Bayesian reconstruction based on flexible prior models

K. M. Hanson

Los Alamos National Laboratory, MS P940, Los Alamos, New Mexico 87545

Received July 16, 1992; accepted September 29, 1992; revised manuscript received December 14, 1992

A new approach to Bayesian reconstruction is proposed that endows the prior probability distribution with an inherent geometrical flexibility, which is achieved through a transformation of the coordinate system of the prior distribution or model into that of the reconstruction. With this warping, prior morphological information regarding the object that is being reconstructed may be adapted to various degrees to match the available measurements. The extent of warping is readily controlled through the prior probability distributions that are specified for the warp parameters. The complete reconstruction consists of a warped version of the prior model plus an estimated deviation from the warped model. Examples of tomographic reconstructions demonstrate the power of this approach.

1. INTRODUCTION

Often the geometrical structure or morphology of an object to be reconstructed is known before data are taken. When few measurements are available, the reconstruction is usually underdetermined, which means that many solutions are possible. Bayesian methods of reconstruction can help to identify the best solution by taking into account characteristics of the object being imaged that are known *a priori*. Knowledge of the object's characteristics are incorporated in terms of prior probability distributions on the appropriate physical parameters. These methods have been shown to improve substantially the accuracy of reconstructions obtained from limited data when good geometrical information is employed in the prior model.[1,2] However, if the actual object under study differs even only slightly in size, shape, or position from the assumed geometrical model, use of this kind of prior can lead to poor reconstructions.

The above difficulties arise because the prior model is typically considered to be fixed relative to the spatial coordinate system of the reconstruction.[3] A superior approach is proposed in which the prior model for the object that is being reconstructed is allowed to alter its geometrical characteristics to accommodate the data by warping the coordinate system of the prior model onto the coordinate system of the reconstruction.[4] Changes in size, position, and orientation of the model are accommodated by a linear transformation between the two coordinate systems. Changes in shape are possible with nonlinear transformations. Whereas it may seem counterproductive to add even more parameters to what is already a difficult reconstruction task, such parameters permit the reconstruction procedure to adapt the shape of the prior model to conform to the measurements. Within the Bayesian framework, the parameters that are needed to specify the coordinate transformation are determined as part of the overall estimation–reconstruction problem of finding the maximum of the full posterior probability distribution. One can readily control the degree and type of warping through a judicious choice of the prior probability distribution on the transformation parameters. A warping formulation with many degrees of freedom and wide prior distributions on the parameters results in a flimsy prior model. One can choose strong constraints on the parameters, to maintain closely the initial shape, or weak constraints, to allow the shape to become fairly contorted.

The power of this new approach to prior models is illustrated through a few examples of tomographic reconstruction. For simplicity, the coordinate transformations are restricted to low-order polynomials in these examples.

2. BAYESIAN APPROACH

Fundamental to the Bayesian approach is the posterior probability, which is supposed to summarize the full state of knowledge concerning a given situation. Given the data $\mathbf{g}$, the posterior probability of any image $\mathbf{f}$ is provided by Bayes's law in terms of the proportionality

$$p(\mathbf{f}|\mathbf{g}) \propto p(\mathbf{g}|\mathbf{f})p(\mathbf{f}), \tag{1}$$

where $p(\mathbf{g}|\mathbf{f})$, the probability of the observed data $\mathbf{g}$ given $\mathbf{f}$, is called the likelihood and $p(\mathbf{f})$ is the prior probability of $\mathbf{f}$. The likelihood is specified by the assumed probability distribution of the fluctuations in the measurements regarding their predicted values (in the absence of noise). The prior probability density $p(\mathbf{f})$, or the prior, expresses the prior information in terms of the expected relative frequency of occurrences of all possible images. Any known constraints concerning impossible images ought to be included explicitly or implicitly in $p(\mathbf{f})$. The complete result of the Bayesian approach is constituted in the full $p(\mathbf{f}|\mathbf{g})$. However, we humans have difficulty visualizing such a multidimensional probability distribution. Thus it is normally desired that a single image be presented as the reconstruction. An appropriate choice is the image that maximizes the *a posteriori* probability, called the maximum *a posteriori* (MAP) estimate.[5,6]

Essential to the Bayesian approach is the use of prior knowledge to help guide the result in the desirable direction. Without the prior, the MAP solution would collapse to nothing more than the maximum-likelihood estimate. However, an infinite number of solutions exist when the data are quite limited, which is precisely the situation I aim to address. Thus the prior is indispensable for weed-

0740-3232/93/050997-08$05.00

ing out unlikely solutions. The prior probability distribution may be thought of as a statement of the type of result that is preferred. The problem with the typical Bayesian approach is that the model for the prior is usually considered to be geometrically fixed.[3] This restriction might seem curious because the approach is grounded in probability theory and thus ought to allow for a continuum of possibilities ranked on the basis of their relative likelihood. The possibility of a change in position or shape of the prior model should be an integral part of the Bayesian approach. The proposed extension to the standard MAP technique overcomes its rigid definition of the prior and provides the desired flexibility in geometry as well as amplitude.

A. Standard MAP Formulation

I begin by briefly summarizing the standard MAP approach. For a fuller treatment, see Refs. 2 and 3. Let us assume that the amplitudes of the N pixels of an image are represented by a vector $\mathbf{f}$ of length N. We are given M discrete measurements that are linearly related to the amplitudes of the original image. Assume that these measurements are degraded by additive noise with a known covariance matrix $\mathbf{R_n}$, which describes the correlations between noise fluctuations. The measurements can then be represented by a vector of length M:

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \mathbf{n}, \tag{2}$$

where $\mathbf{n}$ is the random noise vector and $\mathbf{H}$ is the measurement matrix. In computed tomography the elements of the jth row of $\mathbf{H}$ describe the weight of the contribution of each image pixel to the jth projection measurement.

Because the probabilities are a function of continuous parameters, namely, the N pixel values of the image, they are actually probability densities, designated by a small $p(\,)$. From Bayes's law [Eq. (1)] the negative logarithm of the posterior probability density is given by

$$-\log[p(\mathbf{f}|\mathbf{g})] = \varphi(\mathbf{f}) = \Lambda(\mathbf{f}) + \Pi(\mathbf{f}), \tag{3}$$

where the first term comes from the likelihood and the second term comes from the prior probability. Assuming additive Gaussian noise, the negative log(likelihood) is just half of chi squared:

$$\begin{aligned} -\log[p(\mathbf{g}|\mathbf{f})] &= \Lambda(\mathbf{f}) = (1/2)(\chi^2) \\ &= (1/2)\mathbf{g} - \mathbf{H}\mathbf{f})^T\mathbf{R_n}^{-1}(\mathbf{g} - \mathbf{H}\mathbf{f}), \end{aligned} \tag{4}$$

which is quadratic in the residuals. Of course the choice for the likelihood function should be based on the actual statistical characteristics of the measurement noise, which we can assume are known *a priori*.

The second term $\Pi(\mathbf{f})$ represents the prior probability distribution, which should incorporate as much as possible the known characteristics of the original image. For simplicity, one often chooses a Gaussian distribution for the prior probability, whose negative logarithm may be written as

$$-\log[p(\mathbf{f})] = \Pi(\mathbf{f}) = (1/2)(\mathbf{f} - \bar{\mathbf{f}})^T\mathbf{R_f}^{-1}(\mathbf{f} - \bar{\mathbf{f}}), \tag{5}$$

where $\bar{\mathbf{f}}$ is the mean and $\mathbf{R_f}$ is the covariance matrix of the prior probability distribution. Equation (5) can provide beneficial information in at least two ways. First, $\bar{\mathbf{f}}$ can provide geometric or morphological information regarding the object being reconstructed by the way in which it varies with position. Second, the matrix $\mathbf{R_f}$ describes the extent of certainty regarding how close $\mathbf{f}$ is likely to be to $\bar{\mathbf{f}}$. The size of the elements of $\mathbf{R_f}$ determines the strength of the constraint that the prior places on the solution. Larger values indicate greater uncertainty and a weaker constraint.

In the absence of any subsidiary constraints on $\mathbf{f}$, at the minimum of $\varphi(\mathbf{f})$, that is, at the MAP solution,

$$\nabla_{\mathbf{f}}\varphi(\mathbf{f}) = \nabla_{\mathbf{f}}\Lambda(\mathbf{f}) + \nabla_{\mathbf{f}}\Pi(\mathbf{f}) = 0. \tag{6}$$

Under the assumption of Gaussian probability distributions, $\Lambda(\mathbf{f})$ and $\Pi(\mathbf{f})$ are quadratic in $\mathbf{f}$; hence the resulting MAP equations are linear in $\mathbf{f}$. The solution can therefore be obtained by using standard matrix methods, which explains the popularity of the Gaussian assumption.

Consider an imaging situation in which all possible objects have the same shape and a known constant amplitude. In such a case it would be expected that the elements of $\mathbf{R_f}$ would be small. However, if the objects cover a wide range of sizes, the elements of $\mathbf{R_f}$ must be correspondingly increased in the standard approach to express the lack of knowledge as to size. By using an approach that provides one with the ability to vary size and shape flexibly, the elements of $\mathbf{R_f}$ can be kept small, thereby properly expressing one's state of knowledge concerning the amplitude variation for an object of any particular size. The uncertainty in size can then be incorporated properly in the specification of the probability distribution that is associated with geometry rather than amplitude.

B. MAP Based on Flexible Prior

The difficulty with the standard MAP approach described in Subsection 2.A is that $\bar{\mathbf{f}}$ and $\mathbf{R_f}$ are spatially fixed. In any particular situation the actual object may differ in location, size, or shape. Any of these inconsistencies can destroy the usefulness of the prior.[2,7]

To build geometrical flexibility into the prior, for example, we may consider $\bar{\mathbf{f}}$ to be a function of several parameters represented by the vector $\mathbf{a}$, that is, $\bar{\mathbf{f}}(\mathbf{a})$. The parameters in vector $\mathbf{a}$ control the position, size, and shape of the prior distribution in a manner yet to be specified. Writing the reconstruction as

$$\mathbf{f} = \mathbf{d} + \bar{\mathbf{f}}, \tag{7}$$

where $\mathbf{d}$ is the deviation of the reconstruction from $\bar{\mathbf{f}}$, we stress the equal importance of $\mathbf{d}$ and $\mathbf{a}$ by taking them to be the independent variables in the reconstruction problem. Bayes's law [Eq. (1)] must now be restated in terms of $\mathbf{d}$ and $\mathbf{a}$ rather than $\mathbf{f}$. After $\mathbf{d}$ and $\mathbf{a}$ are determined, the full reconstruction is given by Eq. (7). The negative log(likelihood) is expressed as

$$\begin{aligned} -\log[p(\mathbf{g}|\mathbf{d},\mathbf{a})] &= \Lambda(\mathbf{d},\mathbf{a}) \\ &= (1/2)[g - \mathbf{H}(\mathbf{d} + \bar{\mathbf{f}})]^T\mathbf{R_n}^{-1} \\ &\quad \times [\mathbf{g} - \mathbf{H}(\mathbf{d} + \bar{\mathbf{f}})]. \end{aligned} \tag{8}$$

In Bayesian tradition, we must supply a prior probability distribution for all variables. As above, we might use a Gaussian distribution for the prior probability for the parameters **a**, which augments the prior on the pixel values. Then the negative logarithm of the full prior probability may be written as

$$\Pi(\mathbf{d},\mathbf{a}) = (1/2)\mathbf{d}^T\mathbf{R_d}^{-1}\mathbf{d} + (1/2)(\mathbf{a}-\bar{\mathbf{a}})^T\mathbf{R_a}^{-1}(\mathbf{a}-\bar{\mathbf{a}}), \quad (9)$$

where **a** and **d** are assumed to be statistically independent *a priori*. The choice of a Gaussian distribution for the prior on **a** is arbitrary at this point, although its reasonableness will be shown below. The first term is similar to Eq. (5), with $\mathbf{R_d}$ taking the place of $\mathbf{R_f}$. The covariance matrix $\mathbf{R_d}$ for the deviations from the model $\bar{\mathbf{f}}(\mathbf{a})$ is naturally a function of **a** because it presumably follows the model. The optimization function φ is the sum of Eqs. (8) and (9). The choice of the relative weight of the likelihood [Eq. (8)] and the prior [Eq. (9)] is critical because it affects how forcefully the information contained in the data is transferred to the observer of the image.[8]

I note in passing that, although one is primarily concerned with morphological flexibility here, the parameter vector **a** may include other types of flexibility. For example, one element of **a** might be the magnitude of $\bar{\mathbf{f}}$ inside the object, which may not be well known beforehand. This formalism can clearly provide the prior model with many new types of degrees of freedom.

C. Reconstruction Procedure

In reconstruction we seek to estimate all pixel values in the original scene. It is necessary to estimate **d** and $\bar{\mathbf{f}}$ and therefore **a**. The self-consistent Bayesian solution that maximizes the *a posteriori* probability must satisfy

$$\nabla_{\mathbf{d}}\varphi = 0 \quad \text{and} \quad \nabla_{\mathbf{a}}\varphi = 0, \quad (10)$$

provided that the solution is not otherwise constrained. The gradient of the likelihood with respect to **d** and $\bar{\mathbf{f}}$ is

$$\nabla_{\mathbf{d}}\Lambda = \nabla_{\bar{\mathbf{f}}}\Lambda = \mathbf{H}^T\mathbf{R_n}^{-1}[\mathbf{g} - \mathbf{H}(\mathbf{d}+\bar{\mathbf{f}})], \quad (11)$$

from which we obtain, for the gradient of φ with respect to **d**,

$$\nabla_{\mathbf{d}}\varphi = \mathbf{R_d}^{-1}\mathbf{d} + \mathbf{H}^T\mathbf{R_n}^{-1}[\mathbf{g} - \mathbf{H}(\mathbf{d}+\bar{\mathbf{f}})]. \quad (12)$$

In computed tomography the matrix $\mathbf{H}^T$ corresponds to the well-known backprojection operation.

The gradient of φ with respect to parameter a_j is

$$[\nabla_{\mathbf{a}}\varphi]_j = \frac{\partial\varphi}{\partial a_j} = [\mathbf{R_a}^{-1}(\mathbf{a}-\bar{\mathbf{a}})]_j + \sum_i \frac{\partial\Lambda}{\partial\bar{f}_i}\frac{\partial\bar{f}_i}{\partial a_j}, \quad (13)$$

where the sum is over the pixels of the reconstruction. The first term comes from the prior [Eq. (9)], and the second comes from the likelihood [Eq. (8)]. The first quantity inside the sum is given by Eq. (11), and the second is given by the functional dependence of $\bar{\mathbf{f}}$ on a_j.

The MAP solution characterized by Eqs. (10) can be found by the method of steepest descent with the use of Eqs. (12) and (13) for the gradients. Although this method is computationally inefficient, it demonstrates the usefulness of the flexible prior method. The inclusion of parameters on which the solution may depend in a nonlinear fashion can lead to a situation in which the solution is not unique. This result seems to be a necessary consequence of more complex, and realistic, modeling of the solution. We suppose that it is feasible to make an intelligent choice for the initial guess, one whose basin of convergence includes the most appropriate solution.

3. WARP

One particularly advantageous way to introduce geometrical flexibility into a fixed prior probability distribution on the image amplitude is to transform the coordinate system of the reconstruction into that of the prior probability distribution. The benefit of this method is that it can be applied to any prior distribution. The prior distribution on the image **f** need not be given in parameterized form to generate the geometrical distortion.

Assume that the prior is initially specified as a function of the spatial coordinates (x', y'). A warping of that prior is achieved by transforming the (x, y) coordinates of the reconstruction pixels to the original (x', y') coordinates:

$$x' = x + u(x,y), \qquad y' = y + v(x,y). \quad (14)$$

Then the ensemble mean in the reconstruction domain $\bar{\mathbf{f}}(x,y)$ is actually given by $\bar{\mathbf{f}}[x'(x,y), y'(x,y)]$, whose value is determined by interpolation, if $\bar{\mathbf{f}}(x',y')$ is discretely sampled. The variables u and v therefore represent the x and y displacements of a point (x, y) in the new coordinate system (x', y'). Obviously this coordinate mapping can be quite general in nature. However, it should be restricted in some way to reflect the realistic range of possibilities for the warped shape of the prior model. In most cases it is desirable that the coordinate transformation be one to one in the domains of interest, which implies that the transformation is invertible. Another generally desirable property might be that the transformation be continuous, preserving the locality of neighborhoods. The warping can be accomplished equally well by means of the inverse transformation, that is, from (x', y') to (x, y). The choice is merely a matter of computational convenience and depends on the specific situation. More detail regarding the formulation of the warp can be found in Ref. 9.

A. Physical Analogy

It is appealing to interpret a two-dimensional warp in terms of an analogous physical system, a sheet of elastic material that undergoes distortion while being constrained to lie in the plane. Then the constraints that are placed on the warp roughly correspond to properties of the material being distorted, such as its stiffness. However, it must be recognized that it is not the substance of the object that is actually being warped. In material mechanics the strain corresponds to the first derivative of the mapping. For example, $\partial u/\partial x$ specifies the amount of linear stretching in the x direction producing normal strain, which can occur as either expansion (>0) or contraction (<0). The stress that is induced in the material is proportional to the strain. Thus the strain-energy density, found by integrating the stress with respect to the strain, is proportional to the square of the strain (see Ref. 10, p. 150). Then, for small deformations in linear materials

obeying Hooke's law, the normal strain-energy density is written as

$$w_{\text{normal}} = c_1\left(\frac{\partial u}{\partial x}\right)^2 + c_2\left(\frac{\partial v}{\partial y}\right)^2. \tag{15}$$

Material elements can experience another type of strain in the form of twisting, which induces shear. For small distortions the shear strain-energy density is

$$w_{\text{shear}} = c_3\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right)^2. \tag{16}$$

The coefficients c_1 and c_2 are proportional to the effective elastic moduli in the x and y directions, respectively, and c_3 is proportional to the shear modulus of the fictitious material. In an actual physical system, they are a property of the material (see Ref. 10, p. 365). For the flexible prior, they are set to achieve the properties desired for the warping of the prior, which should be based on prior information concerning the types of geometrical variation expected for the objects that are being imaged.

The Jacobian of the transformation from (x, y) to (x', y') is the determinant

$$J \equiv \frac{\partial(x', y')}{\partial(x, y)} = \frac{\partial x'}{\partial x}\frac{\partial y'}{\partial y} - \frac{\partial x'}{\partial y}\frac{\partial y'}{\partial x}$$

$$= 1 + \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial u}{\partial x}\frac{\partial v}{\partial y} - \frac{\partial u}{\partial y}\frac{\partial v}{\partial x}, \tag{17}$$

which gives the ratio of the change in area of a differential element produced by the transformation. For the transformation to be invertible, the Jacobian must be nonzero.

This conceptual physical model of geometrical distortion may have no connection with the material from which the object being reconstructed is actually composed. Indeed, the choices for the c_i in the above equations are not restricted by the usual constraints that regulate physical systems.[10] Instead, the choices should reflect the range of reasonable configurations that the prior distribution can assume for the class of objects being imaged. The Poisson ratio ν, which specifies the amount of contraction of an unconstrained material element in the direction perpendicular to an applied tension, may well take on values that would be considered unphysical for real materials. It might be set to zero or even to a negative value. For example, $\nu = -1$ indicates that expansion in one direction is most likely accompanied by expansion in the orthogonal direction as well. Such a value corresponds to a preference for similarity transformations, those that maintain shapes. Also, because material is not actually being distorted in the warping process, it is not necessary to scale the amplitude of the prior by the Jacobian of the transformation to conserve mass.

In this physical analogy, the MAP equation [Eq. (6)] corresponds to the solution of a problem in static equilibrium. The $\Lambda(\mathbf{f})$ represents the potential energy associated with the measurements, and $\Pi(\mathbf{f})$ represents the potential energy associated with the prior information. The gradients of these potentials represent forces. The force that moves the solution away from its default, which is defined by $\mathbf{d} = 0$ and $\mathbf{a} = \bar{\mathbf{a}}$, is provided by the data in the form of $\nabla_{\mathbf{d}}\Lambda$ and $\nabla_{\mathbf{a}}\Lambda$, respectively. Equation (6) states that the MAP solution balances the force of the data against the force of the prior. We may again note that linear dependence of the force on displacements from default values arises directly from the assumption of Gaussian-shaped probability distributions.

B. Priors on the Warp

We can take a cue from the above physical model as to how to specify constraints on the warp. Clearly the degree of local distortion is related to the first derivatives of the mapping. Since the strain-energy density is proportional to the square of the first derivatives, a reasonable way to control the amount of distortion is by limiting the strain energy. For example, we may wish to minimize the total normal strain energy, given for small deformations by

$$W_{\text{normal}} = \int w_{\text{normal}}\,dx\,dy$$

$$= \int\left[c_1(x, y)\left(\frac{\partial u}{\partial x}\right)^2 + c_2(x, y)\left(\frac{\partial v}{\partial y}\right)^2\right]dx\,dy, \tag{18}$$

and similarly for the total shear energy:

$$W_{\text{shear}} = \int w_{\text{shear}}\,dx\,dy = \int c_3(x, y)\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right)^2 dx\,dy. \tag{19}$$

One chooses the region of integration in Eqs. (18) and (19) in a manner that is consistent with the problem. For large deformations a more complete calculation must be done. In accordance with the physical interpretation of the preceding section, we take the negative logarithm of the prior probability on the warp to be proportional to the total strain energy $W_{\text{total}} = W_{\text{normal}} + W_{\text{shear}}$. In other words, the prior probability on the warp is a Gibbs distribution:

$$p(\text{warp}) = \exp(-W_{\text{total}}). \tag{20}$$

The role of the coefficients c_i in Eqs. (15) and (16) is then clearly identified as that of specifying the strength of the prior on the warp relative to the prior on the amplitude and relative to the likelihood. Whereas the preceding expressions are satisfactory for our present purposes, a more satisfactory formulation in which the expressions are independent of the choice of the initial orthogonal coordinate system is possible.[9]

Selection of these coefficients amounts to choosing the priors on the warp and should correspond to knowledge of the relative degrees of variability that are encountered for the objects under study. If it were deemed desirable in the warping to maintain right angles between grid lines, as in conformal mapping, then no shear would be allowed, even locally. This constraint could be enforced by requiring the shear energy density [Eq. (16)] to be zero, which would implicitly place constraints on the parameters of the warp. Alternatively, conformality can be achieved by making c_3 very large compared with c_1 or c_2.[11]

Because constraints of this type may be expressed in general terms, one need not be limited to the simple polynomial transformations that are assumed in Subsection 3.C below. Quite general forms are possible because the constraint of minimizing the total strain energy of the warp will sufficiently control the warp parameters. If extreme local distortion were desired, one could use a finite-element representation to describe the mapping.[12] Such an approach, taken by Brackbill and Saltzman,[11] seems particularly well suited to the present method because

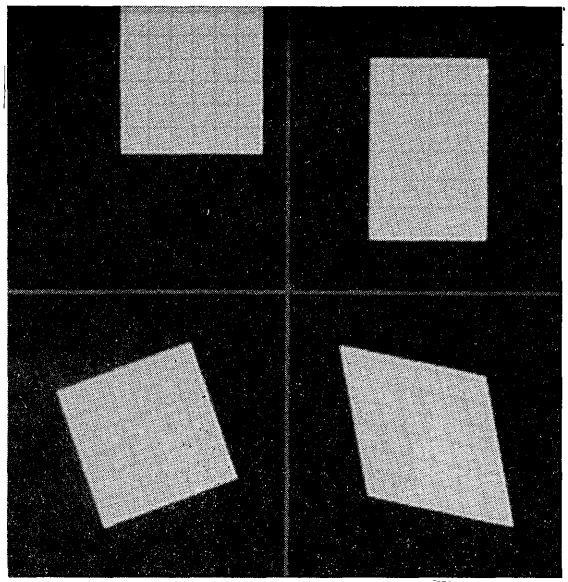

Fig. 1. Examples of warps achieved through a coordinate transformation consisting of the polynomial warp, Eqs. (21), including only the constant terms (upper left) and the linear terms (others). The scene that is being warped consists of a square, centered in the (x', y') plane, with a superimposed Cartesian grid. The lower-right image demonstrates the shearing effect that occurs when the mapping is not conformal.

images are typically represented by pixels that give the image values on Cartesian grids. The finite elements that one uses to represent the image need not coincide with the pixel representation. However, such a representation for the warp would preserve line elements that were located between pixels and thus would be useful in maintaining a prior that was defined on those line elements.[13]

C. Polynomial Warp

Some simplicity is achieved if we express the coordinate transformation [Eqs. (14)] as a polynomial expansion:

$$x' = \sum_{mn} a_{mn} x^m y^n; \qquad y' = \sum_{mn} b_{mn} x^m y^n, \tag{21}$$

where the coefficients a_{mn} and b_{mn} are represented as elements in the parameter vector $\mathbf{a}$ introduced in Subsection 2.B. It is recognized that the constant terms in Eqs. (21), a_{00} and b_{00}, amount to a simple shift in the position of the prior. The linear terms a_{10}, a_{01}, b_{10}, and b_{01} can produce change of scale, rotation, and skewing of the coordinates, which is called shear. The effects elicited by these terms are shown in Fig. 1. The quadratic terms can give rise to bending of the coordinate grid, as shown in Fig. 2. While the quadratic terms can severely skew the coordinates, certain combinations of coefficients can result in conformal mapping, which locally preserves the right angles between the original coordinates.

With Eqs. (21) the second factor inside the sum in Eq. (13) is

$$\frac{\partial \bar{\mathbf{f}}_i}{\partial a_{mn}} = \frac{\partial \bar{\mathbf{f}}_i}{\partial x'} \frac{\partial x'}{\partial a_{mn}} = x^m y^n \frac{\partial \bar{\mathbf{f}}_i}{\partial x'},$$
$$\frac{\partial \bar{\mathbf{f}}_i}{\partial b_{mn}} = \frac{\partial \bar{\mathbf{f}}_i}{\partial y'} \frac{\partial y'}{\partial b_{mn}} = x^m y^n \frac{\partial \bar{\mathbf{f}}_i}{\partial y'}. \tag{22}$$

Use of the polynomial expressions of the warp [Eqs. (21)] results in expressions for the strains that are likewise polynomials with coefficients that are quadratic in the a_{mn} and b_{mn} parameters. Then the total strain energy W_{total} will also be quadratic in these warp parameters. The proposed prior on the warp reproduces the form of the second term of Eq. (9), which was based on a Gaussian prior probability distribution. The matrix $\mathbf{R}_\mathbf{a}^{-1}$ is therefore related to the c_i in Eqs. (15) and (16). The degree to which distortion occurs is governed by the balance between $\mathbf{R}_\mathbf{a}^{-1}$ and $\mathbf{R}_\mathbf{d}^{-1}$ in Eq. (9) and on the strength of these priors relative to the likelihood. If the range of integration in Eqs. (18) and (19) subtends a rectangular region, e.g., the full reconstruction, and the c_i are constant, then the integrals are easy to perform. However, in many applications it might make more sense to integrate only over the extent of the object that is being warped, which could complicate the analytic evaluation of Eqs. (18) and (19).

Although the warp that is given by the polynomial expansion is convenient, it suffers from a few fundamental drawbacks. First, it does not provide much local flexibility without including high orders. This limitation might not be a significant problem for applications that involve only small distortions. Second, when second- or higher-order terms are admitted, the mapping will fold back on itself at some value of x and y. Severe mapping distortions are bound to occur, although they may happen only outside the support of the reconstruction and hence may be inconsequential. A minor computational difficulty of the polynomial warp is that each parameter a_{mn} or b_{mn} globally affect $\bar{\mathbf{f}}$, so the derivative with respect to each parameter in Eq. (13) involves a sum over all pixels. In contrast, if each warp parameter affects only a local region, as in a finite-element representation of the warp, the sum would be limited to that local region.

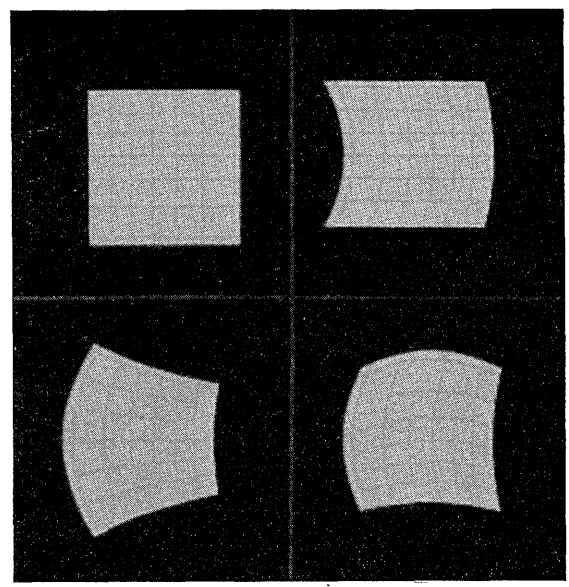

Fig. 2. Examples of warps achieved through a coordinate transformation consisting of the polynomial warp, Eqs. (21), including only the quadratic terms. Although the shape of the square (upper left) appears to be unaltered, the interior grid is strained. The lower-left image shows a conformal mapping in which the right angles between grid lines are locally preserved.

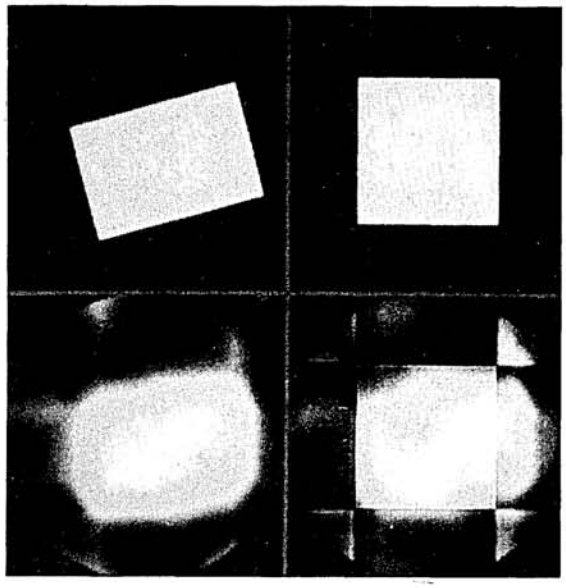

Fig. 3. Some results for tomographic reconstruction of an original scene (upper left) based on four parallel views that are equally spaced over 180°. The minimum-norm solution is provided by the ART algorithm (lower left); the MAP reconstruction (lower right) is obtained from ~~a flexible prior~~ (upper right).
an inflexible prior described in Sect. 2.A
(correction made 20-03-2006 KMH)

4. EXAMPLES

To demonstrate the proposed approach, I first present a simple example of reconstruction from a limited number of views. Figure 3 shows the original scene, which consists of a tilted rectangle. The images in this example are all 128 × 128 pixels in size. Four noiseless parallel projections of this object, taken at 45° angular increments, are assumed to be available. The result of the unconstrained algebraic reconstruction technique (ART),[14] which is known[15] to converge to a minimum-norm solution of the measurement equations [Eq. (2)], is predictably quite poor. If it were known beforehand that the object being imaged looked something like a rectangle and that its amplitude was unity, the square shown in the upper-right panel might be hypothesized as a prior. The resulting standard MAP solution (lower right) is even worse than the ART reconstruction because the square does not provide a good approximation of the actual rectangle. It differs from the actual rectangle in position, size, and orientation. The imprint of this incorrect guess on the MAP reconstruction is obvious and disastrous.

Figure 4 (lower right) shows the MAP reconstruction that was obtained from the same data and the same prior as was used in Fig. 3, but it includes the flexibility provided by the linear warp. Adding flexibility to the square prior permits the algorithm to shift, rotate, lengthen one dimension, and shorten the other to match the data. Even though the warping did not preclude shear, the force of the data was sufficient to rule it out. The final result agrees quite closely with the original scene. In some sense this result is trivial because the actual object can be perfectly matched by the linear warp that was applied to the square prior.

A nontrivial example of tomographic reconstruction is shown in Fig. 5. In this example five sets of parallel projections, equally spaced over 180°, of the sausage-shaped object are given. I assume that it is known that the object has right–left symmetry. This assumption of symmetry makes two of the five views redundant because their mirrored projections are just opposite to one of the other views. Therefore these reconstructions are effectively based on only three distinct views.

The 32 × 32 ART reconstruction presents only a vague result. To exaggerate the extent to which flexibility can be incorporated in the reconstruction process I use a circle for the prior probability distribution. The important aspects that I aim to include in the reconstruction are the expected sharp boundary of the object and its known constant amplitude. To permit a fair degree of flexibility I use a third-order polynomial coordinate transformation. The assumed right–left symmetry reduces the number of warp coefficients from 20 to 11. In this example only diagonal elements are nonzero in the $\mathbf{R}_a^{-1}$ matrix, and these are chosen to be quite small to permit maximum warping. The resulting MAP reconstruction (Fig. 5, lower right) reasonably matches the original, given that the available data consist of only three distinct views. If the shape of the object were better known beforehand, presumably one would obtain a superior reconstruction.

Even the cubic polynomial warp seems to provide insufficient flexibility to make the circular prior match the original sausage-shaped object. This example illustrates the fact that the reconstruction not only comprises the warped model but also includes deviations from the model. Because it is unable to achieve a match to the limited projection data through a warping of the circular prior, the reconstruction algorithm must provide deviations from the model in the form of a slight depression inside the cusp of the object and additional contributions near the tips.

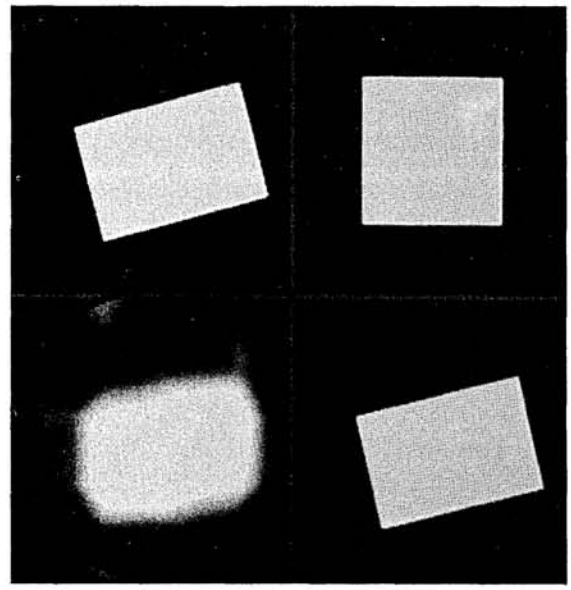

Fig. 4. Results for tomographic reconstruction as described in Fig. 3. However, the prior (upper right) that was used for the MAP reconstruction (lower right) is rendered flexible by the proposed warp process that includes the constant and linear terms in Eqs. (21). The left-hand images are the same as in Fig. 3. The six warp parameters are determined as part of the reconstruction process.

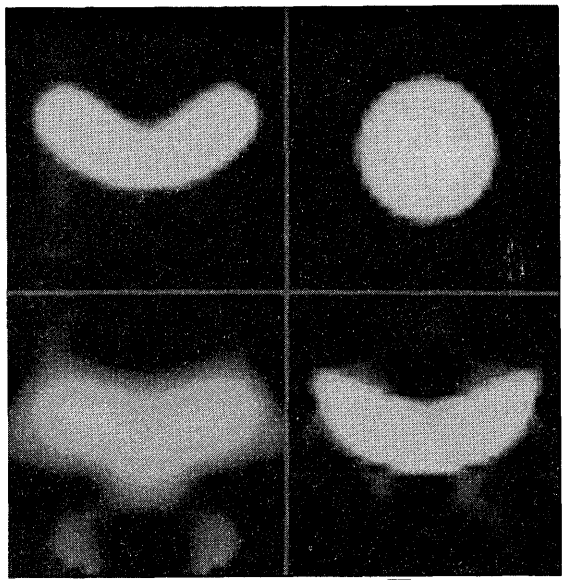

Fig. 5. ART reconstruction (lower left) that was obtained from five noiseless parallel views poorly reproduces the original sausage-shaped object (upper left). The MAP reconstruction (lower right) that was obtained from the same data is based on the circular prior distribution (upper right) subjected to a polynomial warp of third order, constrained to possess left–right symmetry.

5. DISCUSSION

In the examples presented here no constraining limits on the reconstruction values were employed. The use of constraints, such as nonnegativity, has been shown to provide a *bona fide* benefit for reconstruction from limited data.[16–19] The use of such amplitude constraints in combination with the flexible prior could prove to be extremely powerful. The constraints for each pixel might depend on its position relative to the warped prior model. For instance, the reconstruction values in the above example could be required to be zero outside the initial object model and lie between zero and an upper limit inside. These constraints would carry over into the final reconstruction coordinates.

The polynomial mapping that I used here was chosen for its simplicity. The ultimate implementation of the proposed warp would be in terms of a finite-element representation.[11,12] Then every aspect of the warp might depend on the position in the prior model. For example, $\mathbf{R}_a^{-1}$ in Eq. (9) or, equivalently, the c_i in Eqs. (15) and (16) that describe the rigidity of the warp might be a function of (u,v) and therefore a function of (x,y). With this type of model, some regions of the object could be permitted to distort considerably while others remained stiff. With the latitude that is available in the flexible prior approach, the algorithm developer gains exquisite control over the reconstruction process.

I have not addressed the issue of how to choose the strength of the priors on the warp parameters. This choice is critical. The amount of warping is regulated by the strength of the priors on the warp relative to likelihood and by the balance between the deviation from the warped model and the amount of warp regulated by the priors' strength relative to the strength of the prior on the deviation. The relative strengths of the individual terms in Eq. (9) affect the shape of the warped grid lines.[11] If one assigns too much strength to the prior, the final reconstruction will generally be pulled toward the default solution, which results in a significant bias away from the truth. These complex relationships must be dealt with in the future.

In this discussion I have assumed that the fictitious warping material possesses only a linear response to deformations. Many other types of nonlinear behavior are possible. For example, the material might yield above a critical strain, permitting it to flow when the data demand large warps. After yielding, the material could even heal itself to take on its original stiffness. If the reconstruction process is viewed as a time-dependent problem, then similar effects can be attained with elastoviscous materials for which the solution is allowed to creep into its final form.

I have only considered the use of bulk properties of the warped material. In some three-dimensional situations, it may be most appropriate to warp surfaces or curves (two- and one-dimensional manifolds). What happens inside the volume may be unimportant or of secondary interest. The strain-energy expressions would then be changed to focus on these aspects of the geometry. An example of a one-dimensional manifold is the snake model that was proposed for use in computer vision by Kass *et al.*,[20] which mimics an elastic rod that can be bent into a curve (in two or three dimensions) to conform to some specified image features. Such an approach is used there to regularize an otherwise ill-posed problem in much the same way that the flexible prior models used here regularize the reconstruction problem.

The notion of introducing geometrical flexibility into Bayesian estimation clearly extends to all types of reconstruction problems, including deblurring, decoding of coded-aperture images, and the solution of inverse problems in general. This new approach to image reconstruction has ramifications in virtually every other field of imaging. Flexible models are quickly finding use in many aspects of computer vision.[21–24] They are being used to match magnetic resonance images to generic shapes from a brain atlas.[25,26] Indeed, the flexibility that warping provides will likely become an essential tool in every area of image analysis and image recognition. Without this flexibility computer models cannot capture the essential features of the real objects that they are supposed to represent.

ACKNOWLEDGMENTS

I benefited enormously from many stimulating discussions with Robert F. Wagner and Kyle J. Myers. Jerimiah U. Brackbill provided useful insight into proper constraints on warping. James Gee shed light on other ways of achieving the warp. This research was supported by the U.S. Department of Energy under contract W-7405-ENG-36.

REFERENCES

1. K. M. Hanson, "Limited angle CT reconstruction using *a priori* information," Proc. Int. Symp. Med. Imaging Image Interp. **1**, 527–533 (1982).

2. K. M. Hanson, "Bayesian and related methods in image reconstruction from incomplete data," in *Image Recovery: Theory and Application,* H. Stark, ed. (Academic, Orlando, Fla., 1987), pp. 79–125.
3. B. R. Hunt, "Bayesian methods in nonlinear digital image restoration," IEEE Trans. Comput. **C-26,** 219–229 (1977).
4. K. M. Hanson, "Reconstruction based on flexible prior models," in *Medical Imaging VI: Image Processing,* M. H. Loew, ed., Proc. Soc. Photo-Opt. Instrum. Eng. **1652,** 183–191 (1992).
5. H. L. Van Trees, *Detection, Estimation, and Modulation Theory* (Wiley, New York, 1968), Vol. I.
6. A. P. Sage and J. L. Melsa, *Estimation Theory with Applications to Communications and Control* (Krieger, Huntington, N.Y., 1979).
7. K. M. Hanson and G. W. Wecksung, "Bayesian approach to limited-angle reconstruction in computed tomography," J. Opt. Soc. Am. **73,** 1501–1509 (1983).
8. R. F. Wagner, K. J. Myers, and K. M. Hanson, "Task performance on constrained reconstructions: human observers compared with suboptimal Bayesian performance," in *Medical Imaging VI: Image Processing,* M. H. Loew, ed., Proc. Soc. Photo-Opt. Instrum. Eng. **1652,** 352–362 (1992).
9. K. M. Hanson, "Flexible prior models in Bayesian image analysis," in A. Mohammad-Djafari, ed., *Maximum Entropy and Bayesian Methods* (Kluwer, Dordrecht, The Netherlands, 1992).
10. S. C. Hunter, *Mechanics of Continuous Media* (Horwood, Chichester, U.K., 1976).
11. J. U. Brackbill and J. S. Saltzman, "Adaptive zoning for singular problems in two dimensions," J. Comput. Phys. **46,** 342–368 (1982).
12. R. D. Cook, *Concepts and Applications of Finite Element Analysis* (Wiley, New York, 1974).
13. S. Geman and D. Geman, "Stochastic relaxation, Gibb's distributions, and the Bayesian restoration of images," IEEE Trans. Pattern Anal. Mach. Intell. **PAMI-6,** 721–741 (1984).
14. R. Gordon, R. Bender, and G. Herman, "Algebraic reconstruction techniques for three-dimensional electron microscopy and x-ray photography," J. Theor. Biol. **29,** 471–481 (1970).
15. G. T. Herman and A. Lent, "Iterative reconstruction algorithms," Comput. Biol. Med. **6,** 273–294, 1976.
16. K. M. Hanson, "Method to evaluate image-recovery algorithms based on task performance," in *Medical Imaging II,* R. H. Schneider and S. J. Dwyer, eds., Proc. Soc. Photo-Opt. Instrum. Eng. **914,** 336–343 (1988).
17. K. M. Hanson, "Optimization for object localization of the constrained algebraic reconstruction technique," in *Medical Imaging III: Image Formation,* S. J. Dwyer, R. G. Jost, and R. H. Schneider, eds., Proc. Soc. Photo-Opt. Instrum. Eng. **1090,** 146–153 (1989).
18. K. M. Hanson, "Optimization of the constrained algebraic reconstruction technique for a variety of visual tasks," in *Proceedings of Information Processing in Medical Imaging XI,* D. A. Ortendahl and J. Llacer, eds., (Wiley–Liss, New York, 1990), pp. 45–57.
19. K. M. Hanson, "Method of evaluating image-recovery algorithms based on task performance," J. Opt. Soc. Am. A **7,** 1294–1304 (1990).
20. M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: active contour models," Int. J. Comput. Vision **1,** 321–331 (1988).
21. B. C. Vemuri, ed., *Geometric Methods in Computer Vision,* Proc. Soc. Photo-Opt. Instrum. Eng. **1570** (1991).
22. R. Szeliski, "Probabilistic modeling of surfaces," in *Geometric Methods in Computer Vision,* B. C. Vemuri, ed., Proc. Soc. Photo-Opt. Instrum. Eng. **1570,** 154–165 (1991).
23. R. Szeliski and D. Terzopoulos, "Physically-based and probabilistic models for computer vision," in *Geometric Methods in Computer Vision,* B. C. Vemuri, ed., Proc. Soc. Photo-Opt. Instrum. Eng. **1570,** 140–152 (1991).
24. S. M. LaValle and S. A. Hutchinson, "Considering multiple surface hypotheses in a Bayesian hierarchy," in *Stochastic and Neural Methods in Signal Processing, Image Processing, and Computer Vision,* S. Chen, ed., Proc. Soc. Photo-Opt. Instrum. Eng. **1569,** 2–15 (1991).
25. R. Bajcsy and S. Kovačič, "Multiresolution elastic matching," Comput. Vision, **46,** 1–21 (1989).
26. J. C. Gee, M. Reivich, L. Bilaniuk, D. Hackney, R. Zimmerman, S. Kovačič, and R. Bajcsy, "Evaluation of multiresolution elastic matching using MRI data," in *Medical Imaging V: Image Processing,* M. H. Loew, ed., Proc. Soc. Photo-Opt. Instrum. Eng. **1445,** 226–234 (1991).