Proc. of SPIE Vol. 1652, Medical Imaging VI: Image Processing, ed. M H Loew (Jun 1992) Copyright SPIE

Reconstruction based on flexible prior models

K. M. Hanson *

Los Alamos National Laboratory, MS P940
Los Alamos, New Mexico 87545 USA
email: kmh@lanl.gov

ABSTRACT

A new approach to Bayesian reconstruction is introduced in which the prior probability distribution is endowed with an inherent geometrical flexibility. This flexibility is achieved through a warping of the coordinate system of the prior distribution into that of the reconstruction. This warping allows various degrees of mismatch between the assumed prior distribution and the actual distribution corresponding to the available measurements. The extent of the mismatch is readily controlled through constraints placed on the warp parameters.

1. INTRODUCTION

Often the approximate shape of an object to be reconstructed is known beforehand. Bayesian methods of reconstruction can incorporate the structural characteristics of an object that are known *a priori*. These methods have been shown to substantially improve the accuracy of reconstructions obtained from very limited data. However, if the object under study differs even only slightly in size, shape, or position from the assumed model, use of this kind of prior can lead to very poor reconstructions [1, 2].

The above difficulties arise because the prior model typically is held fixed relative to the spatial coordinate system of the reconstruction [3]. A superior approach is proposed in which the prior model for the object being reconstructed is allowed to alter its dimensional characteristics to accommodate the data. Such an accommodation is facilitated by warping the coordinate system of the prior model onto the coordinate system of the reconstruction. A linear transformation between the two coordinate systems can accommodate changes in size, position, and orientation of the model. Furthermore, changes in shape are allowed by nonlinear transformations. Within the Bayesian framework, the parameters needed to specify the coordinate transformation are determined as part of the overall estimation/reconstruction problem. Through the judicious choice of the priors on the transformation parameters, the degree and type of warping is readily controlled. Weak priors on the transformations with many degrees of freedom result in a flimsy prior model. Constraints on the parameters can be employed to approximately maintain the initial shape or to allow it to become fairly contorted.

The power of this new approach to prior models is demonstrated with a simple example. For simplicity, the coordinate transformations are restricted to low-order polynomials in the present example.

2. THE BAYESIAN APPROACH

Fundamental to the Bayesian approach is the posterior probability, which is assumed 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 given by Bayes' 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 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 about their predicted values (in the absence of noise). The prior probability $p(\mathbf{f})$

*Supported by the United States Department of Energy under contract number W-7405-ENG-36.

encompasses the full prior information about the relative frequency of occurrences of all possible images. Any known constraints concerning impossible images ought to be included explicitly or implicitly in $p(\mathbf{f})$. It is often desired that a single image be quoted as the 'result'. We humans have difficulty visualizing the full multidimensional probability distribution that is the complete $p(\mathbf{f}|\mathbf{g})$. An appropriate choice for a single image is that which maximizes the *a posteriori* probability, called the MAP estimate.

The essence of the Bayesian approach involves 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 (ML) result. Unfortunately an infinite number of the ML solutions exist when the data are very limited, which is the situation we wish to address. Thus the prior is indispensable. However, the problem that arises when using the Bayesian approach is that the model for the prior is usually considered to be geometrically fixed. This restriction might seem curious as the approach is based on probability theory and so ought to allow for a continuum of possibilities that are ranked on the basis of their relative likelihood. Thus 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 latitude in geometry as well as amplitude.

2.1 Standard MAP formulation

We assume that 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. We assume that these measurements are degraded by additive noise with a known covariance matrix $\mathbf{R_n}$, which describes the correlations that exist between noise fluctuations. The measurements can then be represented by a vector of length M

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

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

Because the probability is a function of continuous parameters, namely the N pixel values of the image, it is actually a probability density, designated by a small $p()$. From Bayes law, the negative logarithm of the posterior probability density is given by

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

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

$$-\log[p(\mathbf{g}|\mathbf{f})] = \Lambda(\mathbf{f}) = \tfrac{1}{2}\chi^2 = \tfrac{1}{2}(\mathbf{g} - \mathbf{Hf})^{\mathrm{T}}\mathbf{R}_{\mathbf{n}}^{-1}(\mathbf{g} - \mathbf{Hf}) , \tag{4}$$

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 assume are known *a priori*.

The second term $\Pi(\mathbf{f})$ comes from the prior-probability distribution. It should incorporate as much as possible the known characteristics of the original image. Here we use a Gaussian distribution for the prior, whose negative logarithm may be written as

$$-\log[p(\mathbf{f})] = \Pi(\mathbf{f}) = \tfrac{1}{2}(\mathbf{f} - \bar{\mathbf{f}})^{\mathrm{T}}\mathbf{R}_{\mathbf{f}}^{-1}(\mathbf{f} - \bar{\mathbf{f}}) , \tag{5}$$

where $\bar{\mathbf{f}}$ is the mean and $\mathbf{R_f}$ is the covariance matrix of the prior-probability distribution. The difficulty with this standard MAP approach is that $\bar{\mathbf{f}}$ is spatially fixed. In any particular situation, the actual object may differ in location, size, or shape. Any of these errors can destroy the usefulness of the prior [4, 2].

2.2 MAP based on flexible prior

To build flexibility into the prior, we therefore consider $\bar{\mathbf{f}}$ to be a function of several parameters, represented by the vector $\mathbf{a}$, that is, $\bar{\mathbf{f}}(\mathbf{a})$. For convenience we consider the reconstruction to be given in terms of the deviation of $\mathbf{f}$ from $\bar{\mathbf{f}}$:

$$\mathbf{d} = \mathbf{f} - \bar{\mathbf{f}} \ , \tag{6}$$

keeping in mind that the full reconstruction is really the sum $\mathbf{f} = \mathbf{d} + \bar{\mathbf{f}}$. We now consider $\mathbf{d}$ and $\mathbf{a}$ to be the independent variables in the reconstruction problem. The negative log(likelihood) is now expressed as

$$-\log[p(\mathbf{g}|\mathbf{d},\mathbf{a})] = \Lambda(\mathbf{d},\mathbf{a}) = \tfrac{1}{2}[\mathbf{g} - \mathbf{H}(\mathbf{d}+\bar{\mathbf{f}})]^{\mathrm{T}}\mathbf{R}_{\mathbf{n}}^{-1}[\mathbf{g} - \mathbf{H}(\mathbf{d}+\bar{\mathbf{f}})] \ , \tag{7}$$

realizing that $\bar{\mathbf{f}}$ is a function of $\mathbf{a}$. In Bayesian tradition, we must supply a prior for all variables. As before, we use a Gaussian distribution for the prior, whose negative logarithm may be written as

$$\Pi(\mathbf{d},\mathbf{a}) = \tfrac{1}{2}\mathbf{d}^{\mathrm{T}}\mathbf{R}_{\mathbf{d}}^{-1}\mathbf{d} + \tfrac{1}{2}(\mathbf{a}-\bar{\mathbf{a}})^{\mathrm{T}}\mathbf{R}_{\mathbf{a}}^{-1}(\mathbf{a}-\bar{\mathbf{a}}) \ , \tag{8}$$

where we have assumed that $\mathbf{a}$ and $\mathbf{d}$ are statistically independent. The first term is equivalent to Eq. (5), so $\mathbf{R_d}$ takes the place of $\mathbf{R_f}$. As before, the optimization function ϕ is the sum of Eqs. (7) and (8). The choice of the relative weight of the likelihood (7) and the prior (8) is critical, as it affects how well information is transfered to the observer of the image [5].

2.3 Reconstruction Procedure

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

$$\nabla_{\mathbf{d}}\phi = 0 \ \text{ and } \ \nabla_{\mathbf{a}}\phi = 0 \ , \tag{9}$$

provided the region of support for the solution is unlimited, that is, the solution is unconstrained. The gradient with respect to the likelihood is

$$\nabla_{\mathbf{d}}\Lambda = \nabla_{\bar{\mathbf{f}}}\Lambda = \mathbf{H}^{\mathrm{T}}\mathbf{R}_{\mathbf{n}}^{-1}[\mathbf{g} - \mathbf{H}(\mathbf{d}+\bar{\mathbf{f}})] \ , \tag{10}$$

from which we obtain for the gradient of ϕ with respect to $\mathbf{d}$,

$$\nabla_{\mathbf{d}}\phi \ = \ \mathbf{R}_{\mathbf{d}}^{-1}\mathbf{d} \ + \ \mathbf{H}^{\mathrm{T}}\mathbf{R}_{\mathbf{n}}^{-1}[\mathbf{g} - \mathbf{H}(\mathbf{d}+\bar{\mathbf{f}})] \ . \tag{11}$$

In computed tomography (CT), the matrix operation $\mathbf{H}^{\mathrm{T}}$ is the familiar backprojection process.

The gradient of ϕ with respect to parameter a_j is

$$[\nabla_{\mathbf{a}}\phi]_j = \frac{\partial\phi}{\partial \mathrm{a}_j} = [\mathbf{R}_{\mathbf{a}}^{-1}(\mathbf{a}-\bar{\mathbf{a}})]_j \ + \ \sum_i \frac{\partial\Lambda}{\partial\bar{\mathrm{f}}_i}\frac{\partial\bar{\mathrm{f}}_i}{\partial \mathrm{a}_j} \ , \tag{12}$$

where the sum is over the pixels of the reconstruction. The first term comes from the prior (8) and the second from the likelihood (7). The first quantity inside the sum is given by Eq. (10) and the second is given by the functional dependence of $\bar{\mathbf{f}}$ on a_j.

The MAP solution characterized by Eq. (9) can be found by the method of steepest descent, using (10) and (12) for the gradients. Although this method is computationally inefficient, it suffices for the present study.

3. THE WARP

One particularly interesting way to introduce flexibility into a fixed prior distribution is to warp the coordinate system of the prior into the coordinate system of the reconstruction. This method has the advantage that it can be applied to general prior distributions. The prior itself need not be given in a parametrized form.

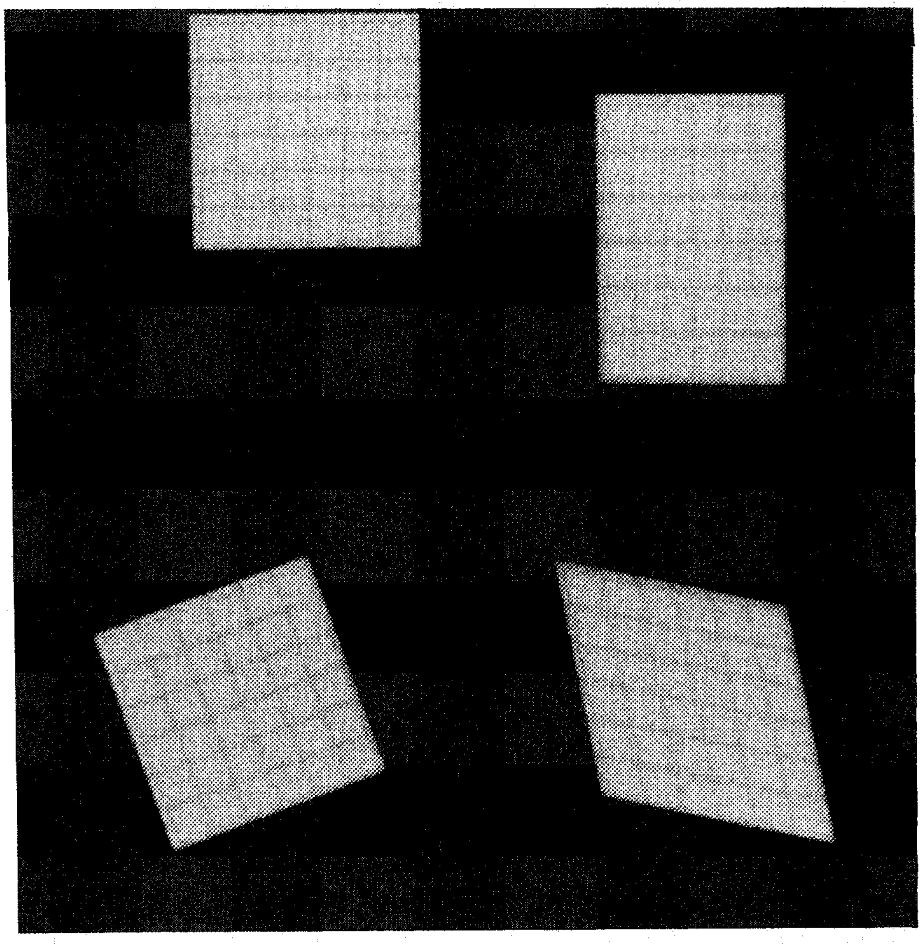

Figure 1: Example of warp achieved through a coordinate transformation consisting of the polynomial warp Eq. (14) including only the constant terms (upper-left) and the linear terms (others). The lower-right image demonstrates the shear effect that occurs when the mapping is not conformal.

3.1 Flexible prior

Suppose that the prior $\bar{\mathbf{f}}$ is specified as a function of the coordinates (u, v). If the coordinate system of the reconstruction is (x, y), we can effect a warping of $\bar{\mathbf{f}}(u, v)$ by means of a transformation of points (x, y) into the original (u, v) coordinates:

$$u = u(x, y)\ ; \quad v = v(x, y)\ . \tag{13}$$

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.

For simplicity, we will assume that the coordinate transformation is given as a polynomial expansion

$$u = \sum_{mn} a_{mn} x^m y^n\ ; \quad v = \sum_{mn} b_{mn} x^m y^n\ , \tag{14}$$

where the a_{mn} and b_{mn} coefficients are represented as elements in the parameter vector $\mathbf{a}$ used in Sec. 2.2. It is recognized that the constant terms in (14), 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}, correspond to a change of scale, rotation, and skewing of the coordinates, called shear. The effects of these term are shown in Fig. 1. The quadratic terms give rise to bending of the coordinates, as shown in Fig. 2. While the quadratic terms can skew the coordinates severely, certain combinations of coefficients can result in conformal mapping, which locally preserves the right angles between the original coordinates.

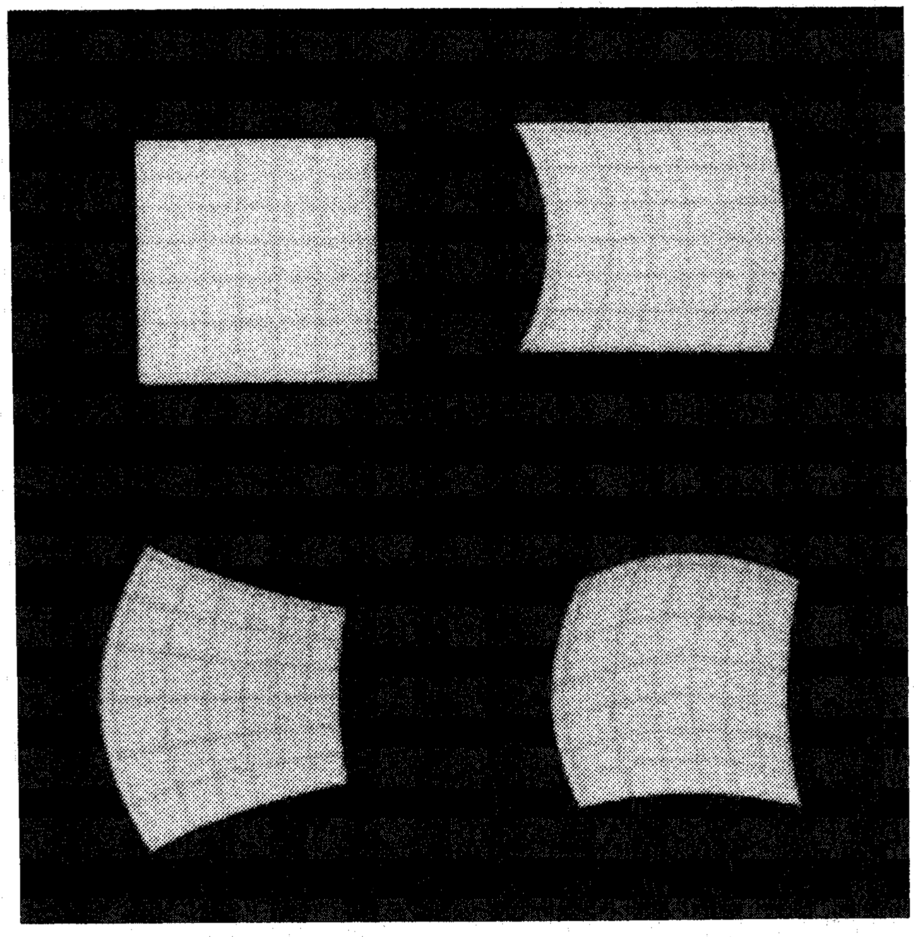

Figure 2: Example of warp achieved through a coordinate transformation consisting of the polynomial warp Eq. (14) including only the quadratic terms. The lower-left image shows a conformal mapping in which the right angles between grid lines are locally preserved.

With (14), the second term inside the sum in Eq. (12) is

$$\frac{\partial \bar{\mathrm{f}}_i}{\partial a_{mn}} = \frac{\partial \bar{\mathrm{f}}_i}{\partial u}\frac{\partial u}{\partial a_{mn}} = x^m y^n \frac{\partial \bar{\mathrm{f}}_i}{\partial u}\ ; \quad \frac{\partial \bar{\mathrm{f}}_i}{\partial b_{mn}} = \frac{\partial \bar{\mathrm{f}}_i}{\partial v}\frac{\partial v}{\partial b_{mn}} = x^m y^n \frac{\partial \bar{\mathrm{f}}_i}{\partial v}\ . \tag{15}$$

Although the warp given by the polynomial expansion is convenient, it suffers from a few fundamental drawbacks. First, it does not provide much local flexibility without including higher orders. This might not be a significant problem for applications in which only a small distortion is desirable. The second difficulty, which occurs when second- or higher-order terms are admitted, is that the mapping will fold back on itself at some value of x and y. This results in severe distortions, although they may occur only outside the support of the reconstruction, and, hence, be unnoticeable.

3.2 Physical analog

It is appealing to interpret the warp in terms of an analogous physical system, a sheet of material that undergoes distortion. Then the priors 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, $\frac{\partial u}{\partial x}$ corresponds to normal strain; either expansion (> 0) or contraction (< 0).

The induced stress is proportional to the strain. So the strain energy density is given by the product of the induced stress times the strain and contains terms like:

$$w_{\text{normal}} = c_1 \left(\frac{\partial u}{\partial x}\right)^2 + c_2 \left(\frac{\partial v}{\partial y}\right)^2 , \tag{16}$$

and

$$w_{\text{shear}} = c_3 \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right)^2 . \tag{17}$$

The coefficients c_1 and c_2 are proportional to the effective elastic modulus 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. For the flexible prior, they are set to achieve the properties desired for the warping of the prior.

It should be emphasized that this conceptual physical model may have no connection with the material from which the object being reconstructed is composed. Indeed, the choices for the c_i in the above equations are not restricted by the usual constraints that regulate physical systems [6]. Instead, the choices should reflect the range of reasonable configurations for the object being imaged. The Poisson ratio, which specifies the amount of contraction perpendicular to the tension, may perfectly well be zero even though such a value might be 'unphysical'. Also note that, 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 analog, the MAP equations (7) and (8) correspond to a problem in static equilibrium. The equations represent potential energies and their gradients represent forces. The force that moves the solution away from the default solution, $\mathbf{d} = 0$ and $\mathbf{a} = \bar{\mathbf{a}}$, is provided by the data in the form of $\nabla\Lambda$.

3.3 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 through the strain energy. For example, we may wish to minimize the total strain energy, given 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. The region of integration is chosen in a manner consistent with the problem. If the integral is over the full rectangular reconstruction and the c_i are constant, (18) is easy to perform. The resultant expression for the total energy has the same form as the scond term of Eq. (8). Then the matrix $\mathbf{R}_{\mathbf{a}}^{-1}$ would be related to the c_i in Eqs. (16 and 17). 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. (8). However, it might make more sense to integrate only over the extent of the object being warped, which could complicate the evaluation of (18).

Priors on these coefficients should correspond to knowledge of the relative degrees of variability encountered for the objects under study. If it were deemed desirable in the warping to maintain right angles between grid lines, that is, that the mapping be conformal, then no shear would be allowed, even locally. This constraint could be enforced by requiring the shear energy density (17) to be zero, which would place restrictions on the parameters of the warp.

Because constraints of this type may be expressed in general terms, one need not be limited to the simple polynomial transformations given in Eq. (14). Quite general forms are possible as the constraint of minimizing the total strain energy of the warp will sufficiently control the warp parameters. One could perhaps employ splines in which the coefficients at a moderate number of control points comprise the variables. Ultimately, if extreme local distortion were desired, one could use a finite-element representation to describe the mapping [7].

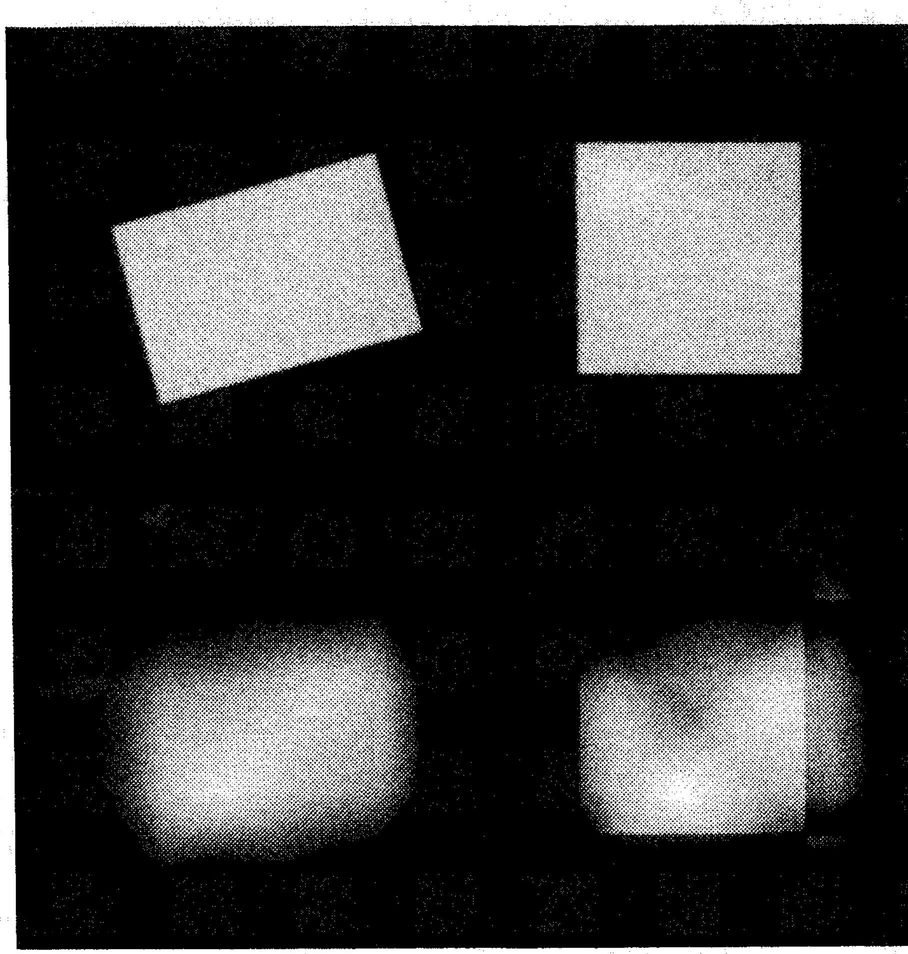

Figure 3: Some results for tomographic reconstruction of an original scene (upper-left) based on four parallel views, equally spaced over 180°. The minimum-norm solution is provided by the ART algorithm (lower-left); the MAP reconstruction (lower-right) is obtained from an inflexible prior (upper-right).

4. A SIMPLE EXAMPLE

To demonstrate the proposed approach, we use a simple example of the reconstruction from a limited number of views. Figure 3 shows the original scene consisting of a tilted rectangle. All the images in this example are 64×64 pixels in size. Four noiseless parallel projections of this object, taken at 45° angular increments, are assumed to be available. The result of the Algebraic Reconstruction Technique (ART) [8], which is known [9] to converge to a minimum-norm solution of the measurement equations (2), is predictably very poor. If it were known beforehand that the object being imaged looked something like a square, the square shown (UR) might be hypothesized as a prior. The resulting MAP solution (LR) is even worse than the ART reconstruction, basically because the square does not approximate very well the actual rectangle. It differs from the actual rectangle in position, size, and orientation. The imprint of this wrong guess on the MAP reconstruction is obvious.

Figure 4 shows the MAP reconstruction obtained from the same data and same prior as used in Fig. 4, but with the flexibility provided by the linear warp included. Adding flexibility to the square prior allows the algorithm to shift, rotate, lengthen one dimension, and shorten the other to match the data. We note that even though the warping did not preclude shear, the force of the data was sufficient to rule it out. The final result agrees very closely with the original scene.

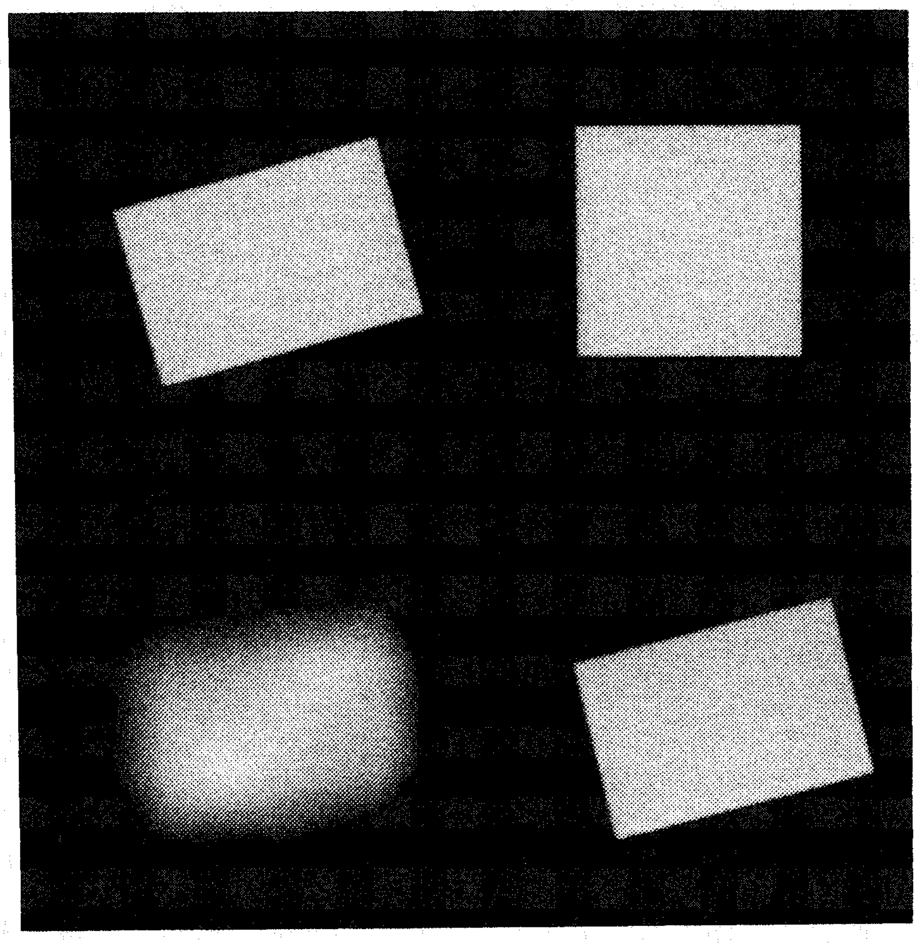

Figure 4: Results for tomographic reconstruction as described in Fig. 3. However, the prior (upper-right) used for the MAP reconstruction (lower-right) is rendered flexible by the proposed warp process that includes the constant and linear terms in Eq. (14). The six warp parameters are determined as part of the reconstruction process.

5. DISCUSSION

For the simple example used here, we have not employed any constraining limits on the reconstruction. The use of constraints, such as nonnegativity, have been shown to provide bona fide benefit for reconstruction from limited data [10, 11, 12, 13]. Use of such constraints in combination with the flexible prior could prove to be extremely powerful. The constraints could vary with position relative to the prior model. For example, the reconstruction might be required to be nonnegative outside an object, and between 0 and an upper limit inside.

The polynomial mapping used here was chosen for its simplicity. A much more general approach could be gained through the use of a finite-element method to describe the warp. Then, every aspect of the warp could be made a function of position in the prior model. For example, $\mathbf{R}_{\mathbf{a}}^{-1}$ in Eq. (8) or, equivalently, the c_i in Eqs. (16 and 17) that describe the rigidity of the warp, could be a function of u and v and, therefore, a function of x and y. With this type of model, some regions of the object could be allowed to distort considerably while others remain stiff. With the latitude available in such an approach, the algorithm developer gains exquisite control over the reconstruction process.

This new approach to image reconstruction has ramifications in all fields of imaging. Flexible models are quickly finding use in many aspects of computer vision [14]. They are being used to match MRI images to generic shapes from a brain atlas [15]. Indeed, the flexibility provided by warping will likely become an essential tool in every area of image analysis and image recognition. Without this flexibility, computer models can not capture the essential features of the real objects they are supposed to represent.

ACKNOWLEDGEMENTS

I acknowledge many spirited and provocative discussions with Robert F. Wagner and Kyle J. Myers.

References

[1] K. M. Hanson. Limited angle CT reconstruction using a priori information. *Proc. Inter. Symp. Med. Imag. and Imag. Interp.*, 1:527–533, 1982.

[2] K. M. Hanson. Bayesian and related methods in image reconstruction from incomplete data. In Henry Stark, editor, *Image Recovery: Theory and Application*, pages 79–125. Academic, Orlando, 1987.

[3] B. R. Hunt. Bayesian methods in nonlinear digital image restoration. *IEEE Trans. Comp.*, C-26:219–229, 1977.

[4] K. M. Hanson and G. W. Wecksung. Bayesian approach to limited-angle reconstruction in computed tomography. *J. Opt. Soc. Amer.*, 73:1501–1509, 1983.

[5] R. F. Wagner, K. J. Myers, and K. M. Hanson. (in press) Task performance on constrained reconstructions: human observers compared with suboptimal Bayesian performance. *Proc. SPIE*, vol. 1652, 1992.

[6] S. C. Hunter. *Mechanics of Continuous Media.* Ellis Horwood Ltd., Chichester, 1976.

[7] R. D. Cook. *Concepts and Applications of Finite Element Analysis.* John Wiley & Sons, New York, 1974.

[8] 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.

[9] G. T. Herman and A. Lent. Iterative reconstruction algorithms. *Comput. Biol. Med.*, 6:273–294, 1976.

[10] K. M. Hanson. Method to evaluate image-recovery algorithms based on task performance. *Proc. SPIE*, 914:336–343, 1988.

[11] K. M. Hanson. Optimization for object localization of the constrained algebraic reconstruction technique. *Proc. SPIE*, 1090:146–153, 1989.

[12] K. M. Hanson. Optimization of the constrained algebraic reconstruction technique for a variety of visual tasks. In D. A. Ortendahl and J. Llacer, editors, *Proc. Information Processing in Medical Imaging*, pages 45–57. Wiley-Liss, New York, 1990.

[13] K. M. Hanson. Method to evaluate image-recovery algorithms based on task performance. *J. Opt. Soc. Amer.*, A7:45–57, 1990.

[14] B. C. Vemuri, ed. *Geometric Methods in Computer Vision. Proc. SPIE*, vol. 1570, 1991.

[15] J. C. Gee, M. Reivich, L. Bilaniuk, D. Hackney, R. Zimmerman, S. Kovačič, and R. Bajcsy. Evaluation of multiresolution elastic matching using MRI data. *Proc. SPIE*, 1445:226–234, 1991.