ON DIGITAL IMAGE INVERSION

Yong-Jian Zheng
Institute for Photogrammetry and Remote Sensing, University Karlsruhe
Englerstrasse 7, D-7500 Karlsruhe 1, Germany
Email: zheng@ipf.bau-verm.uni-karlsruhe.de

Abstract

Image understanding is the enterprise of automating and integrating a wide range of processes and representations used for vision perception. It includes techniques not only for geometric modeling but also for inference and reasoning. In this paper, we look at the issue of inductive inference in digital image understanding. Many goals in both low-level and high-level image analysis can be formulated generally as a problem of inferring object-properties from image data, having assistance of *a priori* knowledge. This process of information processing would be called image inversion, as the desired information about the objects is derived from image data. Based on the inverse problem theory, we provide a sound theoretical basis for determination of generalizations, descriptions, rules, and laws, from a set of raw data, observations, features or facts. To demonstrate this approach, we present its application in the limited domain of surface reconstruction from multiple images.

1 Introduction

Computer Vision includes techniques not only for geometric modeling but also for inference and reasoning. Many of its tasks require the ability to create explicit representations of knowledge from implicit ones and they can be therefore formulated as problems of inference drawing. Drawing inference from image data is only plausible as the available information is incomplete or inexact and it is inadequate to support the desired sorts of logical inferences.

Plausible inference is a basic issue, of which we are all aware through our own experience in research on many vision problems, including feature extraction, image and boundary segmentation, object reconstruction and image interpretation. In these cases, problem solvers have to reason with inconsistent and incomplete information on the basis of beliefs, not only true (or false) facts.

In this paper, we think of inference drawing from digital images as an inverse process which we call *digital image inversion* (Zheng, 1990). Drawing plausible inference is therefore solving ill-posed inverse problems. Although this kind of problems have been considered for a long time almost exclusively as mathematical curiosities, it is now clear that many inverse problems have ill-posed nature and their solutions are of great practical interest (Herman, 1980; Fawcett, 1985; Poggio et al., 1985). To deal with ill-posed inverse problems, one has to deal with several questions including: What is the nature of inverse problems? How about their solvability? How to integrate *a priori* knowledge to deal with the ill-posedness of inverse problems? And how to evaluate the quality of solutions?

We begin by introducing a paradigm for digital image inversion, which has three steps of representation, forward modeling and inversion. Then, we discuss the theory of inductive inference and inverse problems. Based on the Maximum A Posteriori (MAP) we describe a framework for integrating *a priori* knowledge to solve the decision problem in the ill-posed inverse process. Next, we formulate the problem of surface reconstruction from digital images within this framework. We then demonstrate shortly the result of our implementation and experiment results using real image data.

2 Digital Image Inversion

Generally, vision can be regarded as an inference process in which a description of the outside world is inferred from images of the world, having the aid of *a priori* knowledge about the world and about imaging process. Here, three kinds of information have to be dealt with, i.e. the desired information about the outside world, the available information contained in images, and the a priori information of image interpreters.

Now, let S represent a physical system (for instance the earth's surface, or an object in an image). Assume that we are able to define a set of model parameters which completely describes S, to some extent. These parameters may not all be directly measurable. We

can operationally define a set of some observable parameters $\mathcal{Y}$ whose actual values hopefully are relatable to a set of the model parameters $\mathcal{X}$. To solve the forward problem is to predict the values of the observable parameters $Y \in \mathcal{Y}$, given arbitrary values of the model parameters $X \in \mathcal{X}$. To solve the inverse problem is to infer the values of the model parameters X from given observed values of the observable parameters Y (cf. Tarantola, 1987).

Obviously, many problems in computer vision can be formulated as such inverse problems (a particular kind of inference process called induction). The scientific procedure to solve these inverse problems distinguishes the following three steps:

1. **Representation** (parameterization) of System S: Designing a language to represent the characteristic features of S. That is, establishing a minimal set of model parameters X whose values completely characterize the system (from a given point of view).
2. **Forward modeling**: Identification of the physical laws (constraints) which, for given values of model parameters X, allow predictions as to the results of measurements on some observable parameters Y.
3. **Inversion**: Use of the actual results of some measurements of the observable parameters to infer (estimate) the actual values of the model parameters.

3 Inductive Inference

The term *inference* refers generally to effective procedures for deriving new facts from known ones. To draw an inference is to come to believe a new fact on the basis of other information. There are many kinds of inference. The best understood is *deduction*, which proceeds from a set of assumptions called axioms to new statements that are logically implied by the axioms. The deductive inference is logically correct as deduction from true premises is guaranteed to result in a true conclusion. The standard way to characterize deduction is by using a system called predicate calculus which consists of a language for expressing propositions and rules for how to infer new facts (propositions) from those we already have. To deduce new facts from the axioms, we use one or more so called *rules of inference.*

A second kind of inference, on the other hand, is called *induction*, which is a calculus for inferring generalizations from particular observations. This inductive inference process could be thought of as having the form " from: if $(X \rightarrow Y)$ and Y, infer: X " and it performs abstraction, producing generalities from specifics. The inductive inference can be illustrated using a simple example of geometrical reasoning from which we wish to answer a question:

Variables		Premises		Conclusions	
X	Y	$X \rightarrow Y$	$\neg Y$	X	$\neg X$
T	T	T	F	T	F
T	F	F	T	T	F
F	T	T	F	F	T
F	F	T	T	F	T

Table 1: Truth table used to draw inference

> Given a set of geometrical points $\mathcal{P} = \{x_i, y_i\}$, $i = 1, ..., n$. Infer if this set of points depicts a straight line.

To answer this question, we can use some statements that express information during inference:

> If $\mathcal{P}$ represents a straight line, then $y = a\ x\ +\ b$ is valid for all points of $\mathcal{P}$, where a and b are two constants. Some points of $\mathcal{P}$ do not fulfill $y = a\ x\ +\ b$. Does $\mathcal{P}$ depicts a straight line?

In order to express these statements, we have to agree on a suitable set of atomic propositions like:

- X: $\mathcal{P}$ depicts a straight line.
- Y: $y = a\ x\ +\ b$ is valid for all points of $\mathcal{P}$.

The original statements expressing information during inference are called premises and can be described as follows:

$$X \rightarrow Y,\ \neg Y$$

So, the question would be answered if we could prove the proposition X from the premises, or alternatively if we could prove $\neg X$. Since this is a small problem, we can easily employ an exhaustive examination of all possible assignments of truth values to the propositions X and Y to check for the validity of either possible conclusion. Using the so called truth table (cf. Tab. 1) we can list all the possible combinations. Let us first check the validity of X as a conclusion by examining every row in which all two premises are truth. In this example there is only one row where all premises are truth (the bottom row). It is intuitive that the potential conclusion X is false here whereas $\neg X$ is true and this corresponds to the correct answer: $\mathcal{P}$ does not depict a straight line.

The reasoning method just illustrated is called *perfect induction*. Here, the available information which is necessary to support the desired logical inference is *perfect*. This means that all the statements only have two values for their validities, either true or false, and we are able to check exhaustively all the possible combinations. Unfortunately, in many practical problem-solving situations, especially in image analysis and understanding, the available knowledge is incomplete or inexact. The coordinates of the points in $\mathcal{P}$, for example, may contain measure-errors and only from the fact that some points of $\mathcal{P}$ do not fulfill $y = a\ x\ +\ b$ we can not come to the conclusion that $\mathcal{P}$ does not depict a straight line. So, the validity of Y (or $\neg Y$) is not binary and not easy to prove if we do not have knowledge about measure-errors. In cases like this, we need reasoning methods in making *just* decisions.

4 Inverse Problems

Mathematically, the inverse problem can be described as follows. Given a mapping f from set $\mathcal{X}$ into set $\mathcal{Y}$, i.e. $f : \mathcal{X} \rightarrow \mathcal{Y}$. The solution of the inverse problem consists in the interpretation of data $Y \in \mathcal{Y}$ in order to recover the original image $X \in \mathcal{X}$. This is exactly the same goal as that of an inductive inference mentioned above.

Let us now consider a linear mapping $A : \mathcal{X} \rightarrow \mathcal{Y}$. The inverse problem is to identify X from the *data* Y:

$$A\,X = Y. \tag{1}$$

The solvability of this inverse problem could be discussed as follows:

- If A is bijective and A^{-1} is stable one can easily get an unique solution $X = A^{-1}Y$.
- If A is injective but not surjective, the inverse problem is *overdetermined* and has no solution. One can, however, get an unique pseudo solution through minimizing the norm of the residual $\|V\| = \|Y - AX\|$.
- If A is not injective, the inverse problem is *underdetermined* and there is an unique pseudo solution $X = A^{+}Y$, where A^{+} is the so called Moore-Penrose Inversion, which, unfortunately, is only stable if the domain of A is closed in Y.

So, it is quite clear that the ambivalent non-injective inverse problem is practical not solvable through a numeric process, as any few errors in Ycan destroy the solution totally.

Schematically, there are two reasons for the ill-posedness of inverse problems: intrinsic lack of data, and observation uncertainties. With additional information, for instance some *a priori* assumptions on model parameters X or an additional data set, many such problems can be reformulated into well-posed solvable problems. Now, the main question is how to integrate a priori knowledge to solve ill-posed inverse problems. We need criteria to impose constrains on the solution space and a framework to integrate *a priori* knowledge in order to select an unique solution (the so called *best* solution) for given data. Intuitively, the best solution exists only in connection with criteria which are, of course, strongly task dependent.

5 The MAP Criteria

The first criterion which we would introduce in this section is the so called Maximum A Posteriori (MAP) criterion which is based on probability theory (Geman and Geman, 1984). It selects as the best solution the model parameters X that maximizes the conditional probability of X given the data Y: $P(X \mid Y)$, subject to the inverse problem (1). The MAP criterion leads to three important estimation methods, namely the bayesian estimate method (BE), the maximum likelihood method (ML) and least squares method (LS) (cf. Tab. 4), which are widely used in data processing. Using Bayes' theorem gives

$$P(X \mid Y) = P(Y \mid X)P(X)/P(Y), \tag{2}$$

where $P(Y \mid X)$ is the conditional probability of getting data Y given the model parameters X, $P(X)$ are the prior probability of X. The relation (2) shows how the *prior probability* $P(X)$ changes to the *posterior probability* $P(X \mid Y)$ as a result of acquiring new information Y. Intuitively, the MAP criterion will choose X that maximizes

$$P(Y \mid X)P(X), \tag{3}$$

if $P(Y)$ is constant. This is the principle of the Bayesian estimation. Further, under the specification that the prior probabilities $P(X)$ are all the same, i.e. $P(X)$ is constant, the MAP criterion leads to the simpler maximum likelihood principle of selecting that X which maximizes $P(Y \mid X)$. If the random variables to which the data Y refer are normally distributed, the maximum likelihood estimation will give the same results as the least squares estimation which has widely been used in different branches of science and engineering for over a century and a half. If V is the vector of observational residuals, for which $E(V) = 0$, and which is assumed to be normally distributed, and Σ is the covariance matrix of the distribution, then we have

$$P(Y \mid X) = P(V) = C \cdot \exp\left[-\frac{1}{2}V^{T}\Sigma^{-1}V\right], \tag{4}$$

criterion	expression	supposition
MAP	$P(X \mid Y) \rightarrow max$	
BE	$P(Y \mid X)P(X) \rightarrow max$	*
ML	$P(Y \mid X) \rightarrow max$	**
LS	$V^T\Sigma^{-1}V \rightarrow min$	***
* $P(Y)$ =constant ** $P(X)$ =constant and * *** $V \sim N(0,\Sigma)$ and **		

Table 2: The MAP criterion and its progenies

where C is constant. It is to see that the least squares criterion is to minimize $V^T\Sigma^{-1}V$ which is equivalent to using a maximum likelihood estimation to maximize $P(Y \mid X)$.

So far, we have discussed the MAP criterion and its progenies. Obviously, each criterion has its own supposition (cf. Tab. 4). The LS criterion which is so widely used in data processing as a general framework for problem solving is, for instance, only suitable for dealing with over-constrained inverse problems. For under-constrained inverse problems the MAP criterion is more appropriate as it provides a flexible framework to integrate *a priori* knowledge to restrict the solution space and one can take the probability behavior of both the data and the desired solutions into account. There are many problems in computer vision, especially in the low-level image processing, including edge detection, spatial-temporal approximation, image segmentation, image registration, and surface reconstruction (cf. Poggio et al.,1985), which are unfortunately of under-constrained nature and whose solutions demands on new inference techniques beyond the LS estimation.

6 Restricting Solution Space

The MAP criterion provides a general approach to handle the inverse problem in an uncertain environment. It gives a mechanisms to restrict the solution space and to integrate *a priori* knowledge by specifying the appropriate prior probabilities $P(X)$. However, the MAP criterion doesn't tell how to construct $P(X)$. In this section we look at this issue.

The parameter set $\mathcal{X}$, as mentioned earlier, represents a physical system and can be considered as a parameter space. In principle, every point $X \in \mathcal{X}$ represents a possible solution. It can be easily imagined that not all points in the solution space are meaningful. Our job is to explore the solution space to find an appropriate point (solution). So, the first problem is how to measure the appropriateness of a solution and how to describe the solution space. A general way to do this is to define a probability distribution of the solution space $P(X)$ (Tarantola, 1987).

Let X be a set of parameters representing the state of a Markov random field. According to the Hammersley-Clifford theorem (cf. Geman and Geman, 1984; Chou and Brown, 1988), this random field can be described by a probability distribution of the Gibbs form:

$$P(X) = \frac{1}{C} \exp\left[-\frac{1}{T}\mathcal{E}(X)\right], \; X \in \mathcal{X}, \qquad (5)$$

where C is a normalizing constant, T is the so called *temperature* of the field that controls the flatness of the distribution of the configurations X and $\mathcal{E}(X)$ is the energy of X which consists of the sum of the local potential

$$\mathcal{E}(X) = \sum_{x \in X} V_e(x). \qquad (6)$$

The relation (5) suggests that the point in $\mathcal{X}$ with a higher energy occurs less likely.

Now, let us look at the ill-posed inverse problem (1). According to the the least squares criterion (cf. Tab. 4), one can get an unique pseudo solution through minimizing $V^T\Sigma^{-1}V$, with respect to

$$V = A\,X - Y. \qquad (7)$$

This leads to solving the normal equation

$$(A^T\Sigma^{-1}A)\,X = A^T\Sigma^{-1}Y. \qquad (8)$$

Certainly, the normal matrix $N = A^T\Sigma^{-1}A$ is regular only if the problem (1) is overdetermined. This suggests that the least squares criterion can only be used to deal with overdetermined ill-posed inverse problems. For underdetermined ill-posed inverse problems, which emerge so often in image understanding, the least squares criterion can not help us to find a satisfying solution, as it does not have a mechanism to restrict the solution space.

Using the bayesian estimate method (BE) (cf. Tab. 4), we have the following optimizing problem (cf. (3), (4), and (5)):

$$V^T\Sigma^{-1}V + \frac{2}{T}\,\mathcal{E}(X) \rightarrow min, \qquad (9)$$

with respect to (7). Intuitively, this criterion, in comparison with the least squares criterion, is more powerful to deal with underdetermined ill-posed inverse problems, as it gives not only a measure for the quality of the fitting , through the first term in (9), but also a measure for the probability of the solution, through the second term in (9). So we can integrate our *a priori* knowledge into the inverse process by designing

the second term in (9) appropriately. Of course, designing $\mathcal{E}(X)$ is a skill. One needs knowledge about the physical meaning of the solution and the internal coherence of unknown parameters.

Generally, the second term in (9) can be designed to have the form

$$\frac{2}{T}\,\mathcal{E}(X) = E^T\,\Sigma_e^{-1}\,E, \quad E = \Phi\,X - \Psi, \qquad (10)$$

where Φ is an operator, Ψ is a vector, and Σ_e is a matrix. They have to be determined using our *a priori* knowledge. If we, for instance, *a priori* know that the elements $x_i \in X$, $i = 1, ..., m$, should have values around a_i, $i = 1, ..., m$, then we can construct

$$\Sigma_e = \frac{T}{2}\begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_m^2 \end{pmatrix},$$

$$\Phi = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}, \Psi = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix}, \qquad (11)$$

where σ_i, $i = 1, ..., m$, denote the degree of the certainty of our *a priori* knowledge.

Let us solve the ill-posed inverse problem (1) again, but using the new criterion (9) which is equivalent to

$$V^T\Sigma^{-1}V + E^T\Sigma_e^{-1}E \rightarrow min. \qquad (12)$$

This lead to the new normal equation

$$(A^T\Sigma^{-1}A + \Phi^T\Sigma_e^{-1}\Phi)\,X = A^T\Sigma^{-1}Y + \Phi^T\Sigma_e^{-1}\Psi. \qquad (13)$$

It is sure that the new normal matrix $N = A^T\Sigma^{-1}A + \Phi^T\Sigma_e^{-1}\Phi$ is no more singular even for underdetermined ill-posed inverse problems, if Φ, Σ_e and Φ are all appropriately constructed.

7 Surface Reconstruction

There are, as indicated above, many problems in computer vision which can be generally formulated as inverse problems. We have proposed approaches which provide a sound theoretical basis but offer few practical computational methods for dealing with concrete tasks in computer vision. So, in this section, we go further into the application of the inverse problem theory to an elementary problem, i.e. the computing of the representation of visible surfaces from multiple images.

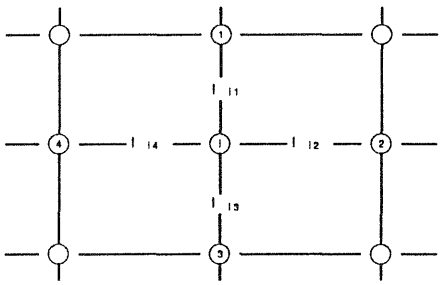

Figure 1: The meaning of the label l_{ij}

7.1 Representation of Visible Surfaces

The role of a representation is to make certain information explicit at an appropriate point in the problem analysis as the abstract information must be expressed by concrete descriptions. Thus, the choice or design of a representation affects the success of analysis. The representation of object surfaces deals with strategies and techniques for describing their geometrical and physical properties in a way appropriate for numerical processing.

Let $\mathcal{S}$ be a set of parameters which describe the geometrical and physical properties of an object surface. An element $S \in \mathcal{S}$ can be a concrete measure, e.g. elevation (depth), deformation, reflectivity, etc.. Each element $S \in \mathcal{S}$ can be mapped onto XY-plane in a 3D coordinate system and represented mathematically as $S = S(X, Y), S \in \mathcal{S}$. For computational reasons, we rather represent S by a grid of square 1×1 elements, where each element is centered at the coordinates (X_i, Y_i) of the i^{th} element. Then, the object surface is described by $m \times n$ elements: $S_i = S(X_i, Y_i), i \in \mathcal{I} = [1, ..., I]$, where $\mathcal{I}$ can be thought of as a vector belonging to the set $(1, ..., m) \times (1, ..., n)$ which has totally $m \times n$ elements.

Sometimes we may be also interested in the spatial coherence (continuity) of S. So we introduce a label set L whose element l_{ij} represents the strength of the spatial coherence between two neighbor S_i and S_j (cf. Fig. 1). The label l_{ij} can be binary: $l_{ij} = 1$ for continuity between S_i and S_j, $l_{ij} = 0$ for discontinuity between S_i and S_j. l_{ij} can also take the value between 0 and 1, i.e. $l_{ij} \in [0, 1]$, for continuously describing the coherence strength.

7.2 Forward Modeling

The purpose of forward modeling is to find constraints linking elements in $\mathcal{S}$ with observations, i.e. image densities (intensities), based on physical properties of imaging. The relationship between the image density

$D(x,y)$ of a photographic image and the exposure H $[lux \cdot sec]$ is

$$D(x,y) = \gamma \log H + D_0 \qquad (14)$$

for normal exposure, where x, y are image coordinates of a pixel, γ (≈ 1) is the gradation and D_0 is a constant. Usually, γ depends on the developer, the development time and temperature, and the photographic material. The exposure H depends, first of all, on the reflection properties of surfaces. Many natural terrain features roughly approximate diffuse reflectors. A Lambertian surface is a perfect diffuse reflector with the property that the radiance L $[cd \cdot m^{-2}]$ is constant for any incident angle.

The relation (14) is very important as it explains the physical meaning of image intensities. From (14) one can derive the following radiometric constraint (cf. Zheng, 1990):

$$D(x,y)/\gamma + 2\cdot\log(c^2+x^2+y^2)+\eta+\psi(x,y) = 0, \quad (15)$$

where η can be considered as a constant for all pixels in the same image; but ψ is a local parameter which changes from pixel to pixel. The physical meaning of ψ in (15) is the logarithm of the luminance intercepted by the lens for a pixel.

The image coordinates x and y in (15) are functions of the object coordinates of the surface element, according to the well known projection equation:

$$\begin{pmatrix} x \\ y \\ -c \end{pmatrix} = \frac{1}{m} R^T \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \begin{pmatrix} X_0 \\ Y_0 \\ Z_0 \end{pmatrix}, \qquad (16)$$

where m is a scale factor, R is the rotation matrix containing three rotation angles (ϕ, ω, κ), (X, Y, Z) are the corresponding ground coordinates of the image point (x, y), and $\Omega = (\phi, \omega, \kappa, X_0, Y_0, Z_0)$ are camera orientation parameters. Besides, $D(x,y)$ has to be re-sampled from the neighboring digitized pixels by using, for instance, a bilinear interpolation

$$D = G^T L, \qquad (17)$$

where L and G denote a set of intensities of neighboring pixels and a corresponding coefficient vector, respectively. Thus, for the ground surface point (X, Y, Z), the left side of (15) is a function of many parameters:

$$f(X, Y, Z, \Omega, L, \gamma, \eta, \psi) = 0. \qquad (18)$$

Now, let us look at the problem of surface reconstruction from multiple images. For the purpose of simplicity, we discuss here only the solution of recovering the surface profile, which is represented using K discrete profile points, from J images, $M_j, j \in \mathcal{J} = [1, ..., J]$,

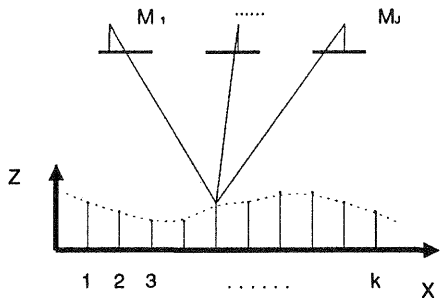

Figure 2: Surface reconstruction from image data

which are taken from different views and depict the same surface (cf. Fig. 2). Besides, we also suppose that the orientation parameters of these images are all known. For the i^{th} profile point, $i \in \mathcal{I} = [1, ..., K]$, we can write J constraints like (18). For K profile points in $\mathcal{I}$ we can write totally $J \times K$ constraints like (18). Supposing a Lambertian surface and $\gamma \simeq 1$, these constraints can be further simplified (cf. Zheng, 1990):

$$D(x_1, y_1) + p_j D(x_j, y_j) + 2 \log \left(\frac{c_1^2 + x_1^2 + y_1^2}{c_j^2 + x_j^2 + y_j^2} \right) + q_j = 0, \qquad (19)$$

where $p_j = -\gamma_1/\gamma_j$, $q_j = \eta_1 - \eta_j$, $j \in [2, ..., J]$, and we have a set of new parameters p_j and q_j, $j \in [2, ..., J]$, which describe approximately the radiometric relationship of the image M_1 with the other images M_j, $j \in [2, ..., J]$. Our *a priori* knowledge about p_j and q_j is that p_j should be around 1 and q_j should be around 0. Linearization of these now constraints gives

$$U\,V + A\,\Delta Z + B\,\Delta Q + W(L, Z_0, Q_0) = 0 \qquad (20)$$

where L and V denote two vectors containing observations, i.e. intensities of image pixels, and their residuals; Z_0 and ΔZ denote two vectors containing approximate elevations of profile points in $\mathcal{I}$ and their corrections; Q_0 and ΔQ denote two vectors containing approximate values of p_j and q_j, $j \in [2, ..., J]$ and their corrections; and U, A, B, and W are corresponding coefficient matrices and the vector of constants, respectively. It is clear that (20) is strongly underdetermined as $V, \Delta Z$ and ΔQ are all unknown. The total number of unknowns is much larger than that of the constraints, and one could generally hypothesize an infinite number of different solutions that would meet (20). So, we have to use criteria to restrict the space of acceptable solutions and to find a unique solution which will be a *best* one to interpret the image data.

7.3 Inversion

According to Table 4, the MAP criterion would choose $Z = Z_0 + \Delta Z$ and $Q = Q_0 + \Delta Q$ in (20) such that the conditional probability $P(Z, Q \mid L)$ is maximized, which is equivalent to maximize $P(L \mid Z, Q)P(Z, Q)$, if $P(L)$ is constant.

As mentioned above, the conditional probability $P(L \mid Z, Q)$ can be simply assumed as the probability that the observational residuals were produced by a normally distributed random variable (cf. (4)). The problem is, now, how to exploit our priori knowledge about Z and Q to constitute their probabilities, i.e. $P(Z, Q)$. If Z and Q are statistically independent, we have $P(Z, Q) = P(Z) \cdot P(Q)$.

To construct $P(Z)$ and $P(Q)$, one has to know the meaning of Z and Q. The vector Z, for instance, represents some elevations of discrete surface profile points. So, our priori assumption about Z is the spatial coherence of its elements. This suggests that the local potential of the element $Z_i \in Z$, $i \in \mathcal{I}$ can be written as

$$V_e(Z_i) = \sum_{\substack{i \neq j \\ j \in N_i}} \left[l_{ij} \frac{Z_i - Z_j}{\sigma_Z} \right]^2 \qquad (21)$$

where N_i denotes the set of totally connected subgraphs (cliques) with respect to the element Z_i, $l_{ij} \in [0, 1]$ is the connection strength between Z_i and Z_j, and σ_Z is a normalizing constant. According to the Hammersley-Clifford theorem, the energy of Z can be computed with

$$\mathcal{E}(Z) = \sum_{Z_i \in Z} \sum_{\substack{i \neq j \\ j \in N_i}} \left[l_{ij} \frac{Z_i - Z_j}{\sigma_Z} \right]^2 = E_Z^T \, \Sigma_Z^{-1} \, E_Z \qquad (22)$$

and

$$E_Z = \Phi_Z \, Z - \Psi_Z, \qquad (23)$$

where

$$\Phi_Z = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 \\ 0 & 1 & -1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & -1 \end{pmatrix}, \quad Z = \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_K \end{pmatrix},$$

$$\Psi_Z = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \Sigma_Z = \begin{pmatrix} \frac{\sigma_Z^2}{l_{12}} & 0 & \dots & 0 \\ 0 & \frac{\sigma_Z^2}{l_{23}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\sigma_Z^2}{l_{(K-1)K}} \end{pmatrix}. \qquad (24)$$

Similarly, since we priori know that p should be around 1 and q should be around 0, so the energy of Q is

$$\mathcal{E}(Q) = E_Q^T \, \Sigma_Q^{-1} \, E_Q, \qquad (25)$$

with

$$E_Q = \Phi_Q \, Q - \Psi_Q, \qquad (26)$$

where

$$\Phi_Q = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}, \quad Q = \begin{pmatrix} p_2 \\ q_2 \\ \vdots \\ p_J \\ q_J \end{pmatrix},$$

$$\Psi_Q = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 1 \\ 0 \end{pmatrix}, \quad \Sigma_Q = \begin{pmatrix} \sigma_p^2 & 0 & \dots & 0 & 0 \\ 0 & \sigma_q^2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma_p^2 & 0 \\ 0 & 0 & \dots & 0 & \sigma_q^2 \end{pmatrix} \qquad (27)$$

and σ_p and σ_q are constants encoding the reliability of our *a priori* knowledge about p and q.

Considering (3), (4), (5), (22) and (25), the MAP based surface reconstruction is to solve the optimizing problem

$$\frac{1}{2} V^T \Sigma^{-1} V + \frac{1}{T} E_Z^T \, \Sigma_Z^{-1} \, E_Z + \frac{1}{T} E_Q^T \, \Sigma_Q^{-1} \, E_Q \rightarrow min, \qquad (28)$$

with subject to (20), (23) and (26). Surely, the surface Z which is inferred in this way is dependent on $\Gamma = (\Sigma, \Phi_Z, \Psi_Z, \Sigma_Z, \Phi_Q, \Psi_Q, \Sigma_Q)$ and they should be determined using *a priori* knowledge, before the inversion process takes place. The quality of Z is sometimes not satisfying if our knowledge is not good enough to ensure an appropriate determination of Γ. So, a very interesting question is how to enlarge our knowledge and how to adapt Γ during the inversion in order to improve the quality of Z. Information processing systems that improve their performance or enlarge their knowledge bases are said to " learn". This ability would clearly have value in digital image inversion. Using parameter estimation technique, we can, for instance, adapt Σ, Σ_Z, and Σ_Q in (28) iteratively during the inversion so that the result is robust against image noises with different properties, against surface discontinuities, and against different reliabilities of our *a priori* knowledge (cf. Zheng and Hahn, 1990).

8 Examples

To demonstrate the feasibility of the methods for the purpose of surface reconstruction an algorithm has been developed based on the MAP criterion (cf. Zheng, 1990). It was tested on a variety of data sets including synthetic and real image data.

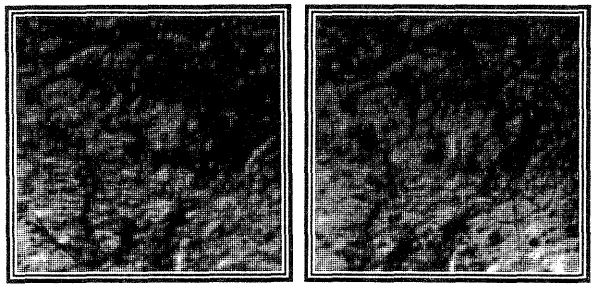

Figure 3: A stereo pair of digital images

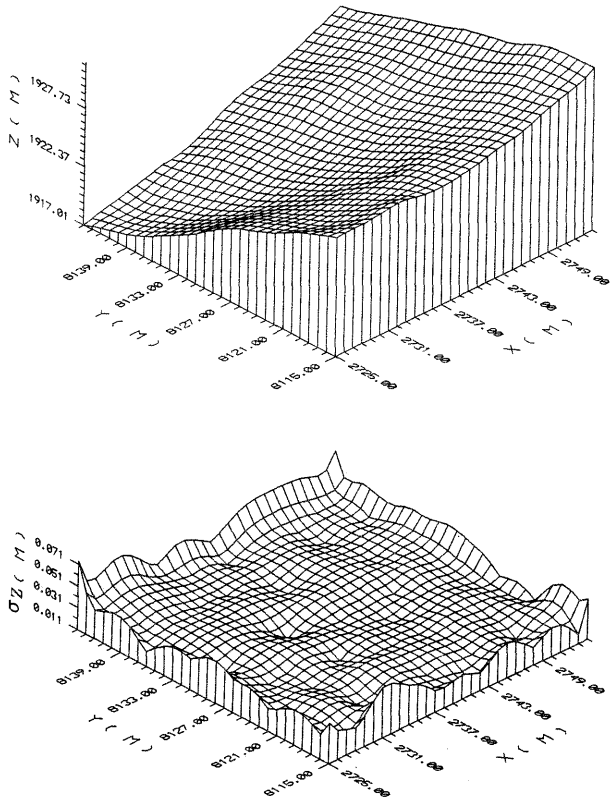

Figure 4: The computed surface and its a posteriori accuracy

Let us firat look at a stereo pair of digital aerial images shown in Figure 3. They represent a piece of steep and rough wilderness with rock-debris. Each of them has 240 × 240 pixels. The image scale is about 1 : 10000. This image material was also used to test the feature based and least squares matching algorithms and is regarded as the hardest one within three selected projects (cf. Hahn/Förstner, 1988).

Figure 4 shows the automaticly generated surface field and its posteriori accuracy. It contains 30 × 30 lattice points with 1 × 1 m^2 lattice size. All surface heights of the same lattice points (900 points) was also manually measured on an analytical measuring device Planicomp C 100 as reference (cf. Fig. 5). The precision of the manual measurements is about

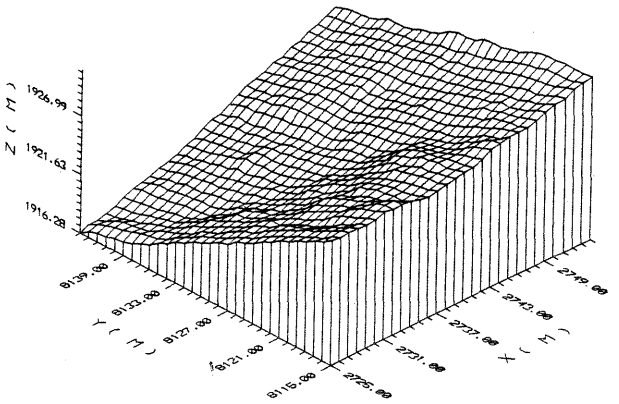

Figure 5: The same surface measured manually

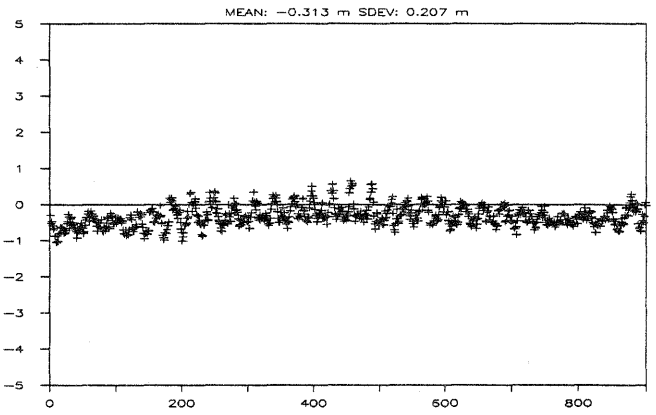

Figure 6: The difference between two surfaces

0.22 m ($\approx 0.14\ ^o/_{oo}$ of flying height). Figure 6 illustrates the difference between the automatically and the manually generated surface fields. This difference can be characterized by its mean (bias) and its standard deviation against the bias. Taking the a posteriori accuracy of the automatically generated surface (cf. Fig. 4) into account, the results are:

MEAN DIFF: −0.313 m (bias),

SDEV: 0.207 m ($\approx 0.13\ ^o/_{oo}$ of flying height),

where the precision of the surface reconstruction using our algorithm is about the same as observed by the operator.

Finally, we look at the image pair "House" (cf. Figure 7), which is one of the twelve image pairs for the test on image matching of the working group III/4 of International Society for Photogrammetry and Remote Sensing (cf. Gülch, 1988). This image pair has been classified by the test organizer into the group of high complexity for image matching, as it contains almost all troubles, including discontinuities, occlusions, shadows, and corruptions. Each image has a

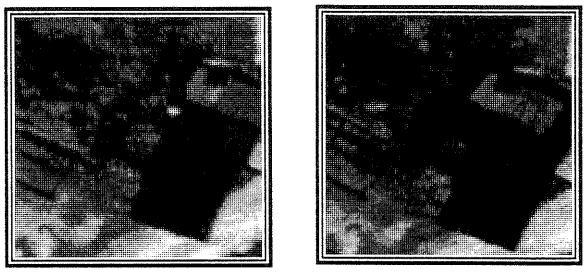

Figure 7: The image pair "House"

size of 240 × 240 pixels and the image scale is about 1 : 3000. In Figure 8, we show the computed surface field, by means of a perspective view and a contour map.

9 Conclusion

In this work, we concentrate our attention on inference processes in computer vision and formulate many of its goals in a general manner as a ill-posed problem of image inversion. Based on MAP criteria, we have introduced a theoretical framework for dealing with ill-posed inverse problems. We have shown how the surface reconstruction from images can be solved under this theoretical basis as an application. Of course, this is only a limited domain of its applications. So, among the goals of future work will be 1) the introduction of a learning mechanism to improve and adapt the *a priori* knowledge during the inverse process, 2) the application of neural network technology to developing parallel algorithms for solving optimizing problem mentioned above, and 3) the extension of the application range of the approach.

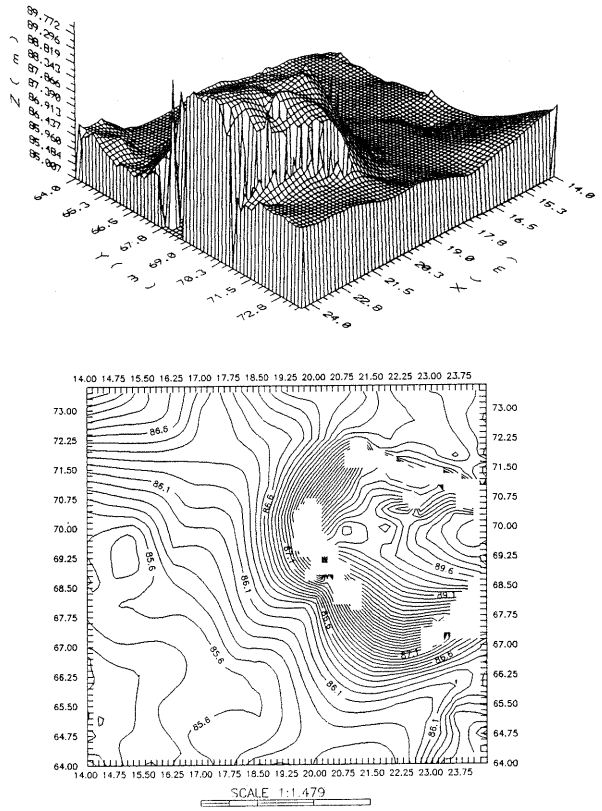

Figure 8: The computed surface

References

[1] Chou, P. B. and C. M. Brown,1988. Multimodal reconstruction and segmentation with Markov random fields and HCF optimization, *Proc. Image Understanding Workshop*, San Mateo, CA: Morgan Kaufmann, pp. 214-221.

[2] Fawcett, J. A.,1985. Inversion of N-dimensional spherical averages, *SIAM J. Appl. Math.* 45, pp. 336-341.

[3] Geman, S. and D. Geman, 1984. Stochastic Relaxation, Gibbs Distribution and Bayesian Restoration of images, *IEEE PAMI*, Vol. 6, pp. 721-741.

[4] Gülch, E., 1988. Results of Test on Image Matching of ISPRS WG III/4, *Intern. Archives of Photogrammetry and Remote Sensing*, vol. 27, Kyoto.

[5] Hahn, M. and W. Förstner, 1988. On the Applicability of a Feature Based and a Least Squares Matching Algorithm for DEM Acquisition, *Intern. Archives of Photogrammetry and Remote Sensing*, 27-3, Kyoto.

[6] Poggio, T., V. Torre, and C. Koch, 1985. Computational Vision And Regularization Theory, *Nature*, vol. 317, pp. 314-319.

[7] Tarantola, A.,1987. *Inverse Problem Theory*, Elsevier, New York.

[8] Zheng, Y.-J., 1990. Inverse und schlecht gestellte Probleme in der photogrammetrischen Objekt Rekonstruktion, *Inaugural-Dissertation*, Fakultät für Vermessungswesen, Universität Stuttgart, Stuttgart.

[9] Zheng, Y.-J. and M. Hahn, 1990. Surface Reconstruction from Images in the Presence of Discontinuities, Occlusions and Deformations, *Intern. Archives of Photogrammetry and Remote Sensing*, vol. 28, part 3/2, pp. 1121-1144, Wuhan.

[10] Zheng, Y.-J., 1992. Inductive Inference and Inverse Problems in Computer Vision, *Robust Computer Vision*, Förstner/Ruwiedel (Eds.), Karlsruhe: Wichmann, pp. 320-350.