is treated like the boundary $\partial\mathcal{G}$.

As occlusions cause an additional curve integral for the derivative of the similarity measure, it has to be expected that the derivative is discontinuous if parts of $\mathcal{G}$ are in a transition state between occluded and visible. Under these circumstances the regularity assumption is violated. Whether there is a possibility to circumvent the discontinuity, for example by means of additional smoothness requirements for $\mathbf{m}$, is not clear. Anyway, violations of the regularity condition indicate problematic situations. An example shall demonstrate the behavior of the similarity measure in the presence of occlusions, Fig. 3.13 and 3.14. It is based on a very simple setup. The texture g consists of a single bright spot with a linear drop off to either side. The map m is assumed piecewise linear and the node value, whose position coincides with the center of the bright spot, is shifted in both directions. This leads to a series of maps m_d parameterized by the disparity d of the central node. m_0 is equal to the identity mapping. As this is a trivial case it is not depicted in Fig. 3.13. For each map m_d the image $g \circ m_d^{-1}$ of the texture g is portrayed. A physical interpretation of each situation is also given

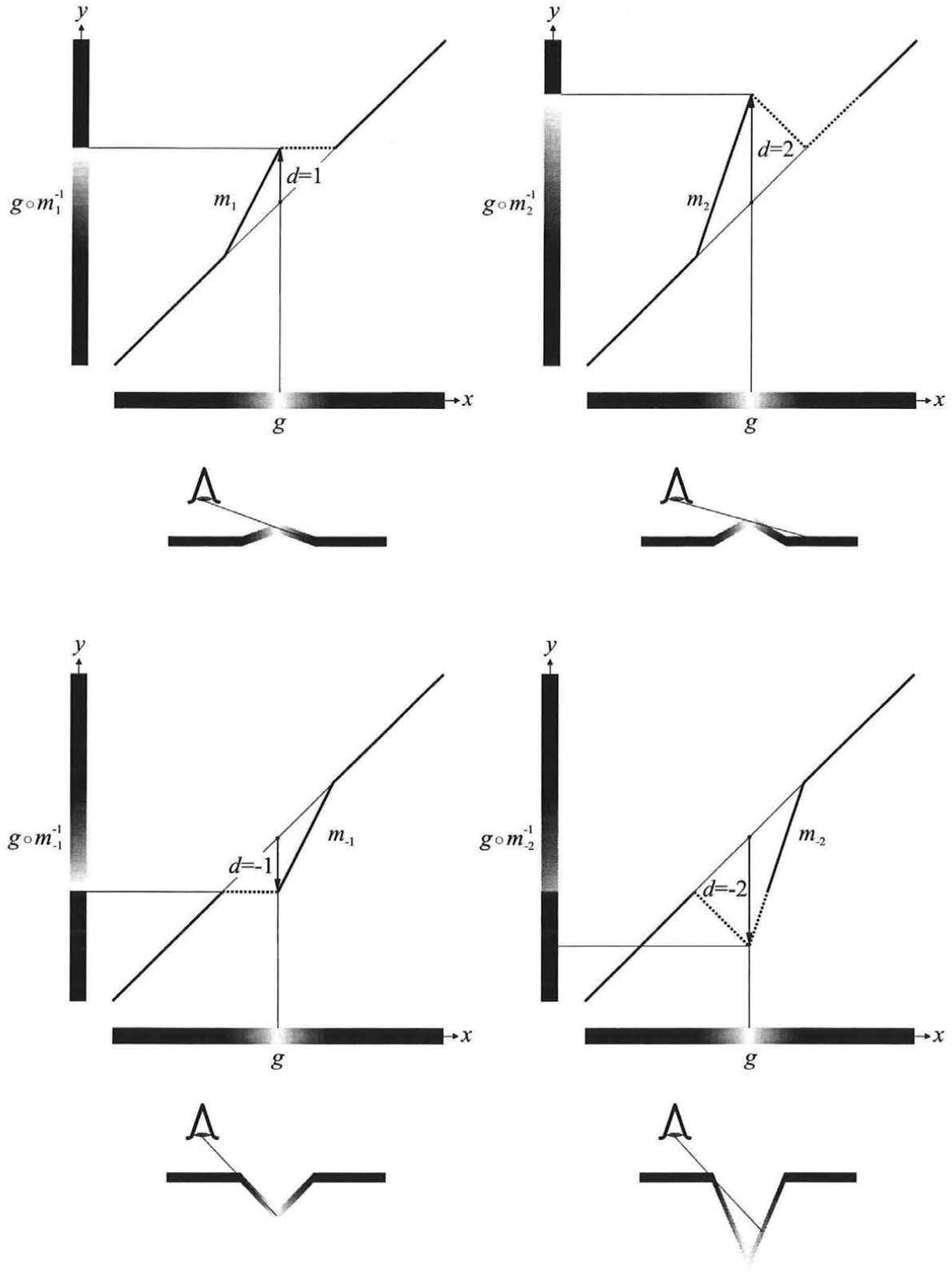

Figure 3.13: The texture g mapped to the image domain by $g \circ m_d^{-1}$ for a series of maps m_d. —— $m_d(x)$, $x \in G^V$, ····· $m_d(x)$, $x \in G^O$.

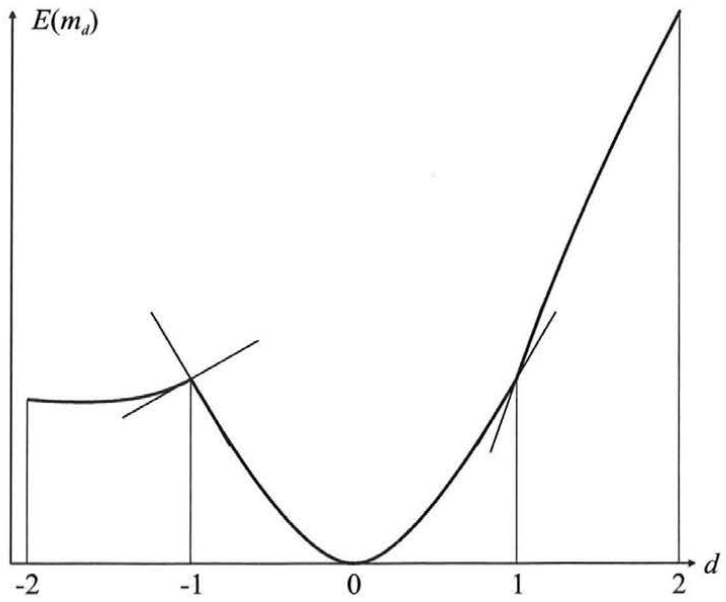

Figure 3.14: The variation of the image energy for the occlusion example. Left and right side tangents are plotted for the critical positions $d = -1$ and $d = 1$.

in Fig. 3.13. These are only a sketches because the disparity and the height are inversely proportional, thus a piecewise linear height does not exactly correspond to a piecewise linear disparity. The image function is assumed to equal $f \equiv g$ and thus $E(m_0) = 0$, as can be seen in Fig. 3.14. The similarity measure is based on the local least-squares measure (3.1.7). For negative values of d, the central node is shifted in the negative y-direction, which corresponds to deforming the surface downwards. This will lead to the critical situation m_{-1}. Shifting the node further, the left slope of the surface and the point of maximum brightness disappear. The brightness stripe indicating the intensity of $g \circ m_{-2}^{-1}$ therefore lacks the region of maximum brightness. Shifting the node in the positive y-direction will cause the right slope of the surface to disappear. The critical state is m_1, at which the transition from visible to invisible happens. The similarity measure Fig. 3.14 exhibits the expected discontinuities of the first derivative. Though without proof, we claim that the fact that f and g are assumed piecewise linear, resulting in only piecewise continuous differentiability of V, has no consequence for the differentiability of the similarity measure. This is ascertained by the fact that the corners of the graph of E appear only at the critical positions.

Chapter 4

Practical Aspects of the Matching Algorithm

The similarity measure constitutes a criterion for good correspondence with respect to intensity values. This chapter introduces the remaining components that are required to construct a matching algorithm. Firstly, there is no doubt that the similarity measure only incompletely specifies what is expected as the outcome of the matching algorithm. This issue is addressed by the modeling of the matching function. Secondly, an appropriate optimization algorithm has to be selected to determine the optimal configuration. All three components, the similarity measure, the modeling of the matching function, and the optimization method determine the eventual performance of the matching algorithm. It is important to separate these issues as clear as possible. The modeling of the matching function has however a significant influence on the performance of the optimization algorithm. In particular, it plays a role for the algorithms ability to avoid suboptimal local minima. The issues of modeling and optimization are thus not that clearly separated. Further, often no quantitative information about the properties of the matching function is available. The parameters of the model are thus adapted heuristically to achieve a reasonable output.

4.1 Modeling of the Matching Function

Intuitively it is clear that the similarity measure does not completely define optimal correspondence, as it is based on the comparison of single intensity values only. Some degree of coupling between the values $\mathbf{m}(\mathbf{x})$ and $\mathbf{m}(\tilde{\mathbf{x}})$ at neighboring positions $\mathbf{x}$ and $\tilde{\mathbf{x}}$ has to be introduced, as arbitrary irregular maps $\mathbf{m}$ can not be accepted as solution. Just consider that the forward similarity measure even requires the Jacobian to exist almost everywhere.

For the following, it is assumed that the correspondence problem is formulated within a particular geometric context, so that the matching function can be assumed of scalar type. The problem, as stated based on the similarity measure only, is given by

$$E_{\text{img}}(u^*) = \min! \,. \tag{4.1.1}$$

The usual notation of optimization, to denote the optimal configuration u^* with an attached asterisk, is adopted. The subscript "img" distinguishes the similarity measure from the regularization term, which is also referred to as deformation energy E_{def}. In its particular form the problem (4.1.1) is under-constrained. Two methods are used to incorporate additional information about the solution:

- *restricting the space of solutions* to a subspace of smooth functions and
- *regularization*, that is the penalization of non-smooth behavior of the solution.

They are both applicable to the variational formulation directly, as well as to the discretized analogue. In case of the forward similarity, it has been helpful to consider the continuous formulation in detail and to transfer the structure to the discretized case. For the modeling of the matching function the continuous formulation can be largely skipped.

4.1.1 Restricting the Space of Solutions

Restricting the space of solutions is simply done by composing u of, not necessarily orthogonal, basis functions b_i such as wavelets, B-splines, bilinear kernels etc. Examples of that strategy can be found in [22, 23, 58]. In [1] the method is considered as part of the optimization algorithm. In this approach, the number of basis functions is increased during the matching process to finally cover the whole space.

The mapping from the coefficient space to U, the space of all possible matching functions u, is a linear mapping denoted by B,

$$u = Bc = \sum_i c_i b_i \,. \tag{4.1.2}$$

The coefficients are denoted by c_i and are assumed to be real. The basis functions b_i are selected to exhibit some reasonable behavior, specifically they are selected to cover a subspace of "smooth" functions. Smoothness in this case refers to the resolution of the smallest details that are capture by the basis functions. Taking the familiar bilinear kernels, the resolution of u is controlled by the node spacing. The function will not be smooth in a mathematical sense, as it will always exhibit discontinuities of the first derivative. The principle is nevertheless the same as when using the Fourier basis and controlling the properties of the function u by the number of admitted frequencies. In this case, the effect is more in line with the intuitive notion of smoothness.

In the following we consider u on a grid $\mathcal{G}_h$ with the index set I. In this case $U = \mathbb{R}^I$ and (4.1.2) is essentially a mapping $\mathbb{R}^p \mapsto \mathbb{R}^I$. Thus the linear operator B can be identified with a matrix, $B = (b_1, b_2, \ldots, b_p)$. The basis functions b_i are elements of $\mathbb{R}^I$. If considered without an additional regularization functional, the dimension of the coefficient space should be smaller than $\#I$, the dimension of U. Otherwise (4.1.2) will just rotate the coordinate system of the problem to a new position leaving the problem unaltered.

For optimization purposes, the influence of the basis functions on the derivatives is of interest. Differentiating the similarity measure with respect to coefficient c_j leads to

$$\frac{\partial E_{\text{img}}}{\partial c_j} = E'_{\text{img}} b_j = \langle \nabla E_{\text{img}}, b_j \rangle \,. \tag{4.1.3}$$

Using a matrix style notation

$$\nabla_c E_{\text{img}} = B^T \nabla E_{\text{img}} \,. \tag{4.1.4}$$

B^T is the transposed operator of B.[1] The Gauss-Newton matrix transforms according to

$$E^{\text{GN}}_{\text{img},c} = B^T E^{\text{GN}}_{\text{img}} B \,. \tag{4.1.5}$$

[1] On the general notion of transposition of linear operators see [4].

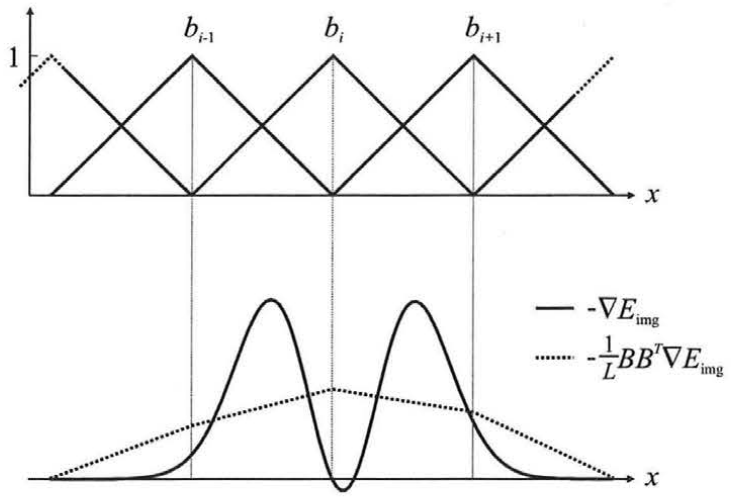

Figure 4.1: The smoothed negative gradient $-\frac{1}{L}BB^T\nabla E_{\text{img}}$ for the 1D example depicted in Fig. 3.4. The piecewise linear basis functions are plotted above.

The gradient with respect to the coefficient space is not easily interpreted. Mapping it to U by

$$B\nabla_c E_{\text{img}} = \sum_i \langle \nabla E_{\text{img}}, b_i \rangle b_i = BB^T \nabla E_{\text{img}} \tag{4.1.6}$$

reveals the influence of the basis functions on the gradient of the image energy term. If the basis consists of orthonormal components, the above equation describes the orthogonal projection of ∇E_{img} onto the range of B. This operation essentially smoothes the gradient. Employing for example the Fourier basis, BB^T is the matrix form of the ideal lowpass filter. The smoothing is not a consequence of the orthonormality of the basis functions, but is effected by other choices for the basis, too. The smoothed negative gradient $-\frac{1}{L}BB^T\nabla E_{\text{img}}$, as obtained using a piecewise linear basis for the example depicted in Fig. 3.4, is portrayed in Fig. 4.1. The basis functions are plotted in the upper part of the figure. The factor $\frac{1}{L}$ is introduced to obtain a result comparable to $-\nabla E_{\text{img}}$. This is necessary because the broader the basis functions are, the stronger is their influence on the matching function u and the larger is the gradient $\nabla_c E_{\text{img}}$. Dividing by L compensates for this effect. Clearly, the region where $-\nabla E_{\text{img}}$ points to the wrong direction is smoothed out. The consequence is a more stable performance of the matching algorithm. Of course, the missing matching information outside the central region, where $-\nabla E_{\text{img}}$ rapidly decays to zero, can not be compensated.

An important subclass of the considered method uses shifted versions of a prototype kernel, such as the bilinear kernel, or B-splines, as basis functions. Ignoring the treatment of boundaries, the parameter function can be written as

$$u_{\mathbf{i}} = \sum_{\mathbf{j}} b_{\mathbf{i}-L\mathbf{j}} \tilde{u}_{\mathbf{j}} \,. \tag{4.1.7}$$

The relation to the above defined basis functions is given by $b_{\mathbf{j}}(\mathbf{x_i}) = b_{\mathbf{i}-L\mathbf{j}}$. The coefficients c are renamed to $\tilde{u}$ to reflect the fact that they are given on a grid related to $\mathcal{G}_h$ the domain of u. The domain of $\tilde{u}$ has a reduced resolution. In particular using bilinear basis functions has quite the same effect as if the discretization had been based on samples of u on a coarser grid $\mathcal{G}_{Lh}$.

4.1.2 Regularization

Regularization is done by penalizing solutions that exhibit a non-smooth behavior. Instead of solving the original problem (4.1.1) the modified problem

$$E(u^*) = E_{\text{img}}(u^*) + \alpha E_{\text{def}}(u^*) = \min! \tag{4.1.8}$$

is solved. $E_{\text{def}}(u)$ is large for functions u that exhibit strong irregularities. It expresses the notion of what is assumed a reasonable solution and which solutions are unlikely. α determines the relative weight of the deformation energy. A meaningful specification of the amount of regularization has to take the magnitude of the image energy into account. This is the reason why we will usually express it in multiples of the variance σ_f^2 of the image function f.

The regularization method has already been mentioned in Section 2.4.1. In the framework of Markov random fields E_{def} specifies the a priori probability of u according to a Gibbs distribution (2.4.1). The concept is used in different contexts under different names. In adjustment calculus the concept is referred to as fictitious observations. The Markov random field framework is however more flexible. Regularization is the term that is commonly used together with ill-posed problems.

A particular important class of regularization terms are the Tikhonov regularizers and generalizations thereof [55, 56]. Some prominent examples are the membrane functional

$$E_{\text{def}}(u) = \frac{1}{2}\int_{\mathcal{G}} u_{x_1}^2 + u_{x_2}^2 \,\mathrm{d}A \tag{4.1.9}$$

and the thin plate functional

$$E_{\text{def}}(u) = \frac{1}{2}\int_{\mathcal{G}} u_{x_1x_1}^2 + 2u_{x_1x_2}^2 + u_{x_2x_2}^2 \,\mathrm{d}A\,. \tag{4.1.10}$$

A third popular variant is the squared Laplacian, which is closely related to (4.1.10) [18]. The above measures are the ones widely used for the problem of stereo correspondence and surface reconstruction [1, 2, 18, 26, 37, 40, 41].

Like the similarity measure these functionals have to be discretized for the implementation. For the standard regularizers the discrete analogues are well known. The desired properties are easily transferred to the discrete case. As an example the discretized version of the membrane functional (4.1.9) reads (cf. [54, pp. 197])

$$E_{\text{def}}(u) = \frac{1}{2}\sum_{\substack{(i,j)\in I\\ i\neq M}} a_j^2(u_{i+1,j} - u_{i,j})^2 + \frac{1}{2}\sum_{\substack{(i,j)\in I\\ j\neq N}} a_i^1(u_{i,j+1} - u_{i,j})^2\,. \tag{4.1.11}$$

The 1D analogue is called the string functional. It is used for 1D test configurations.

The great benefit of this type of models is that the smoothness constraint is imposed by means of local only interactions. Only pair cliques are used in the case of the membrane functional, yet a far range coupling is achieved. This coupling is undesired at edges and discontinuities as it leads to oversmoothing. Modifications of the above regularization terms are primarily targeted at an improvement of this behavior. The advantage of the local coupling is that the interaction can be weakened very easily by changing the weight of clique potentials along discontinuities in depth or slope. This method is used to adapt the a priori

model to fit the particular situation or even to formulate the model such that it is self-adaptive. In [55] continuity control functions are introduced to weaken the constraints where necessary. For a comprehensive overview on discontinuity adaptive models see [36].

The fact that regularizers based on quadratic forms are easier to handle than more general expressions has led to their widespread use. The discretized form of the deformation energy can be written in a matrix style,

$$E_{\text{def}}(u) = \frac{1}{2}\langle u, Ru\rangle = \frac{1}{2}\sum_{\mathbf{i},\mathbf{i}'\in I} u_{\mathbf{i}} R_{\mathbf{i}\mathbf{i}'} u_{\mathbf{i}'} \,. \tag{4.1.12}$$

For a quadratic Gibbs energy the associated probability distribution (2.4.1) is of Gaussian form with R denoting the inverse covariance kernel. This interpretation has to be viewed with criticism in that usually R has a non-trivial nullspace and thus is not invertible. For example in the case of the membrane functional the nullspace is the space of all functions $u \equiv \text{const}$. This is actually a desired property, since certain simple functions should be left unpenalized. Practically the deformation measure is never used alone, but together with data driven quantities, such as the similarity measure E_{img}. Together these terms determine the solution appropriately.

The simple quadratic structure (4.1.12) is of special benefit for optimization purposes. The matrix R directly provides the Hesse matrix of the regularization functional and the first derivative is obtained easily as well,

$$\frac{\partial E_{\text{def}}}{\partial u_{\mathbf{i}}} = \sum_{\mathbf{i}'\in I} R_{\mathbf{i}\mathbf{i}'} u_{\mathbf{i}'} \tag{4.1.13}$$

$$\frac{\partial^2 E_{\text{def}}}{\partial u_{\mathbf{i}} \partial u_{\mathbf{i}'}} = R_{\mathbf{i}\mathbf{i}'} \,. \tag{4.1.14}$$

The regularization method is also applicable in connection with basis functions. For bilinear basis elements it is quite natural to impose the regularization directly in terms of $\tilde{u}$ on the reduced grid. The combination of both methods might provide a compromise between reducing the problem dimension on the one hand and providing a means to steer the amount of smoothness locally on the other hand. However, the local adaptation of the smoothness can be done with the reduced resolution only. In [1] the Fourier basis or alternatively Daubechies wavelets [12] are used. The standard regularizers are related to the Sobolev norms and equivalent formulations with respect to the Fourier or Daubechies basis exist. The matrix R is diagonal with respect to these bases. Using the Fourier basis, the interactions are however not local. Accordingly, it is difficult to perform a local adaptation of the smoothness.

It has to be kept in mind that both methods, the restriction to a subspace of smooth functions and the regularization, follow the same aim. They determine the amount of smoothness of the final solution and thus the amount of information extracted from the correspondence problem. Using no regularization whilst employing basis elements with coarse resolution is not necessarily better than retaining the full resolution of the matching function but imposing a strong regularization and vice versa.

4.2 Optimization Methods

Posing the correspondence problem as an optimization problem, an appropriate algorithm has to be selected to find the minimum. There exists a variety of methods for the solution

of optimization problems. For an overview of algorithms used in Computer Vision see [36]. Each method imposes requirements on the type of objective function. Some are amenable to combinatorial optimization, i.e. each variable may assume values from a finite set only, others require the objective function to be differentiable. The similarity measure can be adapted to most requirements. Specifically, the derivative and the Gauss-Newton term have been computed to support efficient local optimization algorithms [30, 45]. Of course, we can not hope that the objective function is convex, rather multiple suboptimal local minima have to be expected. The standard paradigm when using local optimization methods for image matching is to imbed them in a multiresolution scheme [18, 22, 23, 37, 40, 41, 49]. There is no theoretical guarantee to find a global optimum by that strategy, as for global optimizers like simulated annealing. Nevertheless, the multiresolution approach has proven to effectively solve the problem of the limited region of convergence [22, 23]. In this section, we concentrate on the convergence properties of local methods. It is assumed that a sufficiently good initial approximation is supplied, either from prior knowledge or from the matching result of a lower layer of a multiresolution scheme. Other sources for an initial guess are global optimization algorithms like dynamic programming or simulated annealing. These methods are amenable to combinatorial optimization. If applied to continuous problems, the range of possible values has to be discretized. This has the disadvantage that the computational effort rises with the density of the discretization. Combining these methods with local optimization methods might be an effective way to avoid convergence problems and to achieve high accuracy at a reasonable computational cost.

The methods proposed in the following sections are not special purpose optimizers. The formulation is however adapted to the particular problem of image matching. The problem to be solved is

$$E(u^*) = E_{\text{img}}(u^*) + \alpha E_{\text{def}}(u^*) = \min! \,. \tag{4.2.1}$$

This expresses that the energy assumes its minimum for the optimal matching function u^*. u^* is a vector in $\mathbb{R}^I$, with I the index set of some rectangular grid. Local optimization methods determine the solution iteratively. Starting from an initial guess $u^{(0)}$ a sequence $u^{(k)}$ is computed that converges to u^*. This commonly calls for the repeated evaluation of the derivative at the sequence of intermediate results $u^{(k)}$. For the first derivative of the image energy term at $u^{(k)}$ a shorthand notation is used,

$$L_{\mathbf{i}}^{(k)} = \frac{\partial E_{\text{img}}(u^{(k)})}{\partial u_{\mathbf{i}}} \,. \tag{4.2.2}$$

The quadratic approximation of the similarity measure is based on the Gauss-Newton term. Analogously to the above we write

$$Q_{\mathbf{ii}'}^{(k)} = E_{\text{img}\mathbf{ii}'}^{\text{GN}}(u^{(k)}) \,. \tag{4.2.3}$$

The matching function may be composed of bilinear basis functions. In that case the index set I refers to the coarse grid. It is assumed that the derivatives take the basis functions into account by (4.1.4) and (4.1.5). If it is convenient to use a concrete example for the image energy, the single grid discretization using the local least-squares measure is used (A.0.1), (A.1.1). For the regularization term only quadratic expressions (4.1.12) are considered. This conforms to what is practically used in this work. Additionally, it facilitates the notation. The derivatives are given by (4.1.13) and (4.1.14). If a concrete model is needed, the membrane or string functional is used.

In the sequel two optimization methods are considered: the steepest descent method and the Gauss-Newton method, including a computationally less expensive variant. The Interactive Data Language IDL has been used to practically implement the algorithms. IDL performs most effectively if operations are performed in a vector style. This induces a tendency towards parallel schemes like the Jacobi iteration. It should be no problem to adapt the proposed methods to suit different situations.

4.2.1 The Method of Steepest Descent

A rather broad class of local minimization algorithms can be decomposed into two steps: the specification of a search direction $p^{(k)}$, and the determination of the step size $t_k > 0$. The recurrence formula is given by

$$u^{(k+1)} = u^{(k)} + t_k p^{(k)} \; . \tag{4.2.4}$$

The direction $p^{(k)}$ is required to be a descent direction, which necessitates that the variation with respect to $p^{(k)}$ is negative,

$$\delta E(u^{(k)}; p^{(k)}) < 0 \; . \tag{4.2.5}$$

By means of the gradient, the above condition can be written in the form

$$\langle \nabla E(u^{(k)}), p^{(k)} \rangle < 0 \; . \tag{4.2.6}$$

Using the method of steepest descent (SD) or gradient method, the search direction is chosen to equal the negative gradient, $p^{(k)} = -\nabla E(u^{(k)})$. With respect to the Euclidean norm in the parameter space, this direction has the steepest slope. For the reference problem, the recursion is given by

$$u_{\mathbf{i}}^{(k+1)} = u_{\mathbf{i}}^{(k)} - t_k \left(L_{\mathbf{i}}^{(k)} + \alpha \sum_{\mathbf{i}' \in I} R_{\mathbf{i}\mathbf{i}'} u_{\mathbf{i}'}^{(k)} \right) \; . \tag{4.2.7}$$

As can be seen from Fig. 3.4, the gradient of the image energy term dominantly points to a direction which is considered reasonable. For some regions the indicated direction is however wrong. Additionally, the absolute value of the gradient exhibits strong variations. Both mechanisms described in Section 4.1 provide a means to incorporate a priori knowledge about reasonable solutions and help to stabilize the performance of the optimization algorithm. The regularization term helps to find the correct solution by providing a force that keeps the parameter function u in a smooth shape. The requirements for optimization purposes are however more stringent than those for modeling are. For images with little corruption by noise, a weak regularization will ensure that the global optimum of the energy function does satisfy the expectations with respect to smoothness of the solution. For optimization purposes, it is often required that the a priori model also helps to avoid local minima. Similar to Graduated Non-Convexity (GNC) [7, 36], it should be possible to start with rather strong regularization, facilitating convergence, and to relax the requirements during an iterative process.

The use of basis functions stabilizes the performance of the matching algorithm, too. Like the regularization term, it introduces a coupling between neighboring values of the matching function u. The estimates of the values $u_{\mathbf{i}}$ thus depend on intensity values of an enlarged

neighborhood. (4.1.6) reveals that the introduction of basis functions essentially leads to a smoothing of the gradient. In view of Fig. 4.1, it is plausible that the search direction gets more robust. Similar as above, the requirements of optimization and those to obtain a good solution are different on principle. Amit [1] suggested to use a hierarchical basis and a coarse-to-fine strategy. He proposes to use the Fourier basis or some wavelet basis and to successively increase the number of basis functions. The method starts with a collection of coarse resolution basis functions and adds higher frequency elements during the process. In contrast to the common pyramidal schemes, the resolution of the images is kept fixed. The motivation is not to loose important matching information by resorting to coarse resolutions. The coarse-to-fine strategy is applied to the matching function only. As it turns out, Amit has to use coarse level images, too, to ensure convergence of the algorithm.

The influence of the modeling of the matching function is in any case twofold. Firstly, it affects the performance of the optimization algorithm. Secondly, it accounts for the proper balancing between the matching information provided, determined by the resolution of the input images, and the information extracted, determined by the amount of smoothing of the matching function. We have not addressed this problem separately. When working with different levels of a pyramid, the relative weighting of data term and regularization term is kept constant. Likewise, the resolution of the matching function relative to the resolution of the images is left unchanged for all levels.

The gradient direction plays an outstanding role for all considered optimization methods. Algorithms that are more elaborate modify the gradient direction to achieve faster convergence. If the gradient points to a "wrong" direction, problems will arise for any of the considered methods. Using bilinear basis functions with a node spacing of 8 and a weak regularization of typically $\alpha = 0.1\sigma_f^2$ has been found sufficient to achieve a rather stable convergence to a reasonable solution in many cases. As the use of basis functions leads to a smoothing of the gradient, a previous work [40, 41] proposed to replace the smoothing operation by a positive semi-definite convolution operator. The motivation is that bilinear basis functions introduce a location dependency of the result. The outcome depends on the positioning of the grid of the basis function relative to the grid of the discretization. Convolution operators are shift invariant, thus every pixel in the reference space is treated equally. However, the convergence properties have been found not convincing for some situations. As the efficient optimization is still considered a non-trivial topic, it has been found convenient to resort to situations that are more standard. Further, the convolution approach is computationally more expensive. The idea is not pursued herein.

4.2.2 The Powell-Wolfe Stepsize Rule

For the gradient descent method, appropriate stepsizes t_k have to be selected. The gradient itself gives no hint how to determine the stepsize. t_k can either be chosen based on problem specific considerations or using a stepsize strategy. See [30] for a couple of possible stepsize rules. The strategies usually can be used to determine a reasonable stepsize for any search direction $p^{(k)}$. Their applicability is not limited to the gradient direction. The criterion used herein is the Powell-Wolfe rule. The stepsize t_k is chosen such that for some $\sigma \in]0, \frac{1}{2}[$ and $\beta \in]\sigma, 1[$ the following conditions hold (cf. also [45]): Firstly,

$$E(u^{(k+1)}) \leq E(u^{(k)}) + \sigma t_k \langle \nabla E(u^{(k)}), p^{(k)} \rangle \,, \tag{4.2.8}$$

the decrease of the objective function is within a certain range of what is predicted by means of the linearization. This prevents too large values for t_k. Secondly,

$$\langle \nabla E(u^{(k+1)}), p^{(k)} \rangle \geq \beta \langle \nabla E(u^{(k)}), p^{(k)} \rangle \,, \tag{4.2.9}$$

which prevents too small stepsizes t_k. In [30] an algorithm for finding a stepsize that matches the above condition is given. A flowchart is shown in Fig. 4.2. It requires the specification of a reduction/magnification factor $a \in]0,1[$ and an initial guess t_0 of the stepsize. $a = 0.5$ has been used in this work. The subprocedures too_large and too_small check the conditions (4.2.8) and (4.2.9) respectively. For the Powell-Wolfe rule to be realizable, the derivative has to be continuous and furthermore exactly known. As stressed in Section 3.4.2, E'^h_{img} is not the exact derivative of E^h_{img}. It nevertheless is used for the practical computations. Problems are encountered near the optimum solution. The stepsize search has to be checked for convergence and terminated if necessary. This aspect is not considered by the flowchart.

The gradient method is known for its sometimes slow convergence. If the optimization problem is quadratic $f(x) = \frac{1}{2}x^T Ax + b^T x + c$, the ratio of the largest and smallest eigenvalues of the matrix A determines the speed of convergence. The larger $\frac{\lambda_{max}}{\lambda_{min}}$ the slower proceeds the algorithm. Very different eigenvalues indicate that the objective function varies rapidly along certain directions whilst varying relatively slowly along others. In this case, the level curves of the objective function are elongated ellipses. Provided the second derivative is continuous, general optimization problems are similar to quadratic problems in the neighborhood of the optimum solution, thus the reasoning can be carried over to the general case. For the matching problem, the texture of the images predominately determines the shape of the objective function. For highly textured areas with steep slopes of the intensity function, small stepsizes suffice to achieve a good reduction of the image energy and to reach the minimum fast. For areas with poor texture, an acceleration by using larger stepsizes is necessary. For images containing highly and poorly textured parts, the stepsize strategy leads to a compromise. It is observed that the updates $t_k p^{(k)}$ show oscillatory behavior for highly textured areas whilst little progress is made for poorly textured areas. The methods described in the next section do not exhibit this behavior, or at least mitigate it.

4.2.3 The Gauss-Newton Method

The Gauss-Newton (GN) [6, 30] method may be considered "the" method to solve all kinds of overdetermined systems in geodesy. The basic concept is briefly reviewed in the following. Though applicable to general error norms [52], the Gauss-Newton method for the Euclidean norm is considered only. The traditional setup in adjustment calculus is given by a set of equations

$$\mathbf{f}(\mathbf{x}) = \mathbf{l} + \mathbf{v} \,. \tag{4.2.10}$$

$\mathbf{x} \in \mathbb{R}^m$ are the unknown parameters of the system model $\mathbf{f}$. The observations are denoted by $\mathbf{l} \in \mathbb{R}^n$, $n > m$, and $\mathbf{v} \in \mathbb{R}^n$ accounts for the discrepancies of the overdetermined system. An estimate of $\mathbf{x}$ is obtained as the minimum solution $\mathbf{x}^*$ of the least-squares problem,

$$\frac{1}{2}\mathbf{v}^T\mathbf{v} = \frac{1}{2}\|\mathbf{l} - \mathbf{f}(\mathbf{x}^*)\|^2 = \min! \,, \tag{4.2.11}$$

with $\|.\|$ denoting the Euclidean norm. Whilst given by a set of equations, the mathematical problem is eventually an optimization problem. Nonlinear systems are linearized at an initial

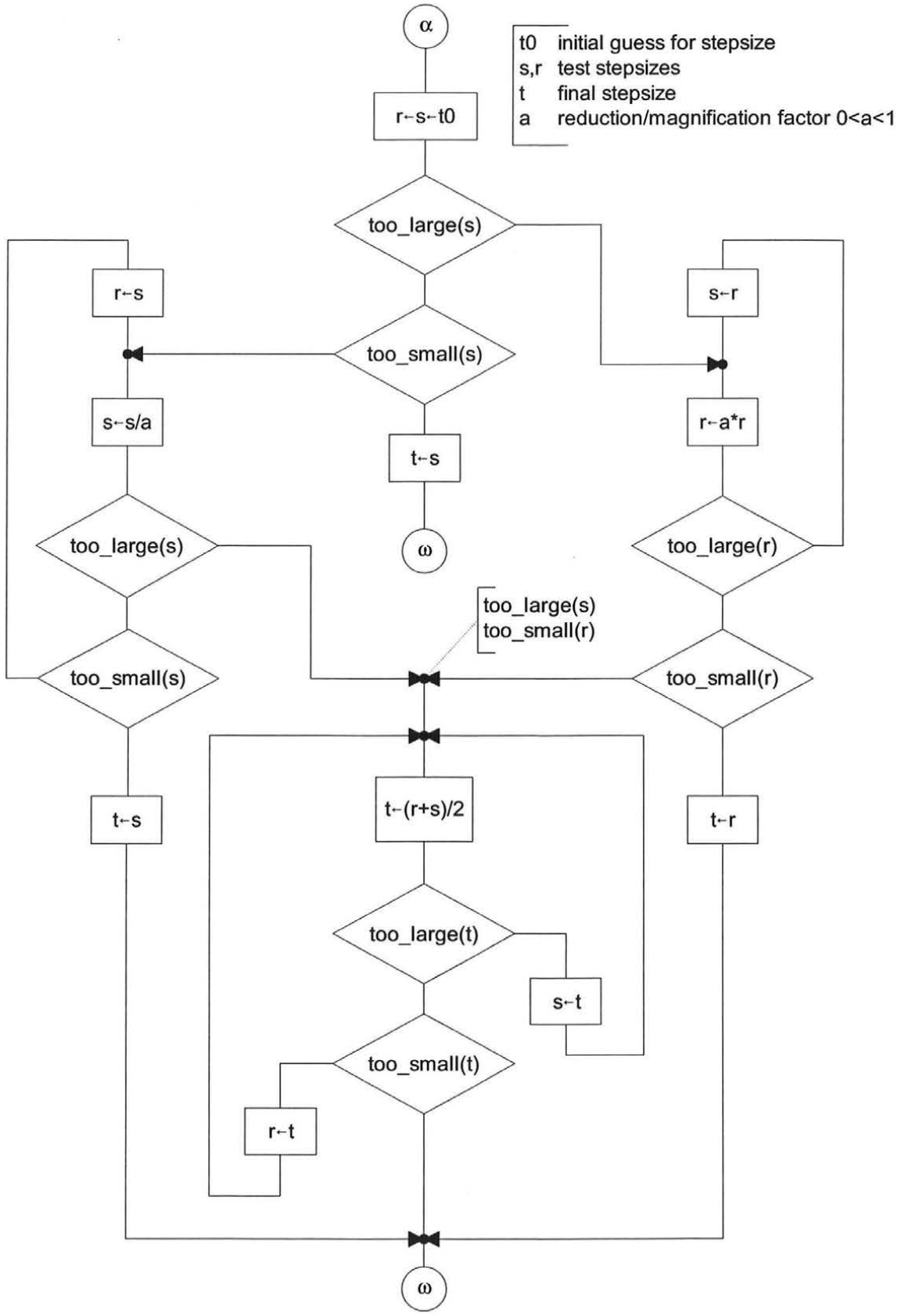

Figure 4.2: Flowchart for the realization of the Powell-Wolfe stepsize rule.

guess $\mathbf{x}^{(0)}$ of the solution. The resulting overdetermined linear system leads to a quadratic optimization problem according to the same principle as above. It is solved to obtain a hopefully better approximation of the solution. This procedure is applied repeatedly until convergence is achieved subject to some termination criterion. Let the Jacobian matrix of $\mathbf{f}$ at $\mathbf{x}^{(k)}$ be denoted by $\mathbf{A}^{(k)}$, $\mathbf{A}_{ij}^{(k)} = \frac{\partial f_i(x^{(k)})}{\partial x_j}$. The linearized system is given by

$$\mathbf{f}(\mathbf{x}^{(k)}) + \mathbf{A}^{(k)}(\mathbf{x} - \mathbf{x}^{(k)}) = \mathbf{l} + \mathbf{v}^{(k)} \; . \tag{4.2.12}$$

The objective function is approximated by

$$\frac{1}{2}\|\mathbf{l} - \mathbf{f}(\mathbf{x})\|^2 \approx \frac{1}{2}\|\mathbf{l} - \mathbf{f}(\mathbf{x}^{(k)}) - \mathbf{A}^{(k)}(\mathbf{x} - \mathbf{x}^{(k)})\|^2 \; . \tag{4.2.13}$$

More explicitly,

$$\begin{aligned} \frac{1}{2}\|\mathbf{l} - \mathbf{f}(\mathbf{x})\|^2 \approx \frac{1}{2}\|\mathbf{l} - \mathbf{f}(\mathbf{x}^{(k)})\|^2 \\ - \big(\mathbf{l} - \mathbf{f}(\mathbf{x}^{(k)})\big)^T \mathbf{A}^{(k)}(\mathbf{x} - \mathbf{x}^{(k)}) \\ + \frac{1}{2}(\mathbf{x} - \mathbf{x}^{(k)})^T \mathbf{A}^{(k)T} \mathbf{A}^{(k)}(\mathbf{x} - \mathbf{x}^{(k)}) \; . \end{aligned} \tag{4.2.14}$$

The minimum is taken as the next value $\mathbf{x}^{(k+1)}$ of the iteration,

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + (\mathbf{A}^{(k)T}\mathbf{A}^{(k)})^{-1}\mathbf{A}^{(k)T}\big(\mathbf{l} - \mathbf{f}(\mathbf{x}^{(k)})\big) \; . \tag{4.2.15}$$

This is the update rule for the Gauss-Newton method. Relating the terms in the formula with the derivatives of the objective function $\frac{1}{2}\|\mathbf{l} - \mathbf{f}(\mathbf{x})\|^2$,

$$\left(\frac{1}{2}\|\mathbf{l} - \mathbf{f}(\mathbf{x})\|^2\right)' = -\mathbf{A}^T\big(\mathbf{l} - \mathbf{f}(\mathbf{x})\big) \tag{4.2.16}$$

and

$$\left(\frac{1}{2}\|\mathbf{l} - \mathbf{f}(\mathbf{x})\|^2\right)'' = \mathbf{A}^T\mathbf{A} - \sum_i f_i''(\mathbf{x})\big(l_i - f_i(\mathbf{x})\big) \; , \tag{4.2.17}$$

reveals the analogy with other optimization methods. The first derivative (4.2.16) appears in the linear part of (4.2.14), just as it would in a Taylor polynomial. The quadratic term of (4.2.14) only incompletely uses the second derivative (4.2.17). In contrast the Newton method for function minimization is based on the complete quadratic term of the Taylor polynomial. The Gauss-Newton step (4.2.15) may thus be interpreted as Newton step with "incomplete" second derivative. Herein we referred to $\mathbf{A}^T\mathbf{A}$ as Gauss-Newton term. Commonly it is easier to calculate than the Hesse matrix. If the residuals $l_i - f_i$ are small, the Gauss-Newton term dominates (4.2.17). If they are zero at $\mathbf{x}^*$, quadratic convergence of the algorithm is achieved [30, 52].

The Gauss-Newton Method for Image Matching

For the reference problem, the quadratic approximation (4.2.14) of the energy function is given by

$$E(u) \approx E(u^{(k)}) + \sum_{\mathbf{i}\in I} L_{\mathbf{i}}^{(k)}(u_{\mathbf{i}} - u_{\mathbf{i}}^{(k)}) + \frac{1}{2}\sum_{\mathbf{i},\mathbf{i}'\in I}(u_{\mathbf{i}} - u_{\mathbf{i}}^{(k)})Q_{\mathbf{i}\mathbf{i}'}^{(k)}(u_{\mathbf{i}'} - u_{\mathbf{i}'}^{(k)}) + \frac{\alpha}{2}\sum_{\mathbf{i}'\in I} u_{\mathbf{i}}R_{\mathbf{i}\mathbf{i}'}u_{\mathbf{i}'} \; . \tag{4.2.18}$$

The minimum condition for the approximated problem is given by

$$L_{\mathbf{i}}^{(k)} + \sum_{\mathbf{i}' \in I} Q_{\mathbf{i}\mathbf{i}'}^{(k)} (u_{\mathbf{i}'} - u_{\mathbf{i}'}^{(k)}) + \alpha \sum_{\mathbf{i}' \in I} R_{\mathbf{i}\mathbf{i}'} u_{\mathbf{i}'} = 0 \qquad \mathbf{i} \in I \,. \tag{4.2.19}$$

This is the Gaussian normal equation for the matching problem, which is obviously a linear equation. The matrix of the system is sparse. For typical grid sizes, the dimension of the problem $\#I$ will easily reach ranges that preclude the use of direct matrix solvers. There exist effective iterative solvers for high dimensional sparse systems, such as the well known Jacobi and Gauss-Seidel relaxation. See [21] for a comprehensive treatment. If the linear system is solved by an iterative procedure, the optimization effectively consists of two cascaded iterations. The system of equations is subject of iterative refinement as the procedure approaches the minimum and the system of equations is solved iteratively itself. This rises questions on how to effectively interleave the two procedures. It hardly makes sense to solve (4.2.19) with extreme precision, considering that the next linearization step will change the equation to be solved. A particular simple strategy is to interleave every linearization step with only one Jacobi step for the linear system. The concept is very similar to the SOR-Newton method described in [63, p. 1140]. The second derivative is however replaced by the Gauss-Newton term. Further we use a Jacobi like iteration rule, but that primarily facilitates analysis and is not fundamental to the concept. The strategy is to keep all values $u_{\mathbf{i}}$, but one, fixed to the current state of iteration $u^{(k)}$. (4.2.19) is then a one-dimensional linear equation. The solution is taken as the next value of the iteration. This recipe is applied to all values $u_{\mathbf{i}}$, $\mathbf{i} \in I$. Including a relaxation parameter this leads to

$$u_{\mathbf{i}}^{(k+1)} = u_{\mathbf{i}}^{(k)} - w \frac{1}{Q_{\mathbf{i}\mathbf{i}}^{(k)} + \alpha R_{\mathbf{i}\mathbf{i}}} \left(L_{\mathbf{i}}^{(k)} + \alpha \sum_{\mathbf{i}' \in I} R_{\mathbf{i}\mathbf{i}'} u_{\mathbf{i}'}^{(k)} \right) \,. \tag{4.2.20}$$

Herein we call this recursion the Gauss-Newton-Jacobi (GNJ) method. Setting $w = 1$ yields the results according to the above description. The relaxation parameter is selected either to accelerate the method or to avoid oscillations. If R is determined by the membrane functional (4.1.11), $w = \frac{1}{2}$ ensures convergence. w serves as stepsize parameter for the method. The selection of the stepsize is less difficult than for the steepest descent method. w primarily corrects the stepsize, which is intrinsically computed by Newton like methods. The steepest descent method in contrast gives no hint how to choose the stepsize, necessitating a stepsize strategy or additional knowledge about the problem. Further, a comparison with (4.2.7) reveals that in (4.2.20) the global stepsize t_k is replaced by a spatially varying factor. As can be seen from Section A, $Q_{\mathbf{i}\mathbf{i}}^{(k)} = E_{\mathrm{img}\mathbf{i}\mathbf{i}}^{\mathrm{GN}}(u^{(k)})$ is proportional to the squared image derivative. Accordingly, the update is adapted to the local texture. An implementation detail concerns the boundary terms of the first derivative of the forward similarity. As already indicated in Section 3.2.2, if the GN or GNJ method are used in connection with the forward similarity, it is recommendable to omit the boundary terms of the first derivative from $L_{\mathbf{i}}^{(k)}$.

The GNJ method has an interesting relation with a rather popular method for solving matching problems, which for example can be found in [18, 26, 37]. The argumentation commonly starts from the discrete Euler-Lagrange equation, which is written more explicitly for the purpose,

$$\frac{\partial E_{\mathrm{img}}(u^*)}{\partial u_{i,j}} + \alpha(4u_{i,j}^* - u_{i+1,j}^* - u_{i-1,j}^* - u_{i,j+1}^* - u_{i,j-1}^*) = 0 \,. \tag{4.2.21}$$

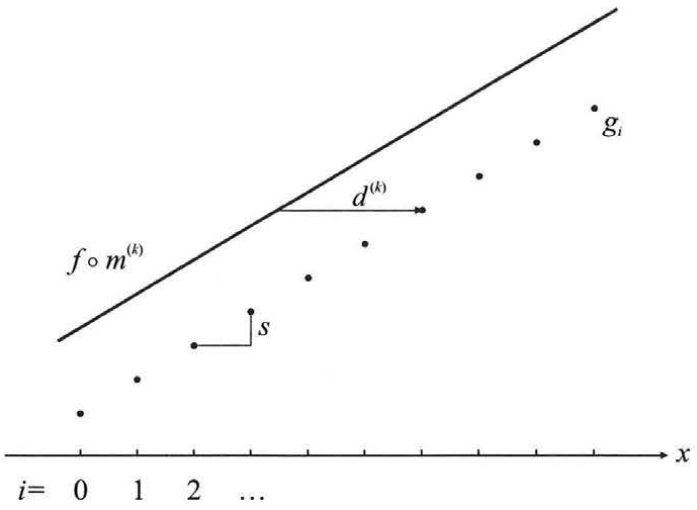

Figure 4.3: A trivial 1D matching constellation.

This is used to construct the following recurrence formula. Given a state of iteration $u^{(k)}$, $u_{i,j}^{(k+1)}$ is determined by

$$u_{i,j}^{(k+1)} = \frac{1}{4}(u_{i+1,j}^{(k)} + u_{i-1,j}^{(k)} + u_{i,j+1}^{(k)} + u_{i,j-1}^{(k)}) + \frac{1}{4\alpha}\frac{\partial E_{\text{img}}(u^{(k)})}{\partial u_{i,j}} . \tag{4.2.22}$$

Similar expression are found for other regularizing functionals. The fix point of this system of equations is obviously a solution of (4.2.21). Variations exist in the way the update is performed. Parallel for all indices I, corresponding to the Jacobi iteration, or sequentially like the Gauss-Seidel iteration. Despite its popularity and the obviously interesting results obtained, the justification of this method is very questionable. Using the abbreviated notation, the update rule can be written as

$$u_{\mathbf{i}}^{(k+1)} = u_{\mathbf{i}}^{(k)} - \frac{1}{\alpha R_{\mathbf{ii}}}\left(L_{\mathbf{i}}^{(k)} + \alpha \sum_{\mathbf{i}' \in I} R_{\mathbf{ii}'} u_{\mathbf{i}'}^{(k)}\right) . \tag{4.2.23}$$

Taking into account that $R_{\mathbf{ii}}$ is constant in the interior of the grid, the recursion is readily identified as gradient update equation (4.2.7) with stepsize $t_k = \frac{1}{\alpha R_{\mathbf{ii}}}$. The method thus proceeds along the gradient direction. It has to be emphasized that linking the stepsize with the regularization parameter α has no justification. The effect is most likely that α has to be adjusted to achieve convergence, which in turn will influence the smoothness of the matching function u. Comparison with (4.2.20) reveals that the term $Q_{\mathbf{ii}'}^{(k)}$, which locally adapts the stepsize, is missing. In other words, the quadratic term of the image energy term is ignored in (4.2.18). The introduction of a relaxation parameter w in (4.2.23) loosens the coupling between the weight of the regularization term and the stepsize. The resulting method is equivalent to the steepest descent method (4.2.7). The following example demonstrates the difference between the GNJ method and (4.2.23).

Example. We assume the trivial 1D constellation $f(x) = sx$ and $g_i = si$ (Fig. 4.3). The graphs $f(x)$ and g_i are straight lines with slope s. The best match is found if every point $x \in \mathcal{G}$ is mapped to $y = x \in \mathcal{F}$. The map m_i is parameterized via the disparity, $m_i = i + d_i$,

thus for $d_i \equiv 0$ perfect matching is achieved. Starting from $d_i^{(0)} \equiv d^{(0)}$, the GNJ method and (4.2.23) both produce a sequence $d_i^{(k)} \equiv d^{(k)}$, i.e. the disparity is constant for all indices at each stage of iteration. The linear and quadratic terms are given by (A.0.1). For the particular constellation,

$$L_i^{(k)} = (f(i + d_i^{(k)}) - g_i)\partial g_i = s^2 d^{(k)} \tag{4.2.24}$$

$$Q_{ii}^{(k)} = \frac{(\partial g_i)^2}{\partial m_i} = s^2 \,. \tag{4.2.25}$$

As regularization functional the string functional is chosen,

$$\sum_{i' \in I} R_{ii'} d_{i'}^{(k)} = 2d_i^{(k)} - d_{i+1}^{(k)} - d_{i-1}^{(k)} = 0 \tag{4.2.26}$$

$$R_{ii} = 2 \,. \tag{4.2.27}$$

As usual, boundaries are not considered explicitly. For the particular simple situation, the relaxation parameter can be assumed $w = 1$. Defining $\gamma = \frac{s^2}{2\alpha}$, the sequence generated by the GNJ method is given by

$$d^{(k+1)} = \frac{1}{1+\gamma} d^{(k)} \,, \tag{4.2.28}$$

whereas (4.2.23) will produce

$$d^{(k+1)} = (1-\gamma) d^{(k)} \,. \tag{4.2.29}$$

Obviously, the GNJ method produces a convergent sequence for all $\alpha > 0$, whilst the sequence generated by (4.2.23) converges only if $\alpha > \frac{s^2}{4}$. Clearly, the second method forces a minimum on the regularization parameter α, no matter how much regularization is actually needed to achieve a "smooth" result. By the way, the GN method solves this problem in one step, simply because the functions f and g are assumed linear and the problem thus is a linear least-squares problem.

The above example demonstrates the advantage of the GNJ method compared with the relaxation scheme (4.2.23). Examples for real imagery are not given, because the convergence of the latter method has been found too problematic. Using the membrane functional (4.1.11) and typical values of α, the Powell-Wolfe rule produces stepsizes for the steepest descent method that are much smaller than $\frac{1}{\alpha R_{ii}}$. Thus for the relaxation scheme (4.2.23) convergence is impossibly achieved. Further, the range of α, for which the method performs well, is influenced by the width of the bilinear basis functions. The average magnitude of the derivative $L_i^{(k)}$ rises with the node spacing L. Therefore, reasonable stepsizes for the steepest descent method do depend on the width of the basis functions, which in turn has influence on the selection of α, as the stepsize is linked to α in (4.2.23). However, it is not necessary to adapt α to the grid spacing hL in order to keep the relative weighting of regularization term and image energy unchanged. The formula of the membrane functional (4.1.11) is independent of the grid spacing. Altogether, the properties of the relaxation scheme (4.2.23) have not been found promising so that further experiments have not been considered worthwhile.

4.2.4 Numerical Tests and Remarks on Performance

It is difficult to judge the performance of matching algorithms on a theoretical basis only. Practical testing and theoretical explanations of the observed behavior complement each other. 1D constellations are particularly useful for numerical testing. They are not only easier and faster implemented, they also do facilitate analysis, generation, and manipulation of data. Using artificially generated data, the influence of certain parameters on the matching result and on the convergence properties can be investigated. In the following, different optimization strategies and different variants of the similarity measure are considered. An exhaustive test series that comprises all permutations of possible arrangements is however not provided. The purpose is twofold. Firstly, some general observations concerning the performance of the matching algorithms are mentioned. Secondly, we want to raise confidence in the proposed matching algorithm, which uses the forward modeling similarity measure. The analysis of the forward similarity requires more effort, in comparison with the reverse similarity. The latter can be formulated in the classical adjustment setup (cf. Section 3.4.2), which may be more familiar to some researchers. Both variants are however very closely related. With respect to the performance of optimization algorithms, comparable results are achieved.

For the 1D case, it is no problem to apply the Gauss-Newton method involving the direct solution of the Gaussian normal equation (4.2.19). As the GNJ method has been rationalized as Gauss-Newton method interleaved with one Jacobi step, shortcomings may be attributed either to the approximation (4.2.18) or to the Jacobi iteration. Solving (4.2.19) directly separates these aspects. The following tests take three optimization methods into account: the Steepest-Descent method in combination with the Powell-Wolfe stepsize rule (SD-PW), the Gauss-Newton method (GN), and the Gauss-Newton-Jacobi method (GNJ). The different variants of the similarity measures are all based on the local least-squares measure. See Appendix A for detailed formulas. Three kinds of discretizations of the forward energy are considered: single and double grid discretization, and a Finite-Elements (FE) variant assuming m, f, and g piecewise linear, as described in Section 3.4.1. The reverse similarity is used as an additional possibility. For the tests, a random function f with standard deviation normalized to $\sigma_f = 1$ has been generated and the template has been derived by $g = f \circ m$, as depicted in Fig. 4.4. Choosing m equal to the identity map has been considered too special. The particular m deviates from identity map by a weak cosine shaped modulation. The selection of m is not immaterial, in particular it is not for the forward modeling similarity measure. The dominant aspect for the performance of the matching algorithm is however the starting value or initial guess $m^{(0)}$. It is set to $m^{(0)} = m + 1$ corresponding to a shift of 1 sampling interval or 1 pixel. This appears to be rather conservative. Fig. 4.4 shows the initial alignment of g and $f \circ m^{(0)}$. It also illustrates the high frequency variations of the intensity functions. In fact, $m^{(0)}$ is located near the extremum of the pull in range of the matching algorithm. Choosing the initial shift to equal ± 2 pixel results in divergence, independent of the optimization method and similarity measure used. The properties of the map m are left unaltered for the tests. It is parameterized via the disparity, which is composed of piecewise linear elements with node spacing 8. The string functional (1D membrane functional) is used to impose a weak regularization ($\alpha = 0.05$). All tests in this section concern the convergence behavior of specific constellations. To that end the distance of $m^{(k)}$ to the final solution m^* is measured by $\|m^{(k)} - m^*\|_\infty$. The solution m^* that is obtained by the matching procedure can not be assumed equal to the map m used to artificially generate the image g. m^* deviates from m because of the influence of the discretization and the regularization. It has to be determined numerically with sufficient accuracy allowing the comparison with the k'th state

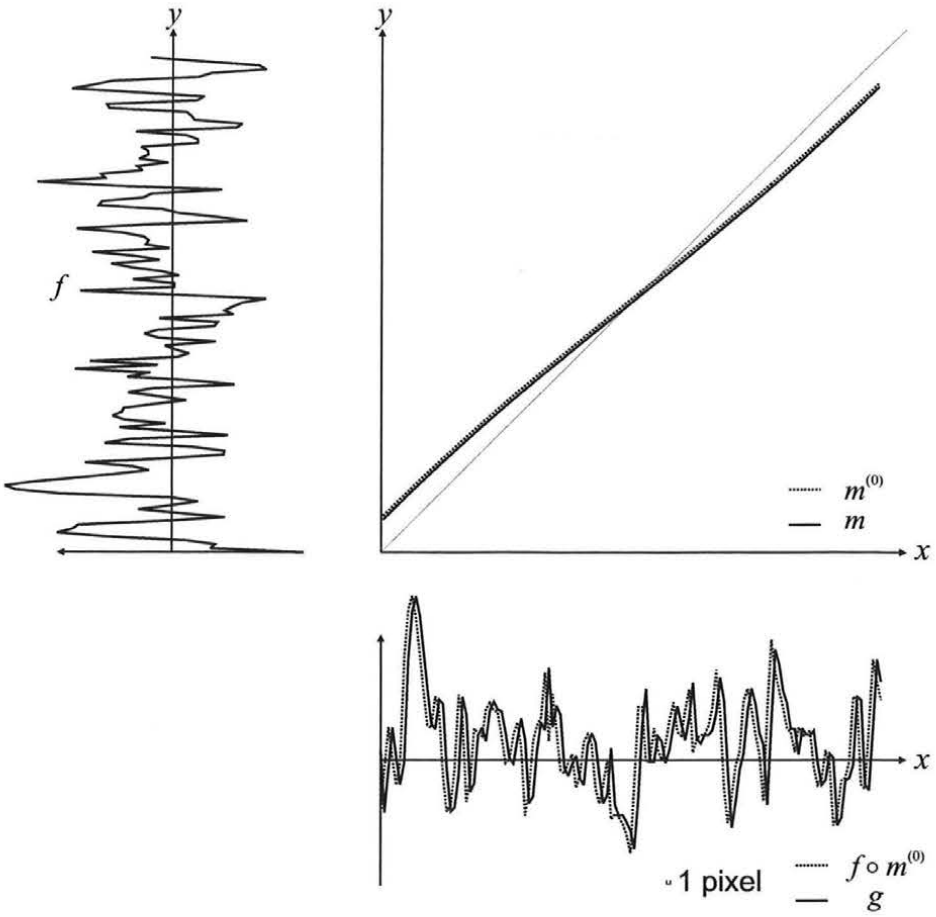

Figure 4.4: An artificially generated 1D matching example, $g = f \circ m$. The initial map $m^{(0)}$ and the initial alignment of g and $f \circ m^{(0)}$ are plotted, too.

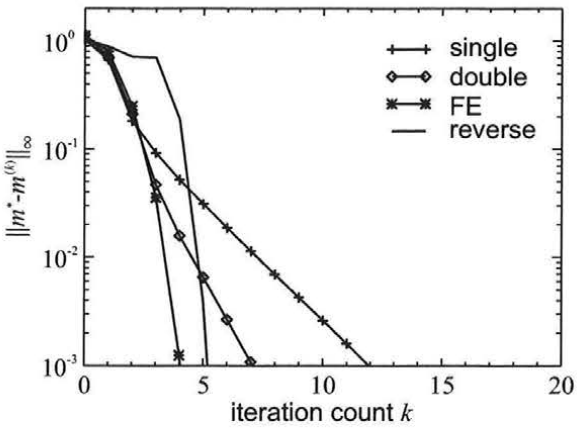

Figure 4.5: Convergence behavior for different similarity measures using the GN method.

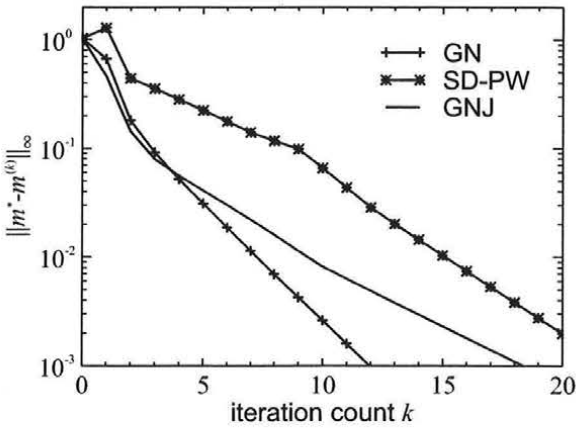

Figure 4.6: Convergence behavior using the single grid discretization together with different optimization methods.

of the iteration $m^{(k)}$.

Fig. 4.5 illustrates the convergence behavior of the Gauss-Newton method for different variants of the similarity measures. For the reverse similarity and the FE discretization of the forward similarity, rapid convergence can be observed. The slow initial progress for the reverse similarity is an indication of the unstable performance for the particular setup. Enlarging the initial offset $m^{(0)} - m^*$ to approximately 1.5 pixel already causes the algorithm to converge to an erroneous solution. Once near the solution, the minimum is reached quickly. For single and double grid discretization the initial convergence is fast. The final settling of the iteration is not that rapid. As has to be expected, the double grid discretization performs better than the single grid discretization. The required subpixel precision is in both cases reached within few iterations. In particular one is content with a precision of 0.1 pixel. The single grid discretization is the one which has the highest relevancy for the practical applications. In Fig. 4.6 it is used to illustrate the performance of different optimization algorithms. The GNJ method achieves the same initial progress as the GN method, though the eventual rate is smaller. The SD-PW method reaches the range of 0.1 pixel deviation within 10 iterations. This can still be considered reasonable. The tests correctly reveal the relative performance, though for real imagery the convergence is strongly influenced by problem areas causing the performance to deteriorate.

The configuration, which has been considered so far, involves no corruption by noise or other disturbances. Fig. 4.7 and 4.8 demonstrate the influence of a constant offset o. Image f is replaced by $f + o$. This test focuses on the influence of the residuals $f \circ m^* - g$, which are enlarged by the offset. Two different values $o = 0.05$ and $o = 0.1$ are used. For real imagery, the residuals are enlarged by noise, outliers, untreated occlusions, or uncompensated radiometric differences. The GN method is used together with the single grid discretization (Fig. 4.7) and with the reverse similarity (Fig. 4.8). For the single grid discretization the performance slightly deteriorates, whereas using the reverse similarity the iteration does not converge. Two effects play a role. Increasing the residuals causes the Gauss-Newton term to be a worse approximation of the second derivative and the convergence properties to get worse (cf. (4.2.17) and the subsequent explanations). This can be seen from Fig. 4.7. It is more clearly observed for the 2D implementations described in Chapter 5. Especially large areas with enlarged residuals have a dominant influence on the convergence behavior. An

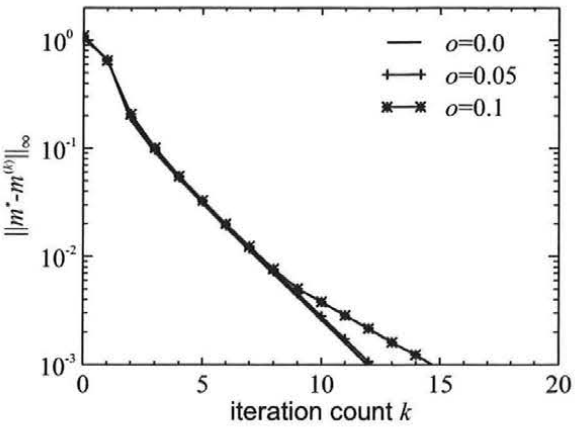

Figure 4.7: Convergence behavior using the single grid discretization and the GN method. The residuals are enlarged by replacing f by $f + o$.

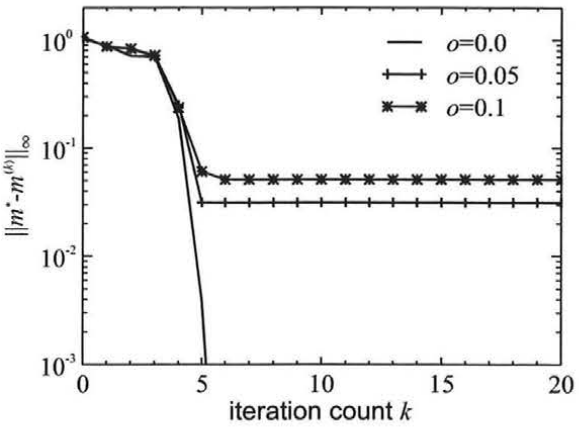

Figure 4.8: Convergence behavior using the reverse similarity and the GN method. The residuals are enlarged by replacing f by $f+o$.

example is given below. The same behavior, as for the forward modeling similarity measure, has to be expected for the reverse similarity (Fig. 4.8). Another effect however dominates the convergence properties. The reason that the algorithm fails to converge is the lack of differentiability of the reverse modeling similarity measure, which is caused by the piecewise linear interpolation of image f. The situation is depicted in Fig. 4.9. The algorithm essentially tries to find a position $m_i \in \mathcal{F}$ for every texture value g_i such that $g_i = f(m_i)$. The model of m will prevent that this can be reached exactly, but that does not affect the principle. Fig. 4.9 illustrates how the GN algorithm determines the next value $m_i^{(k+1)}$ of the iteration based on local linearization at state $m_i^{(k)}$. If f is linear, the GN method succeeds within one step, since in that case the problem is a linear least-squares problem. For the example depicted in Fig. 4.9, the values m_i^k oscillate between the two solutions guessed from the locally valid linear models. If g_i is small enough to be situated beyond the peak of f, the solution depends on the initial guess, but no convergence problems arise, as the support of the local model is not left. The same problems are observed also for the GNJ method. Clearly, choosing a different interpolation type for f, which ensures differentiability, will sidestep the problem. The computational effort is however increased. The forward similarity does not involve the linearization of f. The effect is thus not encountered. Possible difficulties may however arise for multi-image matching, if g is estimated from piecewise linear (bilinear) images f_i.

As mentioned, areas with large residuals can cause slow convergence of the matching algorithm. Fig. 4.10 shows an image with two different regions indicated. It is taken from a configuration that is further explained in Section 5.3.2. Both regions are used as reference domains to recover the respective part of the surface. In Table 4.1 and Figure 4.11 the convergence behavior of the GNJ method, embedded in a multiresolution setup working from coarse (level 4) to fine (level 0), is given. The iteration at each layer is terminated if $\|d^{(k+1)} - d^{(k)}\| < \frac{1}{16}$. d denotes the generalized disparity. For region 1 a rather good convergence behavior is achieved. Region 2 is larger and contains some problem areas. The small shadow area in the upper left corner has no corresponding part in the second image (Fig. 5.18). This is not only cause of a wrong matching result, but also of very slow convergence. This section concentrates on the performance of optimization methods. Clearly, it has to be expected that problem areas caused by untreated occlusions, uncompensated radiometric differences,

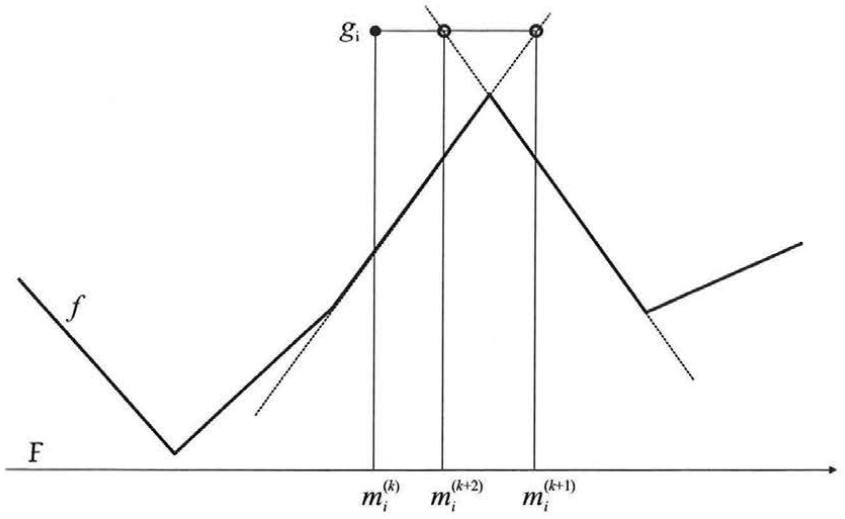

Figure 4.9: Illustration of the oscillation effect for the reverse similarity in connection with the GN method.

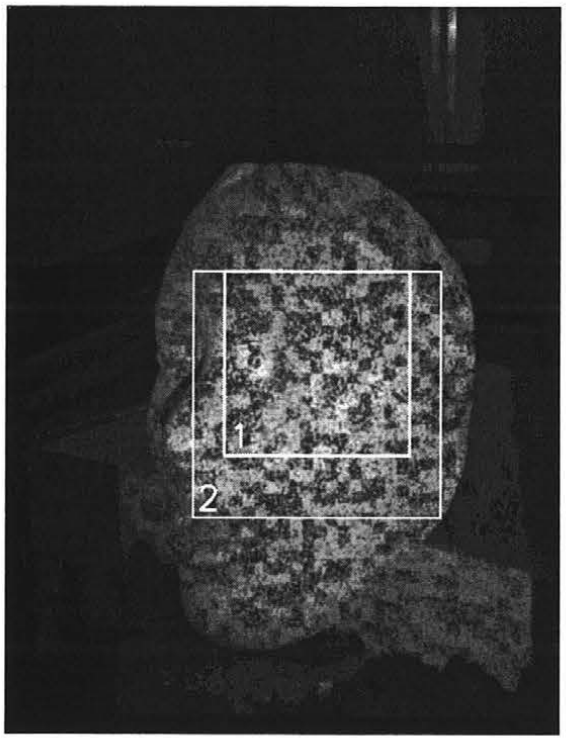

Figure 4.10:

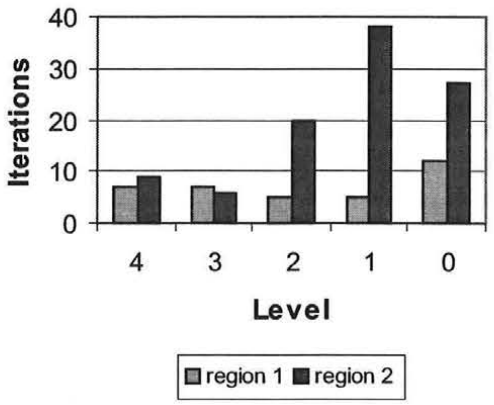

Figure 4.11: Number of iterations for region 1 and region 2 using the GNJ method.

Level	region 1	region 2	
	GNJ	GNJ	GNJ-PW
4	7	9	6 (10)
3	7	6	6 (6)
2	5	20	11(17)
1	5	38	11(20)
0	12	27	19(28)

Table 4.1: Number of iterations for region 1 and region 2 of Fig. 4.10. For the GNJ-PW method, the values in brackets include the intermediate steps for the realization of the stepsize rule.

and the like affect the matching result. An example is given in Section 5.3.1. The essential observation is however that these areas do affect the convergence behavior, too. This may be expected, though it is by no means self-evident. Detailed analysis reveals that the GNJ method performs too small steps for the problem areas. The current implementation of the matching algorithm achieves a slightly better convergence behavior by combining the GNJ method with the Powell-Wolfe stepsize rule (GNJ-PW). For region 2 the number of iterations, as required by the GNJ-PW method for each pyramid layer, are given in Table 4.1. Additionally, the number of iterations including also the intermediate steps for the realization of the Powell-Wolfe rule are given in brackets. The intermediate steps neutralize the benefit of the stepsize rule to some degree. It may be speculated that the problem can not only be addressed by more elaborate optimization algorithms, but also by a refined modeling, for example by incorporating techniques of robust estimation [36, 66] to reduce the influence of outliers.

Chapter 5

A Measurement System for Facial Surgery

In collaboration with the University Clinic of Oral and Maxillofacial Surgery at the General Hospital of Vienna a measurement system for the determination of the shape of human faces has been developed. The geometric information is needed by the surgeon for the planning phase and for the comparison of pre- and postoperative state. The aim is to provide an easy to handle measurement system. The images are acquired by four video cameras connected to a standard PC by a frame grabber card. The camera configuration is shown in Fig. 5.1. The cameras are arranged so that each side of the face is covered by two cameras with vertical baseline. The application of automatic matching algorithms for images of human faces is difficult, as the skin exhibits relatively less texture. No type of matching algorithm is able to perform satisfyingly under these circumstances. It is quite common to circumvent this by applying additional texture by means of a projected light pattern [11, 34, 38]. For close range applications, this is unproblematic. It requires no special technical effort as long as the knowledge about the pattern content is not used as additional input information for the matching algorithm. Two projectors are used, one for each side of the face, centered between the two respective cameras, Fig. 5.1. Fig. 5.2 depicts the frame that carries the cameras and the projectors mounted on a theodolite tripod. Fig. 5.3 portrays the measurement system in operation during the image acquisition phase.

5.1 Random Projection Pattern

The projection pattern is used to substitute missing matching information. To that end, it has to provide many edges. In terms of frequency content, it should cover a rather broad frequency range. The natural choice is to use a random black and white pattern. Using greyvalues other than black or white makes sense only if the knowledge about the content of the pattern is used for the matching algorithm. The proposed algorithm does not distinguish between natural and artificial texture. Accordingly, the pattern is designed to increase the variance of the texture as much as possible. Another question concerns the spectral distribution of the power density. The highest frequency components of the pattern should match the resolution of the imaging system. Signal power in a frequency range that is filtered by the cameras is wasted. A random pattern consists of pixels like any other image. The pattern pixels, once projected and imaged, should have at least the size of the pixels of the camera. With respect to the low frequency range, at least in principle, no enhancement is needed, because the power density

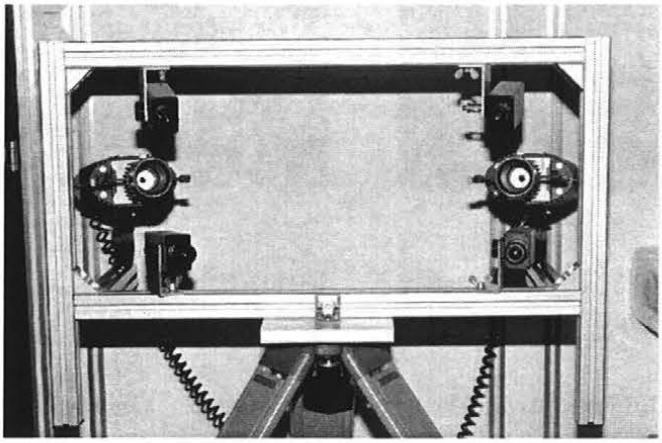

Figure 5.1:

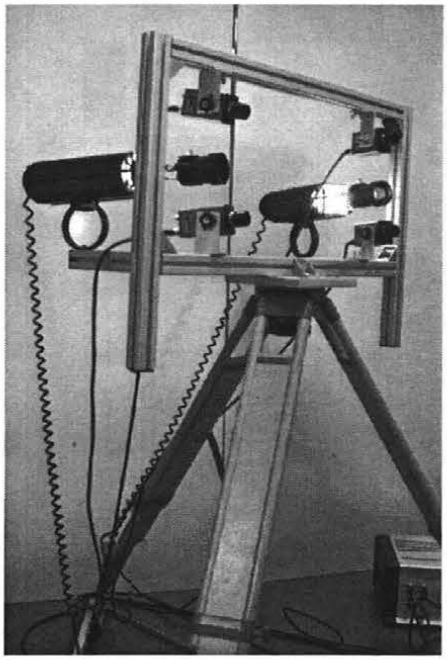

Figure 5.2:

Figure 5.3:

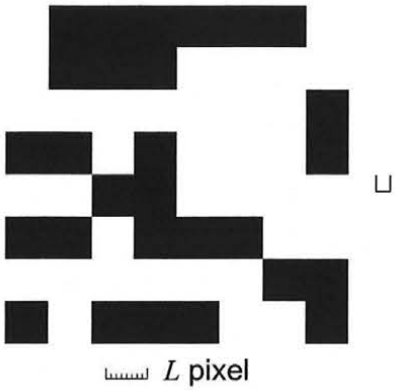

Figure 5.4: Low frequency pattern l.

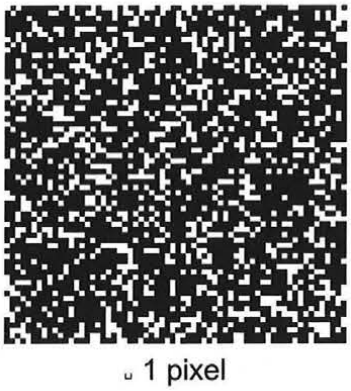

Figure 5.5: High frequency pattern h.

Figure 5.6: Composite pattern p.

spectrum of natural texture usually exhibits a strong lowpass characteristic. On the other hand, it might be useful to enhance the low frequency contributions to grant unproblematic initial convergence of the hierarchical matching algorithm. We propose a simple two level pattern. It enhances the texture in the high frequency range and provides the possibility of an additionally increase in the low frequency range. Bandwidth and relative weighting of high and low frequency components are steered by two parameters. The shaping of spectra of random processes is rather trivial if the range of admitted pattern values is $\mathbb{R}$. If the patter is required to be binary, the matter is more complicate. A general theory seems not worthwhile for the purpose. The problem is considered in more detail in [15] for the design of 1D sequences as needed for communication applications. The proposed pattern is composed of a low frequency pattern l and a high frequency pattern h. We treat them on the discrete domain $\mathbb{Z}^2$. The translation to a stochastic process defined on $\mathbb{R}^2$ is straightforward. Both patterns are binary black and white patterns. For the calculation of autocorrelation functions and power density spectra we assign black the value 0 and white the value 1. The low frequency pattern consists of $L \times L$ square areas with homogenous color (Fig. 5.4). Black and white are distributed equally and are chosen independently for each square area. The values $h_{\mathbf{k}}$ of the high frequency pattern are assumed statistically independent, but not equally distributed, $\mathrm{P}(h_{\mathbf{k}} = 0) = \frac{1-\mu}{2}$ and $\mathrm{P}(h_{\mathbf{k}} = 1) = \frac{1+\mu}{2}$ (Fig. 5.5). Note that the term high frequency pattern does not mean that it has a highpass characteristic, but that it extends to "high" frequencies. Apart of the mean, it is a spectral white pattern with homogeneously distributed power density. The process h is used to introduce additional matching information into the lowpass pattern l. This is done by flipping those pixels of l where h is 1 or white, which is equivalent to a XOR($\sqcup$) operation. The new pattern p thus is given by

$$p_{\mathbf{k}} = l_{\mathbf{k}} \sqcup h_{\mathbf{k}} \; . \tag{5.1.1}$$

The effect of this operation is easily analyzed. If black is mapped to -1 rather than to 0, which is only a linear transformation, the new pattern is simply the product $-hl$. The spectrum is the convolution product of the spectra of h and l. From (5.1.1), it is clear why the pixels of the high frequency pattern are generally not equally distributed. If $\mu = 0$, each pixel of l is flipped independently and the structure of l is lost, resulting in a spectral white pattern p. Fig. 5.7 depicts the autocorrelation function of the composite pattern p

$$R_p[\mathbf{k}] = \mathrm{E}\{p_{\mathbf{i}} p_{\mathbf{i}+\mathbf{k}}\} \; . \tag{5.1.2}$$

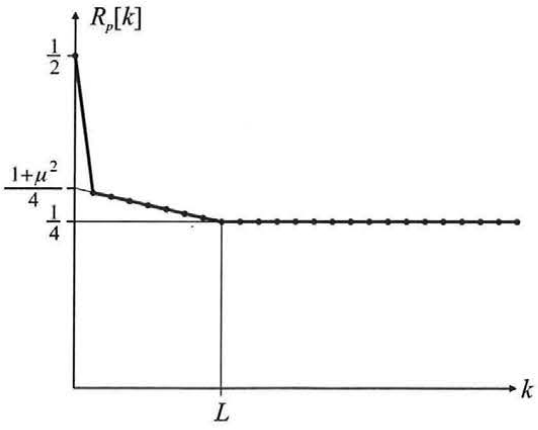

Figure 5.7: Autocorrelation function of the composite pattern, $L = 8$, $\mu^2 = 0.2$.

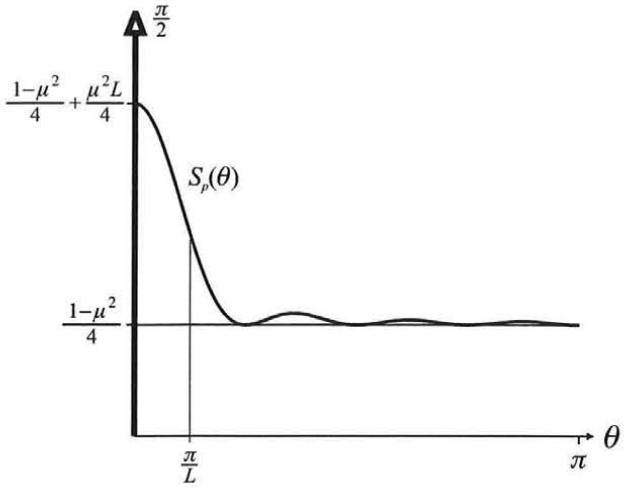

Figure 5.8: Power density spectrum of the composite pattern, $L = 8$, $\mu^2 = 0.2$. A Dirac impulse of weight $\frac{\pi}{2}$ is located at $\theta = 0$.

The pattern is not stationary because of the square patches of l. The autocorrelation function is calculated as usual for cyclostationary processes [35, pp. 66]. The power density spectrum of p is the Fourier transform of R_p

$$R_p \circ\!\!-\!\!\bullet\, S_p \,. \tag{5.1.3}$$

It is depicted in Fig. 5.8. The Dirac impulse at $\theta = 0$ is a consequence of the mean of p. The two parameters μ^2 and L determine the shape of the spectrum. $\mu^2 \in [0, 1]$ controls the relative weight of the lowpass contributions. For $\mu^2 = 0$ the spectrum is white, for $\mu^2 = 1$ it is completely low frequency dominated. The length of side L of the square patches determines the bandwidth of the lowpass pattern l. Projected on a surface, the resulting texture is again a product-like combination of the natural texture and the light pattern.

More elaborate schemes of creating patterns may be found. For example the low frequency pattern can be replaced by realizations of Pickard random fields [46, 47], which are stationary and have a geometric correlation structure. The XOR combination of patterns is also a flexible tool for designing special purpose patterns. The proposed two parameter family of patterns by far covers what is needed for the presented application.

5.2 Implementation of the Algorithm

The matching algorithm presented in this work is an integral part of the measurement system for facial surgery. The algorithm is currently implemented in IDL (Interactive Data Language by Research Systems Inc.). The interactive environment proved to be useful for testing purposes and for the analysis of the algorithm performance. It however puts some limitations on the implementation, because operations are performed efficiently only if specified for whole arrays of data. Using loops like in compiled languages (Pascal, C etc.) diminishes the performance significantly. Some routines have been outsourced to C++. Nevertheless, the style of implementation is similar to parallel processing. As the system is currently in prototype state, modularity and ease of testing have been favored over efficiency. There should be enough space for improvements in speed. The current implementation of the matching

algorithm does not use the full strength of the concept. Only image-to-image matching is performed, using the generalized disparity d (cf. Section 3.3.2) as matching parameter. For the basic constellation one image serves as template image g and the other as data image f. The symmetric variant of image-to-image matching as given by (3.1.18)-(3.1.20) is implemented, too. It is theoretical more founded. However, more experience is available for the basic constellation. The matching is performed for each side of the face independently. The fusion of the two data sets is not done within this work. From the preceding sections, it is clear that configurations that are more elaborate could be used. In particular processing multiple images simultaneously is frequently reported to enhance the robustness of the algorithm [10, 22, 23, 37, 49]. The current implementation nevertheless suffices to perform first test runs.

The following parameters are required as input to the matching algorithm:

- the images and the regions that are to processed,
- the orientation parameters,
- the pyramid levels that are to be used,
- the initial guess of the solution for the lowest pyramid layer,
- the node spacing for the bilinear facets, and
- the amount of regularization.

Further decisions concern

- the type of discretization,
- the optimization method, and
- whether basic or symmetric image-to-image matching is used.

The parameters of the orientation are determined in advance using the hybrid photogrammetric adjustment program ORIENT [28, 32]. In a preprocessing step the nonlinear geometric distortions are removed from the images. This necessitates resampling of the images, which is commonly considered undesirable. The effort of taking the nonlinear distortions into account throughout the whole iterative process is however too high. After the rectification, the image pyramids are generated. If the double grid discretization is used, the image pyramid should be of even type. For the single grid discretization it should be of odd type. This way a correct alignment of the samples $d_{\mathbf{k}}$ of the matching function is granted across pyramid layers. For both pyramids a filter kernel composed in a tensor manner from binomial distributions is used. In detail, the kernel $\frac{1}{8}(1,3,3,1)$ is used for the even pyramid and $\frac{1}{16}(1,4,6,4,1)$ (cf. [22]) for the odd pyramid. They come close to Burt's most Gaussian like filters [8]. Single and double grid discretization are both implemented. Unlike the experimental setup for the 1D case, the particular application did not profit from the double grid discretization, thus the single grid discretization is commonly used. For the initial guess of the matching function, specifying a plane in object space has been found sufficient. The approximate distance between the face and the camera yields enough information to start successful matching. This depends to some extent on the patch size, as it will enforce a limit on the coarsest layer of the pyramid that can be used. The properties of the surface are determined by the node spacing and the regularization parameter α. A node spacing of 8 has been found appropriate

for most applications. The regularization is based on the membrane functional (4.1.11). A typical value for α is $0.1\sigma_f^2$. Though some authors claim that the bias of the membrane functional towards $d \equiv \text{const}$ is not adequate for image matching [18, 26], we have encountered no problems so far. Refinement of this aspect is possible though.

Three different optimization algorithms are available. A steepest descent algorithm using the Powell-Wolfe stepsize rule, the GNJ optimizer with fixed stepsize, and the GNJ optimizer with Powell-Wolfe stepsize control. The iteration is terminated if $||d^{(k+1)} - d^{(k)}||_\infty < \frac{1}{16}$. The convergence of neither method is entirely satisfying. There are always some problem areas in an image that cause a high number of iterations. An example of the influence of problem areas has already been given in Section 4.2.4. For the following examples, the GNJ-PW method is used, since it has been found to terminate after an acceptable number of iterations. The complexity of a single step of the GNJ-PW optimizer is however high. In some cases the algorithm terminates because the stepsize strategy fails to find a suitable stepsize. This can be accepted for two reasons: Firstly, the behavior is encountered only near the solution. Secondly, as the convergence is dominated by problem areas, the termination happens when the majority of values $d_\mathbf{i}$ has settled with high precision. For the problem areas, the quality of the result is doubtable anyway.

At the current state, we can not report facts about the final accuracy of the matching result. For none of the examples have been appropriate reference data available. The proposed method falls in the class of least-squares matching methods. The accuracy potential of these methods has been well investigated by researchers of the photogrammetric field. There is no reason why the herein-proposed variant of global least-squares matching should perform significantly better or worse than related approaches. The achievable accuracy is influenced neither by the usage of the forward similarity instead of the reverse similarity nor by the optimization method. Weisensee [58] claims to reach a standard deviation of 0.2 pixel image space equivalent. Heipke [22, 23] obtains a maximum deviation of 0.1 pixel for an artificially generated example. The example in Section 5.3.1 demonstrates that the radiometric effects can cause deviations up to 1 pixel. This is an extreme case however. It is caused by shortcomings of the radiometric model. Only mean and standard deviation of the two images are mutually adjusted by a simple global linear transformation of intensity values. The majority of methods uses similar simple strategies. Consequently, before trying to achieve an accuracy in the range of 0.1 pixel, the radiometric modeling has to be refined. We are content to reach accuracy comparable to other least-squares methods for regions that are unproblematic with respect to radiometry. Currently no effort is made to deduce the accuracy from the matching result. The method that is used in adjustment calculus to calculate theoretical standard deviations of the unknown parameters appears to be computationally very demanding if large patches are considered. It should however be possible to judge the obtainable accuracy of the matching result, at least qualitatively, from the diagonal terms of the Gauss-Newton term and the residuals of the image intensities.

5.3 Examples

5.3.1 The Author's Face

Fig. 5.9 and 5.10 show two images taken with the medical measurement system. They portray the author's face with a random pattern texture projected onto the skin. For the determination of the surface a patch of 513×513 pixels was processed simultaneously. The indicated region of Fig. 5.9 served as template image. The matching function was parameterized by

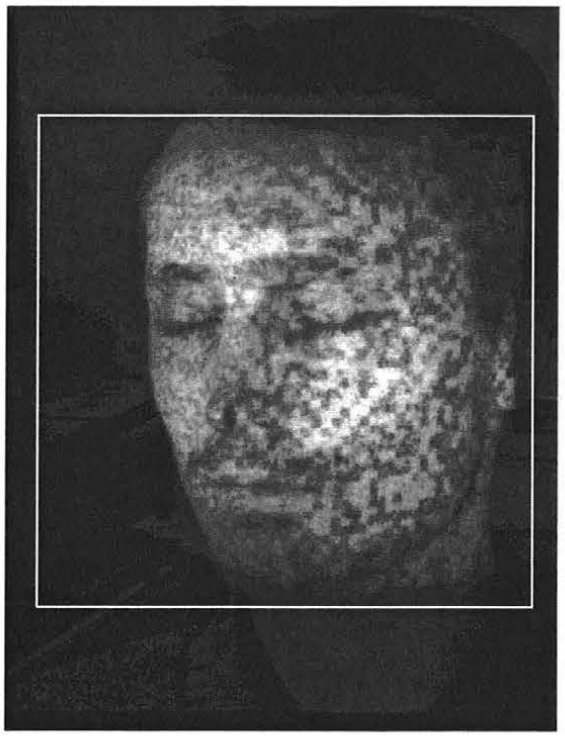

Figure 5.9: Image of the author's face with the reference region indicated.

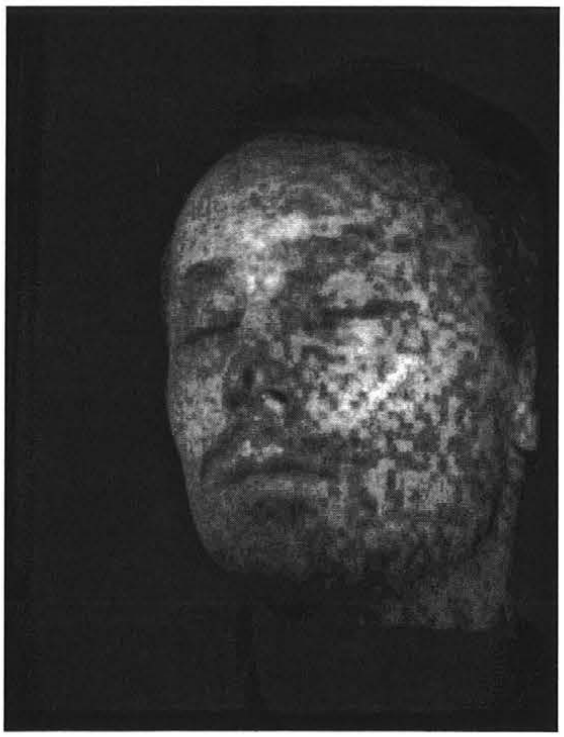

Figure 5.10: Image of the author's face, second view.

bilinear facets with node spacing 8. Eventually 65×65 nodes were determined. Additionally a weak regularization with $\alpha = 0.1\sigma_f^2$ was applied. The matching was performed using the GNJ method together with the Powell-Wolfe stepsize strategy. The computation took about 3 min. on a AMD-K6 PC with 200MHz and 32Mb memory. The required number of iterations for each level is given in Table 5.1. Additionally the number of iterations, including the necessary steps for the realization of the Powell-Wolfe rule, is given in brackets. The performance is adversely affected by areas with large residuals. Especially in the occluded area to the right of the nose[1], the iteration does settle very slowly. The reconstructed surface is depicted in Fig. 5.11. For the regions outside the face no meaningful matching is

[1]The terms left and right are used with respect to the coordinate system of the head. The right side of the face is thus portrayed by the left part of the image.

Level	Iterations
5	7 (7)
4	8 (8)
3	8 (8)
2	16(20)
1	16(24)
0	14(22)

Table 5.1: Number of iterations for the matching of Fig. 5.9 and 5.10 using the GNJ-PW method. The values in brackets include the intermediate steps for the realization of the Powell-Wolfe stepsize rule.

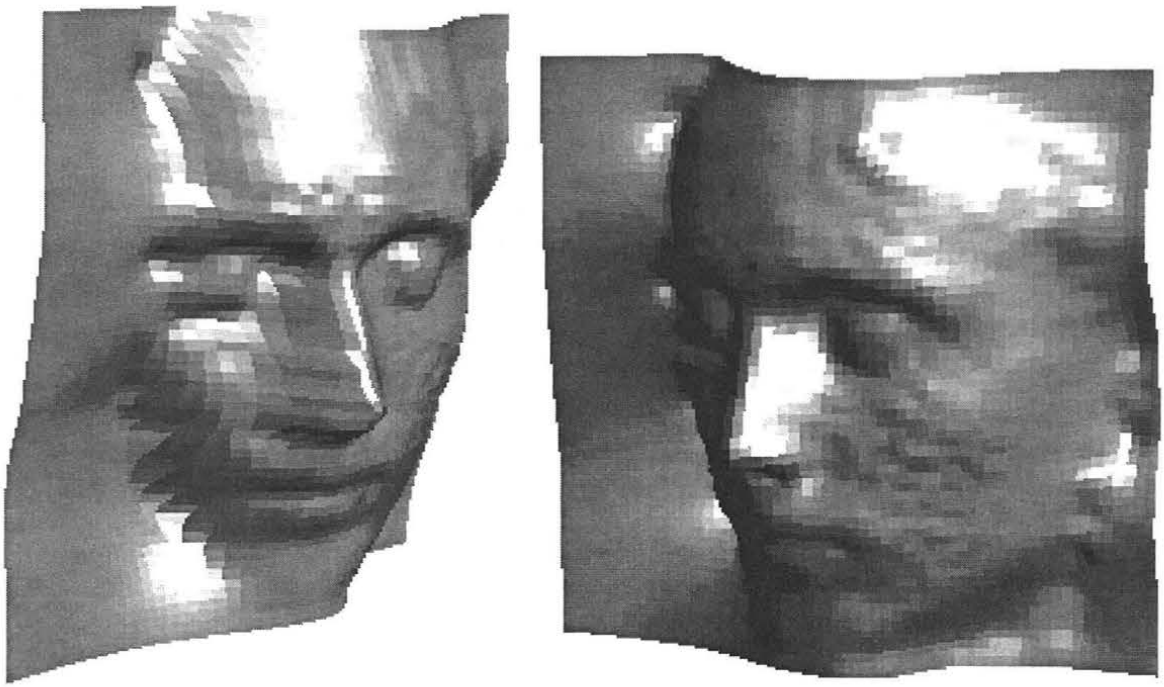

Figure 5.11: Shaded reconstructed surface. View of left and right side.

achieved. The dark background regions do however not disturb the algorithm. According to the input images, the right side of the face is reconstructed only roughly. At the right side of the nose the interpolation enforced by the regularization term is clearly visible. The surface exhibits some distortions at the left cheek. As visible from the images, Fig. 5.9 and 5.10, these correspond to very bright, over-exposed areas. The radiometric model falls short of compensating these effects. The consequence are undulations of the reconstructed surface. In front of the forehead there are some fringes of hair. They are also visible as distortions in the reconstructed surface.

The effect of uncompensated radiometric differences is illustrated in more detail for a 128×128 region located at the left cheek. The transformed image $f \circ \mathbf{m}^*$, the template g, and the absolute value of the difference $|f \circ \mathbf{m}^* - g|$ are depicted in Fig. 5.12-5.14. From the difference image it can be seen that the residuals are not only caused by noise. Some bright spots appear because of radiometric effects. Fig. 5.15 shows intersections of the images and the disparity function d^* along the lines indicated in Fig. 5.12-5.14. Clearly $f \circ \mathbf{m}^*$ and g are aligned geometrically correctly. There are however areas where the intensity differs systematically. These correspond to unnatural deformations of the disparity function d^*. One unit of disparity corresponds to approximately 1 mm in object space and 1 pixel in the domain $\mathcal{F}$. The epipolar direction is approximately parallel to the x_2-direction (vertical direction in Fig. 5.12-5.14). With respect to the reference domain $\mathcal{G}$, increasing the disparity causes a leftwards shift of $f \circ \mathbf{m}^*$. The figure additionally shows a smoothed version d^s of the disparity and the corresponding $f \circ \mathbf{m}^s$. Comparing $f \circ \mathbf{m}^s$ and $f \circ \mathbf{m}^*$ reveals how the radiometric differences cause the distortions of d^*. As the algorithm does not take the source of the residuals into account, it tries to compensate the residuals, caused by the oversimplified radiometric model, by a geometric correction. The driving force is the negative gradient of the image energy term $-\nabla E_{\text{img}}$, which works dominantly at the steep slopes of f. The direction of the force imposed by $-\nabla E_{\text{img}}$ is indicated by arrows. The resulting deformation of d^* is

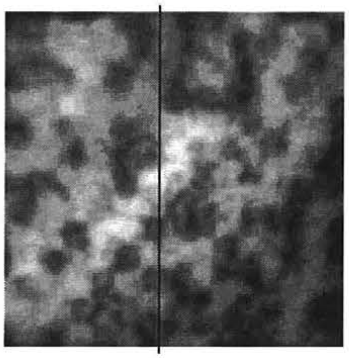

Figure 5.12: Region of the transformed image $f \circ \mathbf{m}^*$.

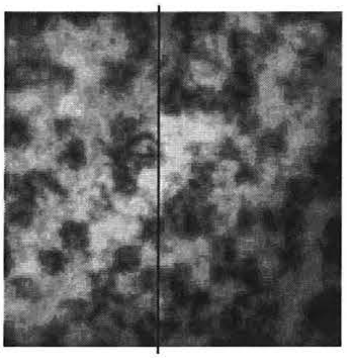

Figure 5.13: Region of the template image g.

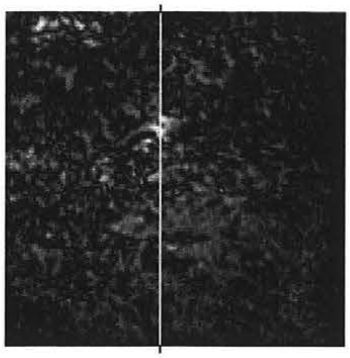

Figure 5.14: Difference image $|f \circ \mathbf{m}^* - g|$. Brightness scale differs from Fig. 5.12 and 5.13.

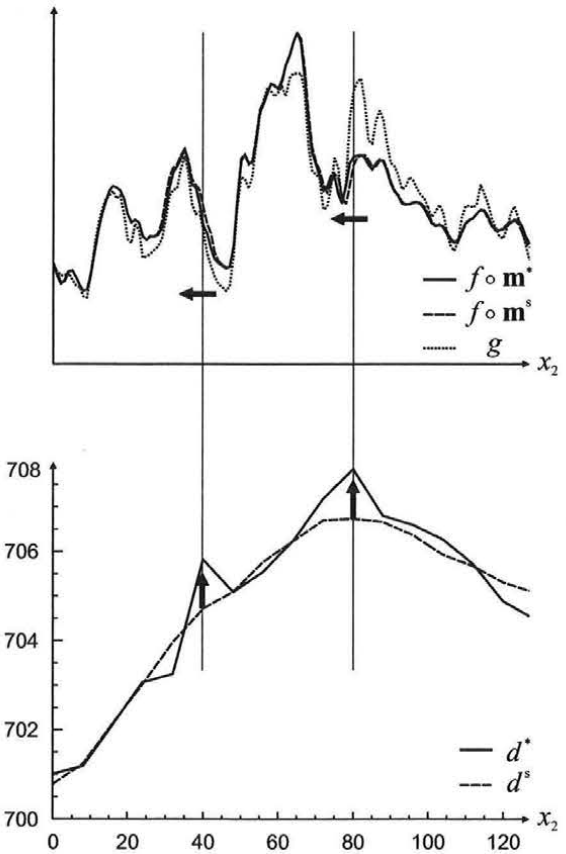

Figure 5.15: Intersections of $f \circ \mathbf{m}^*$, g, and d^* along a line parallel to the x_2-direction. d^s is a smoothed version of the disparity d^*. It leads to the transformed image $f \circ \mathbf{m}^s$.

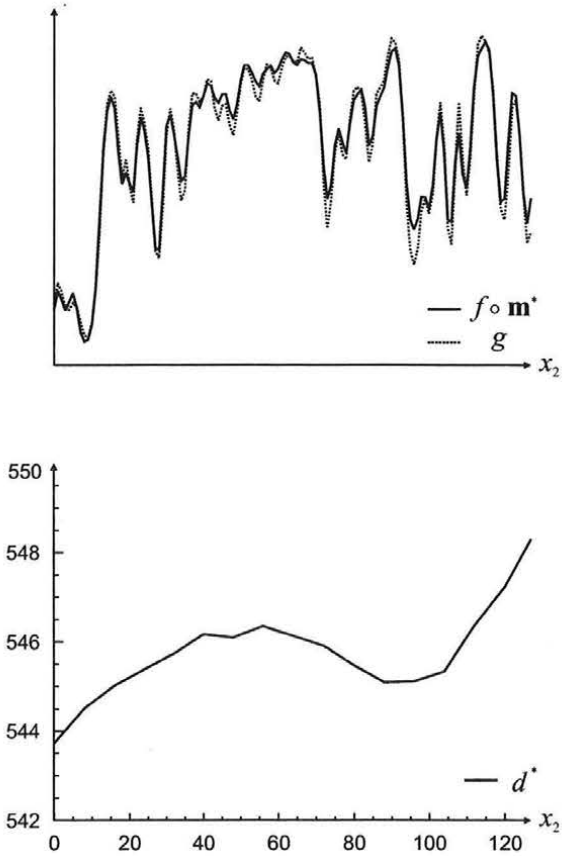

Figure 5.16: Intersections of $f \circ \mathbf{m}^*$, g, and d^* along a line parallel to the x_2-direction for the gypsum face example (Section 5.3.2).

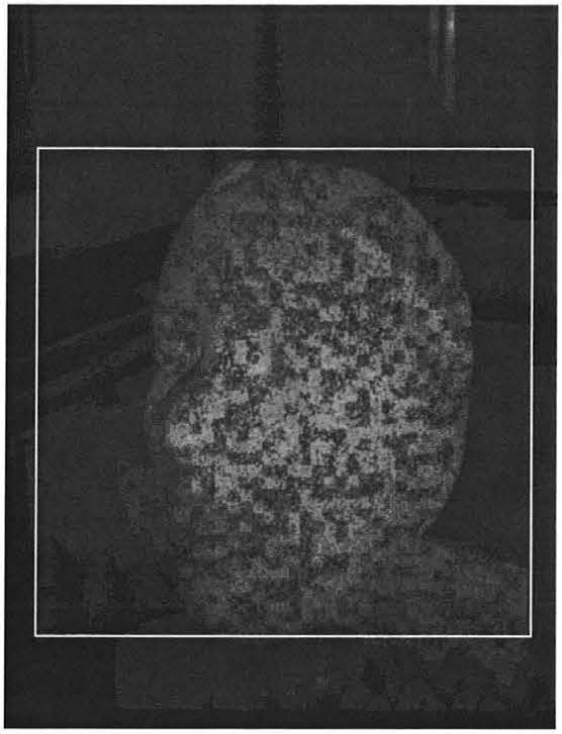

Figure 5.17: Image of a gypsum face with the reference region indicated.

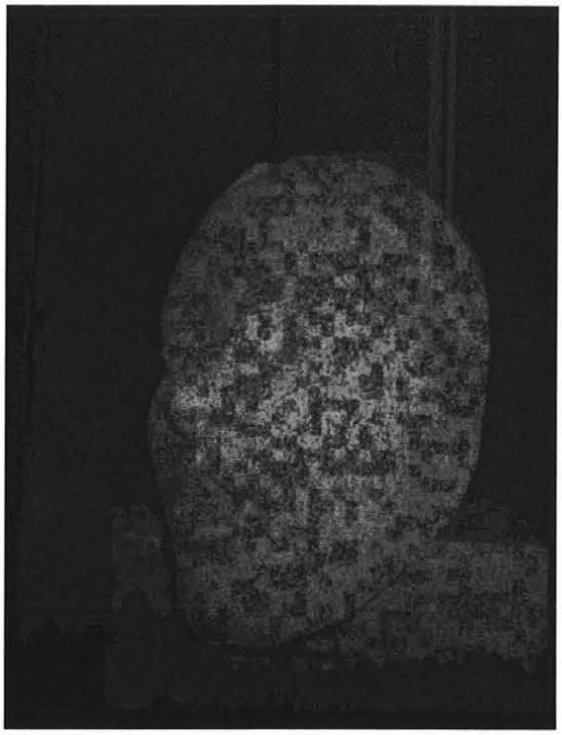

Figure 5.18: Image of a gypsum face, second view.

clearly visibly and is indicated by arrows, too. Fig. 5.16 serves for comparison. It portrays the same quantities for an example that uses images of a gypsum face (Fig. 5.17, 5.18). The configuration is considered in the next section. A similarly located region of the left cheek is used. The radiometric differences are harmless in that case, causing almost no distortions.

So far no refined radiometric model is implemented. The results show however that an improved treatment of the radiometry is necessary. The same relevancy has to be assigned to the robustness with respect to outliers. They have a similar influence on the matching result like unmodeled radiometric differences.

5.3.2 Gypsum Face

Fig. 5.17 and 5.18 portray images of a gypsum model of a face. The model without texture is depicted in Fig. 5.19. The parameters that are used for the reconstruction are the same as in the previous section. For this example the basic similarity measure and the symmetric variant (3.1.18)-(3.1.20) was tested. The basic image-to-image matching used the region indicated in Fig. 5.17 as template image g. The symmetric variant used the same reference domain. However, in this case the region served as image f_0 (cf. Section 3.1.3) and not as template image g. The reconstructed surface as obtained by the symmetric variant is depicted in Fig. 5.20. The results are almost the same for basic and symmetric constellation. For the basic variant, the computation took approximately 7 min. on the same computer environment as above. For the symmetric variant, only 4 min. 30 sec. were needed. The number of iterations for each pyramid level is given in Table 5.2. The convergence is in neither case satisfying. For the basic constellation, about 80% of the time are spent at level 0 of the hierarchical scheme. For the symmetric constellation, a high number of iterations

Figure 5.19: Gypsum face without artificial texture.

Figure 5.20: Reconstructed surface.

Level	Iterations	
	basic	symmetric
5	8 (8)	8 (8)
4	6 (6)	5 (5)
3	12(24)	11(25)
2	15(19)	30(44)
1	18(31)	12(19)
0	31(56)	15(20)

Table 5.2: Number of iterations for the matching of Fig. 5.17 and 5.18 for basic and symmetric constellation using the GNJ-PW method. The values in brackets include the intermediate steps for the realization of the Powell-Wolfe stepsize rule.

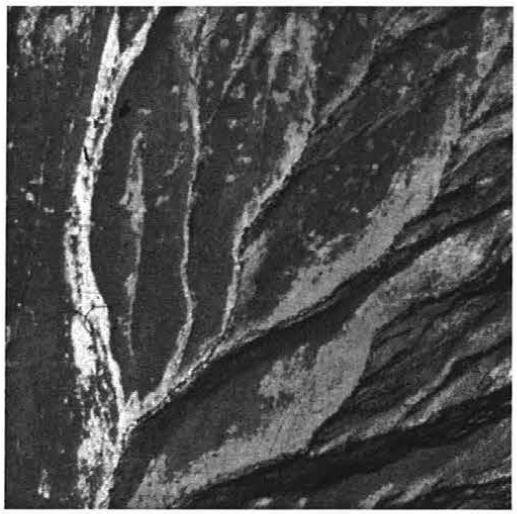

Figure 5.21: Left image of mountainous terrain.

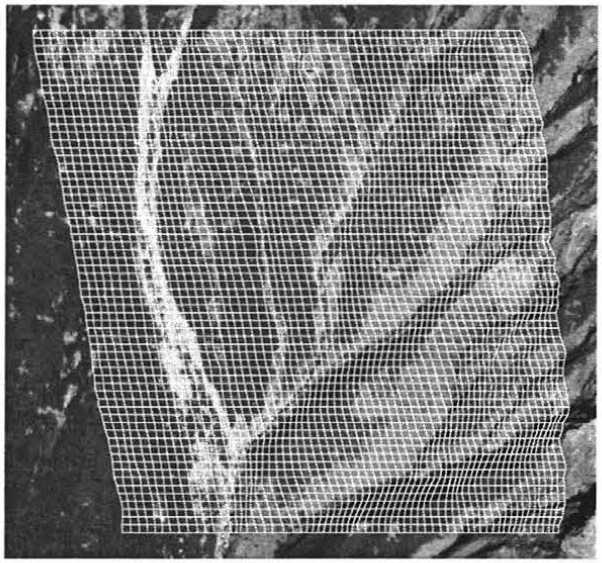

Figure 5.22: Right image of mountainous terrain. Computed map $\mathbf{m}^*$ plotted as an irregular grid.

is necessary at level 2. On lower levels data of smaller size have to be manipulated thus the bad convergence at level 2 has less impact on the total computation time. Again the area to the right of the nose is problematic and causes rather slow convergence. The influence of the shadow areas on the convergence has already been discussed in Section 4.2.4. The current experience is insufficient to conclude whether the symmetric constellation involves practical benefits, though it is obviously theoretically sounder. In contrast to the example of the previous section, no artifacts due to radiometric effects are encountered. Most likely the reason is that the reflectance properties of gypsum obey the Lambertian law better than the human skin.

5.3.3 Matching of Aerial Imagery

To demonstrate the methods capability under varying conditions, the matching of two aerial images of mountainous terrain, Fig. 5.21 and 5.22, is presented. The second image, Fig. 5.22, is overlaid with the computed map $\mathbf{m}^*$ plotted as an irregular grid. $\mathbf{m}^*$ was obtained using a very similar setup as in the previous examples, employing the symmetric variant of image-to-image matching, (3.1.18)-(3.1.20). The image portrayed in Fig. 5.21 is actually a subset of a much larger image. It served as image f_0. The reference domain coincides with the domain of f_0, which has a size of 513×513 pixels. The image portrayed in Fig. 5.22 was used as image f_1. Again the node spacing of the bilinear basis functions was $L = 8$ and the regularization parameter was set to $\alpha = 0.1\sigma_f^2$. The matching was performed using the GNJ method together with the Powell-Wolfe stepsize strategy. The computation took approximately 6 min. The number of iterations is given in Table 5.3. At level 0 of the pyramidal scheme a high number of iterations was necessary. However, at this level after 16(21) steps the termination criterion $|d_{\mathbf{i}}^{(k+1)} - d_{\mathbf{i}}^{(k)}|_\infty < \frac{1}{16}$ is fulfilled for 99% of the nodes. Again a rather small number of points has a strong impact on the convergence behavior. As in the previous examples, the generalized disparity was taken as matching parameter. It is

Level	Iterations
5	7 (7)
4	5 (5)
3	10(11)
2	9 (9)
1	14(15)
0	26(33)

Table 5.3: Number of iterations for the matching of normal case aerial images of mountainous terrain using the GNJ-PW method. The values in brackets include the intermediate steps for the realization of the Powell-Wolfe stepsize rule.

identical to the ordinary disparity for the normal case. It is however possible to reduce the effort for the computation by taking advantage of the specific characteristics of the normal case arrangement. Orientation parameters (projection center, cardinal point, and camera constant) were not available. Fig. 5.23 depicts the surface as reconstructed from the images. Because of the lack of orientation parameters, the drawing is not scaled. The surface is plotted as graph $Z(X, Y)$. The figure must however not be misunderstood in that the sample values in the orthophoto domain (X, Y) do not form a regular grid.

Reference data were not available. A means to detect errors is the absolute difference $|f_1 \circ \mathbf{m} - f_0|$,· which is depicted in Fig. 5.24. Some bright spots are apparent. These are outliers caused by pollution or other imperfections of either of the photographs. For example, a black spot is visible in the upper left part of Fig. 5.21 to the right of the broad vertical white stripe. It is also seen in Fig. 5.24. The structure of the original images Fig. 5.21 and 5.22 is weakly visible in the difference image, because the radiometric correction is not able to remove all radiometric differences between the images. Major problem areas or obviously false matched points were not found.

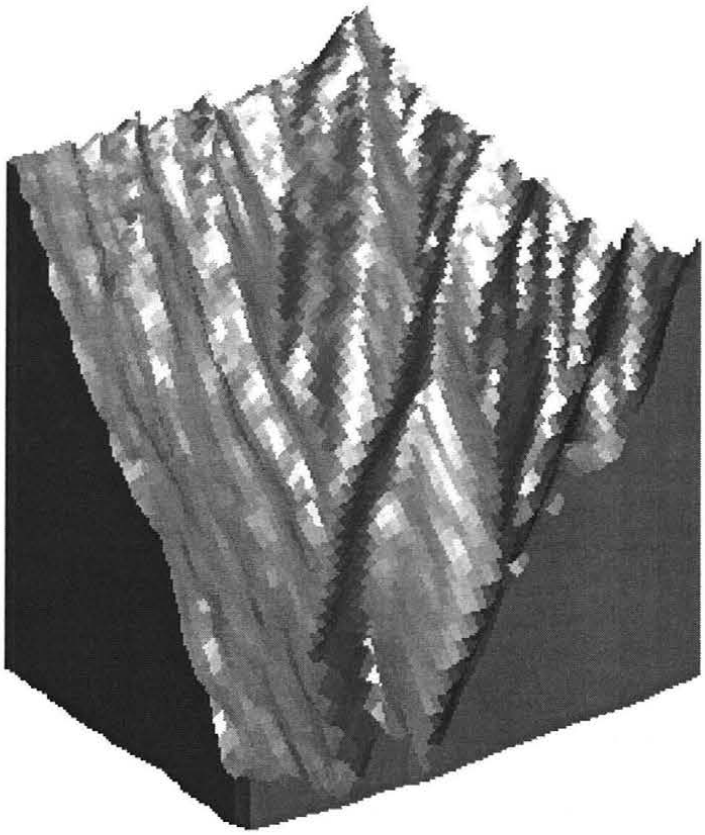

Figure 5.23: Reconstructed surface.

Figure 5.24: Absolute difference $|f_1 \circ \mathbf{m} - f_0|$.

Chapter 6

Summary and Outlook

This work covers a broad range of topics. Starting from the notion of the geometric connection and other foundations, we have discussed the design of a new type of similarity measure. Further, the more practical aspects, such as the modeling of the matching function and the optimization, have been addressed, and lastly a system for measuring the geometric shape of human faces has been presented. Not all aspects have received the same attention. In the following a short summary of the achievements is given. A separate section lists aspects that could be improved in the further development of the image matching algorithm.

6.1 Achievements

Among the foundations presented in Chapter 2 there are some assumptions and modeling choices that follow the mainstream of image matching algorithms. Outstanding is the geometric concept based on the definition of the correspondence relationship, as it extends the commonly used set of notions. It has been found convenient for treating a whole class of problems in a common framework. It clarifies the role of the reference domain and thereby supports the concept of forward modeling. The reference domain, though crucially important for the computations, is recognized to only serve for the parameterization of the geometric connection. This eliminates some aspects of the discussion whether object space based matching algorithms do have a theoretical primacy with respect to image-to-image matching. Further, the concept leads to the requirement that the similarity measure be invariant with respect to the particular choice of reference domain. The forward modeling similarity measure indeed has this property.

In this work, the main emphasis has been laid on derivation and analysis of the forward modeling similarity measure. It is a theoretical concept whose translation into practical benefits is not completely done so far. The conceptual advantages of the forward similarity are however striking. Its properties are considered in contrast to the reverse similarity, as the latter is the natural pendant. The forward modeling similarity measure provides the following advantages:

- regions of an image contribute to the similarity measure according to their size in image space and not according to their extent in the model,
- for multi-image matching applications, the images that portray a particular surface area best, that is with best spatial resolution, dominate the reconstruction process,
- it is invariant with respect to the choice of model domain, and

- it leads to the modeling of occlusions in a natural way.

These properties are not independent. They are all caused by weighting the local contributions by the Jacobian $\det \mathbf{m}'$. The whole concept can indeed be reduced to this rather unspectacular point. Its impact is however significant. Further, it necessitates a much more elaborate analysis than the reverse similarity. Without employing continuous functions to represent the images, the whole concept could not have been obtained. It would be too difficult to see the structure behind the problem. Just consider the derivation of (3.4.22) from (3.4.16). Without knowledge about the eventual structure of the formula, it is hardly found.

For the optimization, a method that is referred to as Gauss-Newton-Jacobi method is proposed. The recurrence formula resembles a popular relaxation scheme for image matching. However, its theoretical and practical properties have been found much better. The convergence properties are convincing for most parts of an image, though problem areas, especially the currently untreated occlusions, affect the performance adversely. The efficient optimization is still a topic with open questions.

Lastly, a concrete example, the measurement system for facial surgery, demonstrates that the matching algorithm is practically applicable. Experience shows that the system works quite stable. Problem areas occasionally cause an increased computation time. The reconstruction of the complete face from a single stereo pair takes currently 3 to 7 min. Uncompensated radiometric differences have been found to cause distortions of the reconstructed surface. The example of matching two aerial images of mountainous terrain demonstrates the algorithms ability to perform under various conditions.

6.2 Future Developments

A couple of assumptions and simplification have been used to facilitate the theoretical and practical treatment of the matching problem. They can be divided into two categories. First, those are listed for which improvements and solutions are available and which have been omitted primarily for practical reasons. The radiometry is one such aspects, which could and, concluding from the practical results, should receive a more refined treatment. Ideas on this topic can be found in [18, 58]. Related is the robustness with respect to outliers, as outliers result in distortions of the matching function similar to those caused by uncompensated radiometric differences. Standard robust estimation techniques may be employed to enhance the performance of the matching algorithm (e.g. [36, 65]). Another point is the modeling of the matching function, which does not take discontinuities into account so far. The problem is encountered in other Computer Vision applications, such as the reconstruction of the visible surface from discrete measurements, too. Approaches for the treatment of discontinuities can be found in [36, 49, 53, 55]. Further, only very simplified geometric constellations have been considered. Particularly for the measurement system for facial surgery, a better solution would use the images from all four cameras simultaneously. The orthophoto domain could serve as a more flexible reference domain. Experience with respect to the usage of multiple images can be found in [22, 23, 37, 49, 58]. Using the orthophoto domain as reference domain is done in [22, 23, 58]. The most rigoros approach, employing a full 3D model, provides also no fundamental, but obviously a practical problem.

In the following two problems are mentioned that may be considered unsolved. The first concerns the treatment of occlusions. Occlusions are treated by the forward similarity measure in a natural way. The practical aspect, namely how to perform the optimization in the presence of occlusions, is however unsolved. Somewhat this problem is linked with

the question, what type of optimization algorithm to use for image matching. Using the forward similarity in the presence of occlusions together with global optimizers like dynamic programming or simulated annealing appears to be unproblematic. Occlusions are treated by the approaches in [5] using dynamic programming and in [51] using mean field annealing. Both applications handle only normal case stereo images and it is not entirely clear whether these methods perform satisfyingly for multi-image constellations with arbitrary mutual orientation. In [37] occlusions are taken into account and a local optimizations algorithm, a relaxation scheme, is used. Considering that occlusions cause discontinuities of the reverse similarity and discontinuities of the first derivative of the forward similarity, it has to be doubted that this strategy is unproblematic. In the course of this work, occlusions have not been taken into account by practical experiments. Contrary to the above listed problems, it is expected that in this respect results of other authors are not easily integrated with the proposed method. The relevancy of this aspect is linked to the question, whether future developments use local optimizers or whether they employ other alternatives.

A definitely unsolved problem is the integration of the blurring operation in the model. A motivation is given why this could be important at all. Recall the performance of the forward similarity for multi-image matching. According to the results of Section 3.1.3, an image contributes to the reconstruction of the correspondence only if the respective part of the model is visible in the image at all. The forward similarity automatically weights the parts of an image, the spatial components, correctly. With respect to frequency components, there is no mechanism to grant such a behavior. Consider two images of an object with different spatial resolution, one with fine resolution and a second, which portrays less details of the object. There is no reason why these two images should not be used together for surface reconstruction, though of course the achievable accuracy is limited by the coarse resolution. If the matching is done with any of both similarity measures and any reasonable local similarity measure, the algorithm will try to match fine scale details of one image with something that is not present in the other image. Like uncompensated radiometric differences, this causes distortions of the matching result. The behavior is not a consequence of optimization. It will result for arbitrarily good initial guesses of the solution. There is further no theoretical justification to increase the requirements on the smoothness of the surface, because the surface properties are independent of the images. Obviously, it is possible to smooth the image that has better resolution. However, why should a problem be solved better if a part of the available information is cut away?

The solution of the problem is to employ strict forward modeling. Just like the forward similarity introduces plausible weighting of the spatial components of an image, including the blurring operation into the model leads to reasonable weighting of the frequency components. Theoretically, this is no problem. Practically, an efficient structure has to be found. Like for the forward similarity, it is desirable to have all computations done in the reference domain. This necessitates the translation of the smoothing operator from the image domain to the reference domain, which bears considerable complexity.

6.3 Concluding Remarks

Image matching is a heavily investigated problem and there is still a tension between theoretical concepts and practical requirements. The problem is harder than it appears at first glance. It necessitates sound theoretical approaches that keep practical aspects in mind. Many ideas have been contributed and have lead to a wide diversity of approaches. Abstraction has to be employed to factor out common aspects of existing concepts. On the other

hand some aspects have not received enough attention. Those require new concepts. From the results of this work, we conclude that there is much benefit to be gained from sound modeling. Simplifying assumptions are part of virtually any model. Particularly for Image Processing and Computer Vision the huge amount of data that have to be dealt with enforces a judicious design of the model. It is however utterly important to be aware of the underlying assumptions. This might appear self evident for scientific work anyway, but practically the impact of some modeling choices is overlooked easily. An example is the reverse modeling similarity measure, which is clearly based on a simplification, yet the consequences are not apparent immediately. The presented forward modeling similarity measure suggests itself as alternative. From the current state of analysis, its benefits do not have to be paid for by any disadvantages. Special care has been taken that the improved properties do not increase the computational burden. We have referred to the paradigm that led to the similarity measure as forward modeling principle. From the discussion in the previous section, it may be speculated that its capabilities are not yet exhausted, but that it bears potential for other aspects of image matching, too. This work has thoroughly investigated theoretical issues of the presented approach and has proved practicability as far as both the quality of the results and the computational efficiency is concerned.

Appendix A

Summary of Discretized Quantities

This section provides a collection of formulas for the discrete similarity measure and related quantities, as needed for experiments and practical implementations. All results are based on the local least-squares measure (3.1.7). The discretization is performed using the Finite-Sums discretization. The diagonality of the Gauss-Newton term is retained, thus only the diagonal terms are given. The summary is intended to form a consistent set of equations. For example the reverse modeling similarity measure is weighted according to the trapezoidal rule in analogy to the forward similarity. The weighting is reasonable though not of intrinsic importance. Some formulas for single and double grid discretization essentially look the same, they are listed anyway. Some of the differences are hidden behind the operators ∂ and δ. Basically the formulas should reveal the important computational steps for each configuration. A compromise between explicit notation, e.g. writing out arguments, and compactness has to be made however. 1D Configurations

A.0.1 Single Grid Discretization

$$E^h = \frac{1}{2}\sum_{k\in I} a_k (f(m_k) - g_k)^2 \partial m_k \tag{A.0.1a}$$

$$E_i'^h = a_i (f(m_i) - g_i)\partial g_i + \begin{cases} -\frac{1}{2}(f(m_1) - g_1)^2 & i = 0 \\ 0 & I\backslash\partial I \\ \frac{1}{2}(f(m_M) - g_M)^2 & i = M \end{cases} \tag{A.0.1b}$$

$$E_{i,i}^{\mathrm{GN}h} = a_i \frac{(\partial g_i)^2}{\partial m_i} \tag{A.0.1c}$$

A.0.2 Double Grid Discretization

$$E^h = \frac{1}{2}\sum_{k\in I^*} (f(\overline{m}_k) - g_k)^2 \, \delta m_k \tag{A.0.2a}$$

$$E_i'^h = (f(m_i) - \overline{g}_i)\, \delta g_i + \begin{cases} -\frac{1}{2}(f(m_0) - g_{\frac{1}{2}})^2 & i = 0 \\ 0 & I\backslash\partial I \\ \frac{1}{2}(f(x_M) - g_{M-\frac{1}{2}})^2 & i = M \end{cases} \tag{A.0.2b}$$

$$E_{i,i}^{\mathrm{GN}h} = \frac{(\delta g_i)^2}{\partial m_i} \tag{A.0.2c}$$

A.0.3 Reverse Modeling Similarity Measure

$$E^h = \frac{1}{2}\sum_{k\in I} a_k h(f(m_k) - g_k)^2 \tag{A.0.3a}$$

$$\frac{\partial E^h}{\partial m_i} = a_i h(f(m_i) - g_i) f'(m_i) \tag{A.0.3b}$$

$$E^{h\mathrm{GN}}_{i,i} = a_i h f'^2(m_i) \tag{A.0.3c}$$

A.1 2D Configurations

A.1.1 Single Grid Discretization

$$\begin{aligned}
\mathbf{m_k} &= \mathbf{m}(\mathbf{x_k}, u_\mathbf{k}) \\
A_\mathbf{k} &= -\frac{1}{h}p_1(\mathbf{x_k}, u_\mathbf{k})\partial_1 u_\mathbf{k} - \frac{1}{h}p_2(\mathbf{x_k}, u_\mathbf{k})\partial_2 u_\mathbf{k} + p_3(\mathbf{x_k}, u_\mathbf{k}) \qquad \mathbf{k}\in I \\
\tilde{\mathrm{d}}_u g_\mathbf{k} &= -\frac{1}{h}p_1(\mathbf{x_k}, u_\mathbf{k})\partial_1 g_\mathbf{k} - \frac{1}{h}p_2(\mathbf{x_k}, u_\mathbf{k})\partial_2 g_\mathbf{k}
\end{aligned}$$

$$E^h = \frac{1}{2}\sum_{\mathbf{k}\in I} a_\mathbf{k} h^2 (f(\mathbf{m_k}) - g_\mathbf{k})^2 A_\mathbf{k} \tag{A.1.1a}$$

$$E'^h_\mathbf{i} = a_\mathbf{i} h^2 (f(\mathbf{m_i}) - g_\mathbf{i})\,\tilde{\mathrm{d}}_u g_\mathbf{i} + \begin{cases} -h\mathbf{n_i}\cdot\mathbf{p_i}\frac{1}{2}(f(\mathbf{m_i}) - g_\mathbf{i})^2 & \mathbf{i}\in\partial I \\ 0 & I\backslash\partial I \end{cases} \tag{A.1.1b}$$

$$E^{\mathrm{GN}h}_{\mathbf{ii}} = a_\mathbf{i} h^2 \frac{(\tilde{\mathrm{d}}_u g_\mathbf{i})^2}{A_\mathbf{i}} \tag{A.1.1c}$$

A.1.2 Double Grid Discretization

$$\begin{aligned}
\mathbf{m_k} &= \mathbf{m}(\mathbf{x_k}, u_\mathbf{k}) \\
A_\mathbf{k} &= -\frac{1}{h}p_1(\mathbf{x_k}, u_\mathbf{k})\partial_1 u_\mathbf{k} - \frac{1}{h}p_2(\mathbf{x_k}, u_\mathbf{k})\partial_2 u_\mathbf{k} + p_3(\mathbf{x_k}, u_\mathbf{k}) \qquad \mathbf{k}\in I \\
\tilde{\mathrm{d}}_u g_\mathbf{k} &= -\frac{1}{h}p_1(\mathbf{x_k}, u_\mathbf{k})\,\delta_1\, g_\mathbf{k} - \frac{1}{h}p_2(\mathbf{x_k}, u_\mathbf{k})\,\delta_2\, g_\mathbf{k} \\
A^*_\mathbf{k} &= -\frac{1}{h}p_1(\mathbf{x_k}, \overline{u}_\mathbf{k})\,\delta_1\, u_\mathbf{k} - \frac{1}{h}p_2(\mathbf{x_k}, \overline{u}_\mathbf{k})\,\delta_2\, u_\mathbf{k} + p_3(\mathbf{x_k}, \overline{u}_\mathbf{k}) \qquad \mathbf{k}\in I^*
\end{aligned}$$

$$E^h = \frac{1}{2}\sum_{\mathbf{k}\in I^*} h^2 (f(\mathbf{m}(\mathbf{x_k}, \overline{u}_\mathbf{k})) - g_\mathbf{k})^2 A^*_\mathbf{k} \tag{A.1.2a}$$

$$E'^h_\mathbf{i} = a_\mathbf{i} h^2 (f(\mathbf{m_i}) - \overline{g}_\mathbf{i})\,\tilde{\mathrm{d}}_u g_\mathbf{i} + \begin{cases} -h\mathbf{n_i}\cdot\mathbf{p_i}\frac{1}{2}(f(\mathbf{m_i}) - \overline{g}_\mathbf{i})^2 & \mathbf{i}\in\partial I \\ 0 & I\backslash\partial I \end{cases} \tag{A.1.2b}$$

$$E^{\mathrm{GN}h}_{\mathbf{ii}} = a_\mathbf{i} h^2 \frac{(\tilde{\mathrm{d}}_u g_\mathbf{i})^2}{A_\mathbf{i}} \tag{A.1.2c}$$

A.1.3 Reverse Modeling Similarity Measure

$$\mathbf{m_k} = \mathbf{m}(\mathbf{x_k}, u_\mathbf{k}) \qquad \mathbf{k} \in I$$

$$E^h = \frac{1}{2} \sum_{\mathbf{k} \in I} a_\mathbf{k} h^2 (f(\mathbf{m_k}) - g_\mathbf{k})^2 \tag{A.1.3a}$$

$$\frac{\partial E^h}{\partial u_\mathbf{i}} = a_\mathbf{i} h^2 (f(\mathbf{m_i}) - g_\mathbf{i}) \frac{\partial f(\mathbf{m_i})}{\partial u_\mathbf{i}} \tag{A.1.3b}$$

$$E^{h\mathrm{GN}}_{\mathbf{ii}} = a_\mathbf{i} h^2 \left(\frac{\partial f(\mathbf{m_i})}{\partial u_\mathbf{i}} \right)^2 \tag{A.1.3c}$$

Appendix B

Additional Analysis of the Similarity Measure

B.1 First Derivative of the Similarity Measure Revisited

In this section the first derivative of the forward modeling similarity measure is computed starting from the reference space formulation (3.1.12). For the Jacobian $\det \mathbf{m}'$ an alternative expression is used

$$\det \mathbf{m}' = \nabla m_1 \times \nabla m_2 \,. \tag{B.1.1}$$

Accordingly, the similarity measure is written as

$$E = \int_{\mathcal{G}} V \, \nabla m_1 \times \nabla m_2 \, \mathrm{d}A \,. \tag{B.1.2}$$

In this section we adopt a somewhat less precise notation. In particular, the transformation of the local similarity measure from $\mathcal{G} \times \mathcal{F}$ to $\mathcal{G}$ by $V \circ (\mathbf{id}_{\mathcal{G}}, \mathbf{m})$ is not explicitly denoted. The problem is of the same structure as (2.4.7) and the steps to calculate the first derivative are the same as needed to establish the Euler-Lagrange equation (2.4.11). The derivative is easily computed by means of (3.2.3),

$$E'(\mathbf{m})\boldsymbol{\mu} = \int_{\mathcal{G}} V_{\mathbf{y}}^T \boldsymbol{\mu} \, \nabla m_1 \times \nabla m_2 \, \mathrm{d}A + \int_{\mathcal{G}} V \, (\nabla m_1 \times \nabla \mu_2 + \nabla \mu_1 \times \nabla m_2) \, \mathrm{d}A \,. \tag{B.1.3}$$

The goal of the following manipulations is to get rid of the derivatives of $\boldsymbol{\mu}$. The preferred form of E' is that of an inner product. Utilizing a kind of partial integration, the operator ∇ can be shifted away from the increments μ_1 and μ_2. Let u, v, w be scalar fields on $\mathcal{G}$. Then, by Stokes equation the following holds,

$$\begin{aligned} \int_{\mathcal{G}} u \nabla v \times \nabla w \, \mathrm{d}A &= \int_{\mathcal{G}} \nabla(uv) \times \nabla w \, \mathrm{d}A - \int_{\mathcal{G}} \nabla u \times v \nabla w \, \mathrm{d}A \\ &= \int_{\mathcal{G}} \nabla \times (uv \nabla w) \, \mathrm{d}A - \int_{\mathcal{G}} \nabla u \times v \nabla w \, \mathrm{d}A \\ &= \int_{\partial \mathcal{G}} uv \nabla w \cdot \mathrm{d}\mathbf{s} - \int_{\mathcal{G}} \nabla u \times v \nabla w \, \mathrm{d}A \,. \end{aligned} \tag{B.1.4}$$

Applying this to the second part of (B.1.3) yields

$$\int_{\mathcal{G}} V\,(\nabla m_1 \times \nabla \mu_2 + \nabla \mu_1 \times \nabla m_2)\,\mathrm{d}A = \int_{\partial\mathcal{G}} V\,(\mu_1 \nabla m_2 - \mu_2 \nabla m_1) \cdot \mathrm{d}\mathbf{s} - \int_{\mathcal{G}} \nabla V \times (\mu_1 \nabla m_2 - \mu_2 \nabla m_1)\,\mathrm{d}A\,. \quad \text{(B.1.5)}$$

Caution is required in that ∇V has to be read as $\nabla\big(V \circ (\mathbf{id}_{\mathcal{G}}, \mathbf{m})\big)$ and hence

$$\nabla V = V_{\mathbf{x}} + V_{y_1} \nabla m_1 + V_{y_2} \nabla m_2\,. \quad \text{(B.1.6)}$$

Recall that $V_{\mathbf{x}}$ is the column vector $(V_{x_1}, V_{x_2})^T$. Inserting this into (B.1.5), it is easy to see that the terms including the partial derivatives V_{y_i} cancel with the first term of (B.1.3). Stacking all results together leads to

$$E'(\mathbf{m})\boldsymbol{\mu} = \int_{\partial\mathcal{G}} V\,(\mu_1 \nabla m_2 - \mu_2 \nabla m_1) \cdot \mathrm{d}\mathbf{s} - \int_{\mathcal{G}} V_{\mathbf{x}} \times (\mu_1 \nabla m_2 - \mu_2 \nabla m_1)\,\mathrm{d}A\,. \quad \text{(B.1.7)}$$

The result is equivalent to (3.2.14), although written in a slightly different form.

Appendix C

C.1 List of Symbols

The following list of symbols covers the most important quantities and mathematical expressions, but is not comprehensive. Only symbols that are relevant for more than one section are included. Related quantities are grouped and listed roughly in order of appearance in the text.

$P \rightleftharpoons Q$	Point P corresponds to point Q, and vice versa.
C	geometric connection
$\mathcal{G}$	reference domain
$\mathcal{F}$, $\mathcal{F}_i$	image domain
$\partial\mathcal{G}$	boundary of domain $\mathcal{G}$
g	surface texture, surface intensity
f	image intensity, image greyvalue
$\mathbf{x}$	coordinate vector in $\mathcal{G}$. Vector-valued quantities are denoted in boldface. The usual case for this work is $\mathbf{x} \in \mathbb{R}^2$.
$\mathbf{y}$	coordinate vector in $\mathcal{F}$
$\mathbf{m}$	map from $\mathcal{G}$ to $\mathcal{F}$
$\mathbf{m}'^{\text{ad}}$	adjunct Jacobi matrix of $\mathbf{m}$
$\text{vis}_{\mathbf{m}}$	visibility indicator for the map $\mathbf{m}$
$\mathbf{id}_{\mathcal{G}}$	identity mapping on the set $\mathcal{G}$
u	generic matching function, i.e. scalar parameter function of $\mathbf{m}$
E_{img}	similarity measure, image energy
E_{def}	regularization term, deformation energy
V	local similarity measure
α	regularization parameter, weight of the regularization term
$E'(u)v$	first derivative of E at u applied to v
$\delta E(u; v)$	variation of E at u in direction v
$\nabla E(u)$	gradient of E at u
$E^{\text{GN}}(u)vv$	Gauss-Newton term of E at u, quadratic in v.
$\mathrm{d}A$	differential element of area
$\mathbf{ds}$	differential element of arc length
$< ., . >$	scalar product $X \times X \mapsto \mathbb{R}$, for some vector space X
$\|.\|_{p,\mathcal{F}}$	L^p norm on domain $\mathcal{F}$
$\tilde{\mathrm{d}}_u g$	reference space derivative of g for the matching function u
$\mathbf{p}$	vector of coefficients for the reference space derivative, $\tilde{\mathrm{d}}_u g = -g_{\mathbf{x}}^T \mathbf{p}$
h	grid spacing h

$\mathcal{G}_h$	grid with spacing h
$\mathcal{G}_h^*$	dual grid of $\mathcal{G}_h$
I	index set of $\mathcal{G}_h$
I^*	index set of $\mathcal{G}_h^*$
∂I	boundary elements of index set I
$a_{\mathbf{k}}$	weight of the trapezoidal rule at $\mathbf{k}$
$\mathbf{n}_{\mathbf{k}}$	normal vector of the boundary of $\mathcal{G}_h$
$\partial g_{\mathbf{k}}$	symmetric difference of g at $\mathbf{k}$
$\delta g_{\mathbf{k}}$	central difference of g at $\mathbf{k}$
$\overline{g}_{\mathbf{k}}$	central mean of g at $\mathbf{k}$
$E(u^*) = \min!$	E assumes its minimum at u^*.
$u^{(k)}$	k'th element of a sequence of matching functions, as generated by an iterative optimization method.
$L^{(k)}$	vector of the linear approximation of E_{img} at iteration state k
$Q^{(k)}$	matrix of the quadratic approximation of E_{img} at iteration state k
R	matrix of E_{def}
$a \propto b$	Quantity a is proportional to b by a factor.
$f(x) \circ\!\!-\!\!\bullet F(\theta)$	$f(x)$ and $F(\theta)$ form a Fourier pair.
$\wedge$	logical AND
$\vee$	logical OR
$\sqcup$	logical XOR (exclusive or)

C.2 Abbreviations

CCD	Charge Coupled Device
DLT	Direct Linear Transform
FE	Finite-Elements
GNC	Graduated Non-Convexity
IDL	Interactive-Data-Language (by Research Systems Inc.)
LSI	Linear Shift Invariant
MAP	Maximum a Posteriori (estimation)
PDF	Probability Distribution Function
SAR	Synthetic Aperture Radar
VLL	Vertical Line Locus

Optimization

GN	Gauss-Newton method
GNJ	Gauss-Newton-Jacobi method
PW	Powell-Wolfe stepsize rule
SD	Steepest Descent method

Bibliography

[1] Y. Amit. A nonlinear variational problem for image matching. *SIAM J. Scient. Comp.*, 15(1):207–224, Jan. 1994.

[2] Y. Amit, U. Grenader, and M. Piccioni. Structural image restoration through deformable templates. *J. Amer. Statist. Assoc.*, 86(414):376–387, June 1991.

[3] J.-P. Aubin. *Applied Abstract Analysis.* Pure and Applied Mathematics. Wiley, New York, 1977.

[4] J.-P. Aubin. *Applied Functional Analysis.* Pure and Applied Mathematics. Wiley, New York, 1979.

[5] P. N. Belhumeur. A bayesian approach to binocular stereopsis. *Int. J. Comp. Vision.*, 19(3):237–260, 1996.

[6] Å. Björck. *Numerical Methods for Least Squares Problems.* SIAM, Philadelphia, 1996.

[7] A. Blake and A. Zissermann. *Visual Reconstruction.* MIT Press, Campbridge, MA, 1987.

[8] P. J. Burt. Fast filter transforms for image processing. *Computer Graphics and Image Processing*, 16:20–51, 1981.

[9] P. J. Burt and E. H. Adelson. The laplacian pyramid as a compact image code. *IEEE Trans. on Communications*, 31(4):532–540, Apr. 1983.

[10] I. J. Cox, S. L. Hingorani, S. B. Rao, and B. M. Maggs. A maximum likelihood stereo algorithm. *Comp. Vision and Image Understanding*, 63(3):542–567, May 1996.

[11] N. D'Apuzzo. Automated photogrammetric measurement of human faces. In *Int. Arch. Photogrammetry and Remote Sensing*, volume XXXII-B5, pages 402–407, 1998.

[12] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Comm. Pure and Appl. Math.*, 41:909–996, 1988.

[13] H. J. Dirschmid. *Mathematische Grundlagen der Elektrotechnik.* Vieweg, Braunschweig/-Wiesbaden, 2 edition, 1987.

[14] H. J. Dirschmid. *Tensoren und Felder.* Springer, Wien, New York, 1996.

[15] R. Eier. *Analyse und Synthese von diskreten Zufallsprozessen mit Hilfe von Markoffschen Ketten.* Verband d. wissenschaftl. Gesellschaften Österreichs Verlag, Wien, 1973.

[16] W. Förstner. On the geometric precision of digital correlation. In *Int. Arch. Photogrammetry and Remote Sensing*, volume XXIV Com. II, pages 176–189, 1982.

[17] W. Förstner. A feature based correspondence algorithm for image matching. In *Int. Arch. Photogrammetry and Remote Sensing*, volume 26 Part 3/3, pages 150–166, Rovaniemi, Finland, Aug. 1986.

[18] M. A. Gennert. Brightness-based stereo matching. In *Proc. of 2nd Int. Conf. on Computer Vision*, pages 139–143, Tampa, FL, 1988. IEEE.

[19] G. L. Gimel'farb. Symmetric bi- and trinocular stereo: Tradeoffs between theoretical foundations and heuristics. In W. Kropatsch, R. Klette, and F. Solina, editors, *Theoretical Foundations of Computer Vision*, volume 11 of *Computing Supplement*, pages 53–71. Springer, Wien, 1996.

[20] A. Gruen and H. Kahmen, editors. *Optical 3D-Measurement Techniques IV*, Zürich, Oct. 1997. Wichmann, Heidelberg.

[21] W. Hackbusch. *Iterative Lösung großer schwachbesetzer Gleichungssysteme.* Teubner-Studienbücher: Mathematik, Stuttgart, 2 edition, 1993.

[22] C. Heipke. *Integration von Bildzuordnung, Punktbestimmung, Oberflächenrekonstruktion und Orthoprojektion innerhalb der digitalen Photogrammetrie.* Number 366 in Reihe C. Deutsche Geodätische Kommission, Bayr. Akad. d. Wiss., München, 1990.

[23] C. Heipke. A global approach for least-squares image matching and surface reconstruction in object space. *Photogram. Engineering & Rem. Sensing*, 58(3):317–323, Mar. 1992.

[24] P. Henrici. *Elements of Numerical Analysis.* Wiley, New York, 3 edition, 1967.

[25] G. Hermanson, J. Hinchman, U. Rauhala, and W. Mueller. Comparison of two terrain extraction algorithms: Hierarchical Relaxation Correlation and Global Least Squares Matching. In *SPIE Proccedings*, volume 1944, pages 31–43, Orlando, Apr. 1993.

[26] B. K. P. Horn. *Robot Vision.* Electrical engineering and computer science series. MIT Press, Massachusetts, 7 edition, 1991.

[27] J.-M. Jolion and A. Rosenfeld. *A Pyramid Framework for Early Vision.* Series in engineering and computer science. Kluwer, Dordrecht, Netherlands, 1994.

[28] H. Kager. ORIENT: A universal photogrammetric adjustment system. In A. Gruen and H. Kahmen, editors, *Optical 3D-Measurement Techniques*, pages 447–455, Zürich, 1989. Wichmann, Karlsruhe.

[29] T. Korten, B. Wrobel, M. Franek, and M. Weisensee. Experiments with facets stereo vision (FAST-vision) for object surface reconstruction. In *Int. Arch. Photogrammetry and Remote Sensing*, volume 27/B3, pages 396–404, Kyoto, 1988.

[30] P. Kosmol. *Methoden zur numerischen Behandlung nichtlinearer Gleichungen und Optimierungsaufgaben.* Teubner-Studienbücher: Mathematik, Stuttgart, 1989.

[31] K. Kraus. *Photogrammetry. Volume 1: Fundamentals and Standard Processes.* Dümmler, Bonn, 4 edition, 1993.

[32] K. Kraus. *Photogrammetry. Volume 2: Advanced Methods and Applications.* Dümmler, Bonn, 4 edition, 1997.

[33] K. Kraus and W. Schneider. *Fernerkundung. Band 1: Physikalische Grundlagen und Aufnahmetechniken.* Dümmler, Bonn, 1988.

[34] P. Krzystek and F. Petran. Automatic reconstruction of concept models by using a digital photogrammetric measurement system. In *Int. Arch. Photogrammetry and Remote Sensing*, volume 30-5W1, From Pixels to Sequences, pages 176–185, Zürich, Mar. 1995.

[35] E. A. Lee and D. G. Messerschmitt. *Digital Communication.* Kluwer, Boston, 4 edition, 1992.

[36] S. Z. Li. *Markov Random Field Modeling in Computer Vision.* Computer Science Workbench. Springer, Tokyo, 1995.

[37] A. Luo and H. Burkhardt. An intensity-based cooperative bidirectional stereo matching with simultaneous detection of discontinuities and occlusions. *Int. J. Comp. Vision.*, 15(3):171–188, 1995.

[38] H.-G. Maas. Robust automatic surface reconstruction with structured light. In *Int. Arch. Photogrammetry and Remote Sensing*, volume XXIX-B5, pages 709–713, Washington D.C., 1992.

[39] K. Oda, H. Kano, and T. Kanade. Generalized disparity and its application for multistereo camera calibration. In Gruen and Kahmen [20], pages 109–116.

[40] C. Öhreneder. Global least squares matching using a steepest descent method. In W. Burger and M. Burge, editors, *Pattern Recognition 1997, 21st Workshop of the Austrian Association for Pattern Recoginition*, pages 79–85, Hallstatt, Upper Austria, May 1997. R. Oldenbourg, Wien, München.

[41] C. Öhreneder. Stochastic modeling and a steepest-descent method for global least squares matching. In Gruen and Kahmen [20], pages 53–60.

[42] S. I. Olsen. Stereo correspondence by surface reconstruction. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 12(3):309–315, Mar. 1990.

[43] H.-P. Pan. General stereo image matching using symmetric complex wavelets. In M. A. Unser, A. Aldroubi, and A. F. Laine, editors, *Wavelet Applications in Signal and Image Processing IV*, volume 2825 of *SPIE Proceedings*, pages 697–720, Denver, Colorado, Oct. 1996.

[44] H.-P. Pan. Uniform full-information image matching using complex conjugate wavelet pyramids. In *Int. Arch. Photogrammetry and Remote Sensing*, volume XXXI-B3, pages 619–625, Vienna, 1996.

[45] A. L. Peressini, F. E. Sullivan, and J. J. Uhl, jr. *The Mathematics of Nonlinear Programming.* Undergraduate texts in mathematics. Springer, New York, 1988.

[46] D. K. Pickard. A curious binary lattice process. *J. Appl. Prob.*, 14:717–731, 1977.

[47] D. K. Pickard. Unilateral markov fields. *Adv. Appl. Prob.*, 12:655–671, 1980.

[48] E. R. Pinch. *Optimal control and the calculus of variations.* Oxford Univ. Press, 1995.

[49] L. Robert and R. Deriche. Dense depth map reconstruction: A minimization and regularization approach which preserves discontinuities. In B. Buxton and R. Cipolla, editors, *Computer vision : ECCV '96 ; 4th European Conference on Computer Vision*, Lecture notes in computer science, pages 439–451, Cambridge, UK, Apr. 1996. Springer.

[50] D. Rosenholm. *Accuracy Improvement in Digital Matching.* Number 52 in Fotogrammetriska meddelanden Photogr. Rep. The Royal Inst. of Technology Dep. of Photogrammetry, Stockholm, 1986.

[51] T. Saito, H. Kudo, T. Anan, and C. Iganami. A stereo matching algorithm based on energy minimization principle in markov random field model. In *Statistical and Stochastical Methods for Image Processing*, volume 2823 of *SPIE Proceedings*, pages 128–137, Denver, Colorado, Aug. 1996.

[52] R. Schaback. Convergence analysis of the general Gauss-Newton algorithm. *Numer. Math.*, 46:281–309, 1985.

[53] C. Schnörr, R. Sprengel, and B. Neumann. A variational approach to the design of early vision algorithms. In W. Kropatsch, R. Klette, and F. Solina, editors, *Theoretical foundations of computer vision*, number 11 in Computing Supplement, pages 149–165. Springer, Wien, 1996.

[54] H. R. Schwarz. *Numerik Symmetrischer Matrizen.* Teubner, Stuttgart, 1968.

[55] D. Terzopoulos. Regularization of inverse visual problems involving discontinuities. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 8(4):413–424, July 1986.

[56] A. N. Tikhonov and V. Y. Arsenin. *Solutions of Ill-Posed Problems.* Scripta Series in Mathematics. V. H. Winston & Sons, Washington, D.C., 1977.

[57] H. Weinrichter and F. Hlawatsch. *Stochastische Grundlagen nachrichtentechnischer Signale.* Springer, Wien, New York, 1991.

[58] M. Weisensee. *Modelle und Algorithmen für das Facetten-Stereosehen.* Number 374 in Reihe C. Deutsche Geodätische Kommission, Bayr. Akad. d. Wiss., München, 1992.

[59] J. Werner. *Numerische Mathematik 1.* Vieweg, Braunschweig/Wiesbaden, 1992.

[60] G. Winkler. *Image Analysis, Random Fields and Dynamic Monte Carlo Methods: A Mathematical Introduction.* Applications of Mathematics. Springer, Berlin, 1995.

[61] B. P. Wrobel. Facets stereo vision (FAST vision) - an new approach to computer stereo vision and to digital photogrammetry. In *Proc. Fast processing of Photogrammetric Data*, pages 231–258, Interlaken, 1987.

[62] B. P. Wrobel. Least squares methods for surface reconstruction from images. In *Int. Arch. Photogrammetry and Remote Sensing*, volume 27/B3, pages 806–820, Kyoto, 1988.

[63] E. Zeidler, editor. *Teubner-Taschenbuch der Mathematik.* Teubner, Stuttgart-Leipzig, 1996.

[64] E. Zeidler et al., editors. *Teubner-Taschenbuch der Mathematik Teil II.* Teubner, Stuttgart-Leipzig, 7 edition, 1995.

[65] Z. Zhang. Parameter estimation techniques: A tutorial with application to conic fitting. Rapport de Recherche 2676, INRIA, Sophia-Antipolis, Oct. 1995.

[66] Z. Zhang. Parameter estimation techniques: A tutorial with application to conic fitting. *Image Vision Comp.*, 15(1):59–76, Jan. 1997. Available as tech. rep. [65].

Curriculum Vitae

Name	Dipl.-Ing. Christian Michael Öhreneder
Date and place of birth	January 31, 1971. Vienna, Austria
Address	Knabstraße 5/1, A-3013 Tullnerbach
Email	`co@ipf.tuwien.ac.at`
Wife	Mag. Gabriele Öhreneder
Parents	Christine and Dipl.-Ing. Hermann Öhreneder
Education	1977-1981 primary school, Purkersdorf 1981-1989 secondary school, BRG XIII Wenzgasse, Vienna graduated with excellence on Jun. 5, 1989
University	1989-1995 study of Electrical Engineering - Communications and Information Technology at the Vienna University of Technology (TU Wien) Diploma thesis in the field of microwave remote sensing titled *Funktionsanalyse des Mikrowellen-Radio-Scatterometers RSM und Signatur-Messung über alpiner Grasheide.* Performed in the course of the "Highalpine SAR Experimen" of the NASA Shuttle-Radar-Mission SIR-C/X-SAR (Oct. 94) at the Inst. of Meteorology and Geophysics of the Univ. of Innsbruck and at the Inst. of Communications and Radio-Frequency Engineering (TU Wien), with support of the Inst. of Photogrammetry and Remote Sensing (TU Wien). graduated with excellence on Jun. 22, 1995 1995-1999 doctoral study at TU Wien, Inst. of Photogrammetry and Remote Sensing. Supported by the Austrian Academy of Science Doctoral Scholarship for highly qualified students. In collaboration with the Austrian Science Fund project, no. S7004 *Theory and Application of Digital Image Processing and Pattern Recognition/Stereovideometry and Spatial Object Recognition.*
Civil service	Oct. 1, 97 - Sep. 30, 98 at the Red Cross
Research interests	Image and Signal Processing, Computer Vision, Applied Mathematics

GEOWISSENSCHAFTLICHE MITTEILUNGEN

Bisher erschienen:

Heft 1 Kolloquium der Assistenten der Studienrichtung Vermessungswesen. 1970-1973, Dezember 1973.

Heft 2 EGGER-PERDICH-PLACH-WAGENSOMMERER, Taschenrechner HP 45 und HP 65, Programme und Anwendungen im Vermessungswesen. 1. Auflage, März 1974, Special Edition in English, Juli 1974, 2. verbesserte Auflage, November 1974.

Heft 3 Kolloquium der Assistenten der Studienrichtung Vermessungswesen 1973-1974, September 1974.

Heft 4 EGGER-PALFINGER-PERDICH-PLACH-WAGENSOMMERER, Tektronix-Tischrechner TEK 31, Programmbibliothek für den Einsatz im Vermessungswesen, November 1974.

Heft 5 K. LEDERSTEGER, Die horizontale Isostasie und das isostatische Geoid, Februar 1975.

Heft 6 F. REINHART, Katalog von FK4 Horrebow-Paaren für Breiten von +30 bis +60, Oktober 1975.

Heft 7 Arbeiten aus dem Institut für Höhere Geodäsie, Wien, Dezember 1975.

Heft 8 Veröffentlichungen des Instituts für Photogrammetrie zum XIII. Internationalen Kongreß für Photogrammetrie in Helsinki 1976, Wien, Juli 1976.

Heft 9 W. PILLEWIZER, Felsdarstellung aus Orthophotos, Wien, Juni 1976.

Heft 10 PERDICH-PLACH-WAGENSOMMERER, Der Einsatz des programmierbaren Taschenrechners Texas Instruments SR-52 mit Drucker PC100 in ingenieurgeodätischen Rechentechnik, Wien, Mai 1976.

Heft 11 Kolloquium der Assistenten der Studienrichtung Vermessungswesen 1974-1976, November 1976.

Heft 12 Kartographische Vorträge der Geodätischen Informationstage 1976, Wien, Mai 1977.

Heft 13 Veröffentlichung des Instituts für Photogrammetrie anläßlich des 80. Geburtstages von Prof. Dr. h. c. K. Neumaier, Wien, Januar 1978.

Heft 14 L. MOLNAR, Self Checking Analytical Relative Orientation and Strip Formation, Wien, Dezember 1978.

Heft 15 Veröffentlichung des Instituts für Landesvermessung anläßlich des 80. Geburtstages von Prof. Dr. Alois Bavir, Wien, Januar 1979.

Heft 16 Kolloquium der Assistenten der Studienrichtung Vermessungswesen 1976-1978, Wien, November 1979.

Heft 17 E. VOZIKIS, Die photographische Differntialumbildung gekrümmter Flächen mit Beispielen aus der Architekturbildmessung, Wien, Dezember 1979.

Heft 18 Veröffentlichung des Instituts für Allgemeine Geodäsie anläßlich des 75. Geburtstages von Prof. Dipl.-Ing. Dr. F. Hauer, Die Höhe des Großglockners, Wien, 1981.

Heft 19 H. KAGER, Bündeltriangulation mit indirekt beobachteten Kreiszentren, Wien, April 1981.

Heft 20 Kartrographische Vorträge der Geodätischen Informationstage 1980, Wien, Mai 1982.

Heft 21 Veröffentlichung des Instituts für Kartographie anläßlich des 70. Geburtstages von Prof. Dr. Wolfgang Pillewizer: Glaziologie und Kartographie, Wien, Dezember 1982.

Heft 22 K. TEMPFLI, Genauigkeitsschätzung digitaler Höhenmodelle mittels Spektralanalyse, Wien, Mai 1982.

Heft 23 E. CSAPLOVICS, Interpretation von Farbinfrarotbildern, Wien, November 1982.

Heft 24 J. JANSA, Rektifizierung von Multispektral-Scanneraufnahmen - Entwicklung und Erprobung eines EDV-Programms, Wien, Mai 1983.

Heft 25 Zusammenfassung der Diplomarbeiten, Dissertationen und Habilitationen an den geodätischen Instituten der TU Wien, Wien, November 1984.

Heft 26 T. WUNDERLICH, Die voraussetzungsfreie Bestimmung von Refraktionswinkeln, Wien, August 1985.

Heft 27 G. GERSTBACH (Hrsg.), Geowissenschaftliche/geotechnische Daten in Landinformationssystmen - Bedarf und Möglichkeiten in Österreich, Juni 1986.

Heft 28 K. NOVAK, Orientierung von Amateuraufnahmen ohne Paßpunkte, Wien, August 1986.

Heft 29 Veröffentlichung des Instituts für Landesvermessung und Ingenieurgeodäsie, Abt. Ingenieurgeodäsie, anläßlich des 80. Geburtstages von Prof. Dipl.-Ing. Dr. F. Hauer, Wien, Oktober 1986.

Heft 30 K.-H. ROCH, Über die Bedeutung dynamisch ermittelter Parameter für die Bestimmung von Gesteins- und Gebirgseigenschaften, Wien, Februar 1987.

Heft 31 G. HE, Bildverbesserung mittels digitaler Filterung, Wien, April 1989.

Heft 32 F. SCHLÖGELHOFER, Qualitäts- und Wirtschaftlichkeitsmodelle für die Ingenieurphotogrammetrie, Wien, April 1989.

Heft 33 G. GERSTBACH (Hrsg.), Geowissenschaftliche/geotechnische Daten in Landinformationssystemen - Datenbestände und Datenaustausch in Österreich, Wien, Juni 1989.

Heft 34 F. HOCHSTÖGER, Ein Beitrag zur Anwendung und Visualisierung digitaler Geländemodelle, Wien, Dezember 1989.

Heft 35 R. WEBER, Lokale Schwerefeldmodellierung unter Berücksichtigung spektraler Methoden zur Geländereduktion, Wien, April 1990.

Heft 36 o. Prof. Dr. Hans Schmid zum 70. Geburtstag. Veröffentlichung der Abteilung für Landesvermessung, Wien, Oktober 1990.

Heft 37 G. GERSTBACH, H. P. HÖLLRIEGL und R. WEBER, Geowissenschaftliche Informationsbörse - Eine Nachlese zu GeoLIS II, Wien, Oktober 1990.

Heft 38 R. ECKER, Rastergraphische Visualisierungen mittels digitaler Geländemodelle, Wien, August 1991.

Heft 39 Kartographische Forschungen und Anwendungsorientierte Entwicklungen, herausgegeben von W. Stams und F. Kelnhofer zum 80. Geburtstag von Prof. Dr. W. Pillewizer, Wien, Juli 1991.

Heft 39a W. RIEGER, Hydrologische Anwendungen des digitalen Geländemodelles, Wien, Juli 1992.

Heft 40 K. STEINNOCHER, Methodische Erweiterungen der Landnutzungsklassifikation und Implementierung auf einem Transputernetzwerk, Wien, Juli 1994.

Heft 41 G. FORKERT, Die Lösung photogrammetrischer Orientierungs- und Rekonstruktionsaufgaben mittels allgemeiner kurvenförmiger Elemente, Wien, Juli 1994.

Heft 42 M. SCHÖNER, W. SCHÖNER, Photogrammetrische und glaziologische Untersuchungen am Gsbre (kErgebnisse der Spitzbergenexpedition 1991), Wien, Mai 1996.

Heft 43 M. ROIC, Erfassung von nicht signalisierten 3D-Strukturen mit Videotheodoliten, Wien, April 1996.

Heft 44 G. RETSCHER, 3D-Gleiserfassung mit einem Multisensorsystem und linearen Filterverfahren, Wien, April 1996.

Heft 45 W. DAXINGER, Astrogravimetrische Geoidbestimmung für Ingenieurprojekte, Wien, Juli 1996

Heft 46 M. PLONER, CCD-Astrometrie von Objekten des geostationären Ringes, Wien, November 1996

Heft 47 Zum Gedenken an Karl Killian "Ingenieur" und "Geodät" 1903-1991, Veröffentlichung der Fachgruppe Geowissenschaften, Wien, Februar 1997.

Heft 48 A. SINDHUBER, Ergänzung und Fortführung eines digitalen Landschaftsmodelles mit multispektralen und hochauflösenden Fernerkundungsaufnahmen, Wien, Mai 1998.

Heft 49 W. WAGNER, Soil Moisture Retrieval from ERS Scatterometer Data, Wien, Dezember 1998.

Heft 50 R. WEBER, E. FRAGNER (Editoren), Prof. Bretterbauer, Festschrift zum 70. Geburtstag, Wien, Juli 1999