

FLEXIBLE PRIOR MODELS IN BAYESIAN IMAGE ANALYSIS

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

ABSTRACT. A new class of prior models is proposed for Bayesian image analysis. This class of priors provides an inherent geometrical flexibility, which is achieved through a transformation of the coordinate system of the prior distribution or model into that of the object under analysis. Thus prior morphological information about the object being reconstructed may be adapted to various degrees to match the available measurements. An example of tomographic reconstruction illustrates the potential of this approach.

1. Introduction

One often encounters problems in the image analysis of objects that are ill-posed because of a lack of data. Bayesian methods help select the best of many possible solutions on the basis of the characteristics of the object that are known *a priori*. The prior knowledge concerning the object is incorporated in terms of prior probability distributions on the appropriate physical parameters. We wish to develop a method to take into account an object's approximate shape, when it is known beforehand. The model for the object being analyzed is allowed to alter its geometry to accommodate the data by warping the coordinate system of the prior model onto the coordinate system of the actual object. Thus the character of the edges of the model, as well as its morphology, are preserved. The proposed extension to the standard MAP technique overcomes its static definition of the prior providing flexibility in geometry and other characteristics. For a fuller description of the use of this flexible model, see [1, 2].

The power of this new approach to prior models is illustrated with an example of computed tomographic reconstruction in which the coordinate transformations are restricted to low-order polynomials.

2. Basic Formulation

We are given M discrete measurements that are linearly related to the amplitudes of the original image vector \mathbf{f} of length N , degraded by additive noise: $\mathbf{g} = \mathbf{H}\mathbf{f} + \mathbf{n}$,

where \mathbf{n} is the random noise vector, and \mathbf{H} is the measurement matrix. From Bayes' law the negative logarithm of the posterior probability density is given by

$$-\log[p(\mathbf{f}|\mathbf{g})] = \phi(\mathbf{f}) = \Lambda(\mathbf{f}) + \Pi(\mathbf{f}), \quad (1)$$

where the first term comes from the likelihood and the second term from the prior probability. For simplicity, a Gaussian distribution is chosen for the prior probability,

which is characterized by the known quantities $\bar{\mathbf{f}}$, the ensemble mean, and $\mathbf{R}_{\mathbf{f}}$, the ensemble covariance matrix. The negative logarithm of the prior probability on \mathbf{d} , the deviation from the model $\mathbf{d} = \mathbf{f} - \bar{\mathbf{f}}$, may be written as

$$\Pi(\mathbf{d}) = \frac{1}{2} \mathbf{d}^T \mathbf{R}_{\mathbf{d}}^{-1} \mathbf{d}, \quad (2)$$

Assuming additive Gaussian noise with a known covariance matrix $\mathbf{R}_{\mathbf{n}}$, the negative log(likelihood) is just one half chi-squared

$$-\log[p(\mathbf{g}|\mathbf{d}, \mathbf{a})] = \Lambda(\mathbf{d}, \mathbf{a}) = \frac{1}{2} \chi^2 = \frac{1}{2} [\mathbf{g} - \mathbf{H}(\mathbf{d} + \bar{\mathbf{f}})]^T \mathbf{R}_{\mathbf{n}}^{-1} [\mathbf{g} - \mathbf{H}(\mathbf{d} + \bar{\mathbf{f}})], \quad (3)$$

which is quadratic in the residuals.

The full Bayesian solution is characterized by the posterior probability $p(\mathbf{f}|\mathbf{g})$. However, to represent the result with a single image, an appropriate choice is the image that maximizes the *a posteriori* probability, called the MAP estimate [3], which equivalently minimizes ϕ . The choice of the relative weights of the log-likelihood (3) and the log-prior (2) is critical, as it affects how well the information contained in the data is transferred to the observer of the image [4].

minimum of $\phi(\mathbf{f})$, that is, at the MAP solution $\nabla_{\mathbf{f}} \phi(\mathbf{f}) = 0$.

The problem with the preceding standard Bayesian formulation is that the model for the prior is usually considered to be geometrically fixed [5, 6]. To build flexibility into the prior, we 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. Under the assumption that \mathbf{a} and \mathbf{d} are statistically independent, the prior on these new parameters $\Pi(\mathbf{a})$ is simply added to $\Pi(\mathbf{d})$. We take \mathbf{d} and \mathbf{a} to be the independent variables in the problem.

In reconstruction, we seek to estimate all pixel values in the original scene, that is \mathbf{f} . If there is no interest in the actual values of the parameters \mathbf{a} and \mathbf{d} , probability theory says we must marginalize over these nuisance parameters. As the integration over so many parameters is computationally very difficult, the solution for \mathbf{f} is calculated from the \mathbf{a} and \mathbf{d} that minimize ϕ , for which $\nabla_{\mathbf{d}} \phi = 0$ and $\nabla_{\mathbf{a}} \phi = 0$, provided the solution is not otherwise constrained. The MAP solution can be found by the method of steepest descent or some better method such as that of conjugate gradients. It is unlikely that one can guarantee that the uniqueness of the solution for ϕ . We must rely on a knowledgeable choice of initial parameters to guide the solution to a meaningful result.

3. Elastic Sheet Analogy

Suppose that the prior on the image amplitude, specified in terms of $\bar{\mathbf{f}}$ and $\mathbf{R}_{\mathbf{d}}$ in the present case, is given as a function of the spatial coordinates (x', y') . Geometrical flexibility of the prior is accomplished by transforming the (x, y) coordinates of the final image \mathbf{f} to the original (x', y') coordinates:

$$x' = x + u(x, y); \quad y' = y + v(x, y), \quad (4)$$

where u and v are the displacements of x and y . The warp could equally well be accomplished by means of the inverse transformation, that is, from (x', y') to (x, y) . The choice is

mostly a matter of computational convenience and depends on the specific situation. Obviously this coordinate mapping can be arbitrary. However, it should be restricted in some way to reflect the realistic range of possibilities for the warped shape of the prior model.

The warp should meet the following general requirements. The coordinate transformation should be one-to-one in the domains of interest, which implies the transformation is invertible. Without any previously known preference for orientation or position, the transformation should be isotropic and stationary (homogeneous). The formulation for the warp should be independent of the rectilinear coordinate system in which it is stated. Neighborhood relationships should be maintained so the transformation should be continuous. Homogeneous translations and rotations should be unimportant. We will propose a mechanism to control the warp that meets all these criteria.

It is natural to draw an analogy between a 2D warp and the distortion of a sheet of elastic material that is constrained to lie in a plane. Then the constraints placed on the warp are analogous to the elastic properties of the material being distorted. In material mechanics the strain corresponds to the first derivative of the mapping. For example, $\frac{\partial u}{\partial x}$ specifies the amount of linear expansion in the x -direction producing normal tensile strain. For deformations in linear materials obeying Hooke's law, the stress induced in the material is proportional to the strain. So the strain energy density, found by integrating the stress with respect to the strain, is proportional to the square of the strain.

We would like to generalize this notion of strain energy density for application to coordinate transformations. We employ tensor notation to achieve independence of the rectilinear coordinate system and allow easy extension to 3D. The displacement vector comprised of u and v is designated u_i . The subscripts run over the coordinate indices, i.e. $1 \sim x, 2 \sim y$. The strain is expressed as a symmetric tensor $e_{ij} = \frac{1}{2}(u_{i,j} + u_{j,i})$, where the comma denotes a derivative with respect to the appropriate coordinate.

plane. Exceptions relevant to 3D will be given in parentheses. Therefore, $i = 1, 2(3)$ and similarly for j .

Thus, the strains are

$$e_{11} = \frac{\partial u}{\partial x}, \quad e_{22} = \frac{\partial v}{\partial y}, \quad e_{12} = e_{21} = \frac{1}{2} \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right). \quad (5)$$

The first two quantities are the normal strain in the x - and y -directions and the last one is the usual expression for shear strain.

For small strains the general expression for the strain energy density in a linear isotropic medium is [7]

$$\begin{aligned} w &= \frac{1}{2} \lambda (e_{ii})^2 + \mu e_{ij} e_{ij} \\ &= \frac{1}{2} \lambda (e_{11} + e_{22})^2 + \mu (e_{11}^2 + e_{22}^2 + 2e_{12}^2), \end{aligned} \quad (6)$$

where λ and μ are the Lamé constants of elasticity. As usual, repeated subscripts within an expression imply summation over them. As μ controls the change in angles induced by the transformation, it is called the modulus of rigidity or shear modulus. We note that by construction the expression for w is invariant under rotation or translation of the (x, y) coordinate system.

The quantity $e_{ii} = \vartheta$ is called the area dilation, or volume dilation in 3D. It is linearly related to the Jacobian of the transformation from (x, y) to (x', y') , which is the determinant

$$J \equiv \frac{\partial(x', y')}{\partial(x, y)} = \frac{\partial x'}{\partial x} \frac{\partial y'}{\partial y} - \frac{\partial x'}{\partial y} \frac{\partial y'}{\partial x} = |\delta_{ij} + u_{i,j}|, \quad (7)$$

where δ_{ij} is the usual Kronecker delta. The Jacobian gives the ratio of the change in area of a differential element produced by the transformation. For small strains, $J \approx 1 + \vartheta$. For the transformation to be invertible (one-to-one), it is necessary that the Jacobian be nonzero.

The Lamé constants are a property of the material. Their relation to Young's modulus E and Poisson's ratio ν , more typically used in engineering, is given by

$$E = \frac{\mu(3\lambda + 2\mu)}{\lambda + \mu}, \quad \nu = \frac{\lambda}{2(\lambda + \mu)}. \quad (8)$$

Both λ and μ are nonnegative for real materials, which assures $w \geq 0$. Poisson's ratio specifies the relative amount of contraction in the direction perpendicular to an applied tension. In 2D the upper limit of $\nu = 1$ is attained by incompressible materials. Its upper limit in 3D is $\frac{1}{2}$.

4. Prior on the Warp

We propose to use a Gibbs' distribution [8] for the prior probability on the warp, which is proportional to $\exp(-W)$, where W is the total strain energy of the warp, given for small deformations by

$$W = \int w \, d\tau, \quad (9)$$

where $d\tau = dx \, dy$ is the differential area in (x, y) . Then the negative log-prior on the warp is simply W . The role of the Lamé constants in Eq. (6) is then clearly identified as that of specifying the strength of the log-prior on the warp relative to the log-prior probability on the amplitude and relative to the log-likelihood. While the above expressions hold only for small strains, valid evaluation for large strains may be achieved by numerically integrating (6) and (9).

It should be emphasized that the elastic constants for the conceptual physical model of the warp are not related to those of the material from which the object being studied is actually composed. Indeed, the choices for the Lamé constants are not restricted by the usual constraints regulating physical systems [7]. Instead, their selection should reflect the range of reasonable configurations the prior distribution can assume for the class of objects being imaged. Poisson's ratio might be set to zero, if a stretch in one direction is not expected to influence what happens in the other direction. Even 'unphysical' values are legitimate. For example, $\nu = -1$, would correspond to a similarity transformation, which maintains shapes, because that would indicate that an expansion in one direction is most likely accompanied by an equal expansion in the orthogonal direction. Also note that, because material is not actually being distorted in this warping process, it is not necessary to scale the amplitude of the prior by the Jacobian of the transformation to conserve mass.

If it were deemed desirable to maintain angles in the warping, as in conformal mapping, then no shear would be allowed, even locally. This constraint could be enforced by requiring the shear (ϵ_{12}) to be zero, which would implicitly place constraints on the parameters of the warp. Alternatively, conformality can be achieved by making μ very large compared to λ [9].

Note that, because (6) is based on a linear stress-strain relationship, there is nothing that precludes the Jacobian from going to zero. By assuming that the stress σ is proportional to a relative change in length of a differential element, that is $\sigma \propto \epsilon/(1 + \epsilon)$, where ϵ is the strain, the strain energy density would be proportional to $\epsilon - \log(1 + \epsilon)$, which avoids the dreaded condition $\epsilon = -1$. This expression is obviously related to the Burg entropy.

The constraint given in Eq. (9) is expressed in a general form that does not imply a representation for the warp. Wherever representation is used, the parameters involved correspond to the vector \mathbf{a} in Sec. 2. Ultimate control over local distortion can be had through the use of a finite-element representation to describe the mapping [10]. Such an approach, taken by Brackbill and Saltzman [9], seems particularly well suited to the present method, because images are typically represented by pixels that give the image values on Cartesian grids. Of course, the finite elements used to represent the image need not coincide with the pixel representation. However, such a representation for the warp would preserve line elements located between pixels, and thus would be useful in maintaining a prior defined on those line elements [8].

Some simplification is achieved if we express the coordinate transformation (4) as a polynomial expansion

$$u = \sum_{mn} a_{mn} x^m y^n ; \quad v = \sum_{mn} b_{mn} x^m y^n , \quad (10)$$

where the coefficients a_{mn} and b_{mn} are represented as elements in the parameter vector \mathbf{a} introduced in Sec. 2. Use of (10) 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 energies W will also be quadratic in these warp parameters, consistent with a Gaussian prior probability distribution. Although the polynomial expansion for the warp is convenient, it suffers from a few fundamental difficulties. First, it does not provide much local flexibility without including high orders. Second, the mapping will inevitably cease to be invertible at some values of (x, y) , when nonzero second- or higher-order terms are included.

5. The Knockwurst Example

In this example we assume that we are given five sets of parallel projections of the sausage-shaped object shown in Fig. 1. The projections contain 128 samples each and are taken at 36° increments in angle. Note that an assumption of right-left symmetry makes two of the five views redundant so these reconstructions are effectively based on only three distinct views!

The tomographic reconstructions are shown in Fig. 1. The 128×128 reconstruction obtained with the unconstrained Algebraic Reconstruction Technique (ART) [11] is predictably very poor. To exaggerate the extent to which flexibility can be incorporated in the reconstruction process, we use a circle of unit amplitude for $\bar{\mathbf{f}}$, the mean of the prior

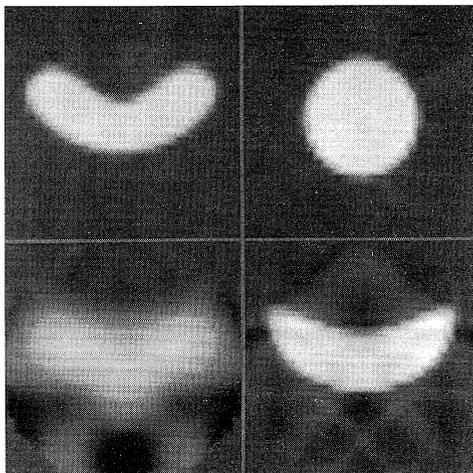


Figure 1:

The ART reconstruction (lower-left) obtained from five noiseless, parallel views taken at angular increments of 36° poorly reproduces the original sausage-shaped object (upper-left). The MAP reconstruction (lower-right) obtained from the same data is based on the circular prior distribution (upper-right) subject to a polynomial warp of third order.

probability distribution on amplitude. The important aspects that we wish to include in the reconstruction are the expected sharp boundary of the object and its constant amplitude. To permit a fair degree of flexibility, a third-order polynomial coordinate transformation is employed. The assumed right-left symmetry reduces the number of warp coefficients from 20 to 11. In the present example, the full expression given by (6) and (9) is replaced by the sum of the squared values of the a_{mn} and b_{mn} . The coefficients in this sum are chosen to be very small to allow maximum warping. The resulting 32×32 MAP reconstruction (LR) reasonably matches the original, given that the available data consist of only three distinct views.

6. Discussion

The choice of the strength of the priors on the warp parameters is critical in determining the shape of the solution. One way to determine the Lamé constants is through the posterior probability, in the same way as Gull suggests for finding the strength of the entropic prior [12]. A better method would be to base the choice on *bona fide* prior knowledge about the objects being imaged. More detailed prior information might even allow specification of the elastic constants as a function of position.

The linear-material approach taken here may be extended in many ways. The assumption of stress being proportional to relative change in dimension, mentioned in Sec. 4, leads to an energy density of entropic form. Other possibilities include endowing the fictitious

material being warped with nonlinear, ductile, or even elasto-viscous behavior. The flexible structures may be 1D (or 2D in a 3D problem) as it may be most appropriate to warp lines or surfaces.

The notion of introducing geometrical flexibility into the priors used in Bayesian analysis clearly extends to all types of image analysis. Flexible models have been used for several years in computer vision [13, 14] and to match MRI brain images to generic shapes from a brain atlas [15, 16].

7. Acknowledgments

I acknowledge many helpful prior discussions with Robert F. Wagner, Kyle J. Myers, Jerry U. Backbill, James C. Gee, and David R. Wolf. This work was supported by the United States Department of Energy under contract number W-7405-ENG-36.

References

- [1] K. M. Hanson. Reconstruction based on flexible prior models. *Proc. SPIE*, 1652:183–191, 1992.
- [2] K. M. Hanson. Bayesian reconstruction based on flexible prior models. *to be published in J. Opt. Soc. Amer.*, 1992.
- [3] A. P. Sage and J. L. Melsa. *Estimation Theory with Applications to Communications and Control*. Robert E. Krieger, Huntington, 1979.
- [4] R. F. Wagner, K. J. Myers, and K. M. Hanson. Task performance on constrained reconstructions: human observers compared with suboptimal Bayesian performance. *Proc. SPIE*, 1652:352–362, 1992.
- [5] B. R. Hunt. Bayesian methods in nonlinear digital image restoration. *IEEE Trans. Comp.*, C-26:219–229, 1977.
- [6] 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.
- [7] I. S. Sokolnikoff. *Mathematical Theory of Elasticity*. McGraw-Hill, New York, 1956.
- [8] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intell.*, PAMI-6:721–741, 1984.
- [9] J. U. Brackbill and J. S. Saltzman. Adaptive zoning for singular problems in two dimensions. *J. Comput. Phys.*, 46:342–368, 1982.
- [10] R. D. Cook. *Concepts and Applications of Finite Element Analysis*. John Wiley & Sons, New York, 1974.
- [11] 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.
- [12] S. F. Gull. Developments in maximum-entropy data analysis. In J. Skilling, editor, *Maximum Entropy and Bayesian Methods*, pages 53–71. Kluwer Academic, 1989.

- [13] R. Szeliski. Probabilistic modeling of surfaces. *Proc. SPIE*, 1570:154–165, 1991.
- [14] R. Szeliski and D. Terzopoulos. Physically-based and probabilistic models for computer vision. *Proc. SPIE*, 1570:140–152, 1991.
- [15] R. Bajcsy and S. Kovačič. Multiresolution elastic matching. *Comput. Vision*, 46:1–21, 1989.
- [16] 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.