Image Feature Identification via Bayesian Hierarchical Models

Colin McCulloch, Jacob Laading, and Valen Johnson
Institute of Statistics and Decision Sciences, Duke University
Colin McCulloch, Box 90251, Duke University, Durham, NC 27708

Key Words: deformable template models, hierarchical normal models, scale-space

Abstract:

A probability density function is proposed to model the task of feature identification in an observed image taken from a class, given a template image for the class. A large number of locations, called "facets", are defined on several grids in the template image and feature identification is accomplished by locating all facets in the observed image. The model is based on a particularly simple hierarchical normal form which enables fast calculation of mode estimates and Monte Carlo sampling. The method is demonstrated on three-dimensional human brain magnetic resonance images.

1 Introduction

The aim of this research is to model the task of feature identification within a class of images given a template image representative of the class. Our goal is to construct a probabilistic model, given the template image and an observed image, whose mode matches the positions of features found in the template with the corresponding positions in the observed image. This paper outlines such a model and illustrates its capabilities in a preliminary set of experiments matching two three-dimensional brain magnetic resonance images. The predictive distribution is maximized and samples are simulated to investigate its shape about the mode.

To begin the construction of the model, a dispersed set of locations are oriented on a grid in the template image. These locations, termed "facets", are numerous enough that any feature in the template image can be represented by one or a set of them. The notion of sufficient facet density is made more explicit in Section 2. Facets are similar to landmarks as defined by Bookstein [1] except that a human observer has not assigned each facet to represent an anatomical structure. A facet represents an image feature once a human observer has associated its predetermined location in the template image with a feature in that image. A facet also differs from a Bookstein landmark in that it is defined to be one random variable in a probability density function (pdf) hierarchically defined on a common tree of facets. The random variables in the facet tree represent each individual facet's position in the template image and its corresponding position in a given observed image from the class. The pdf proposed in this paper is a joint distribution on all facets' locations in an observed image from the class, given their positions in a template image.

Facet template locations are oriented in the template image on several grids of varying densities. Each grid is one level in the facet tree, and the levels are ordered in increasing density so that the most dense grid is the bottom level of the tree. In the model formulation, the bottom level of facets represents the image on its original scale and higher levels can be loosely regarded as lower resolution representations of the image.

The unnormalized joint pdf on facet locations in an observed image from the class given their locations in the template image is the product of two functions. The first is a hierarchical normal model on facet locations that assigns a probability density value to possible facet locations in any observed image from the class, given their locations in the template. This normal model is formulated as a series of regressions of each child facet's predicted position on its parent's predicted position. The model provides that, in expectation, the template parent/child spatial relationships will be preserved in an observed image.

The second term in the product relates the observed image to the template through the facet locations in both and is a function of only the bottom (most numerous) level of the facet tree. This term is a product over all bottom level facets of independent normal distributions on the observed image Laplacian [7] at a facet's location with mean equal to the facet's template image Laplacian.

Maximization of the joint density on the facet tree proceeds hierarchically starting at the top of the tree

and moving down by maximizing a constrained full conditional density on sub-trees of decreasing size given the positions of all facets not contained in the sub-tree. Proceeding in this manner, we address global characteristics of the high-dimensional density before turning to its local characteristics.

In this paper we present a formal definition of the model along with our method for maximizing the distribution. We apply the method to achieve point estimates of feature identifications in a three-dimensional human brain magnetic resonance image (MRI) using another MRI as template and we simulate from the pdf using a Gibbs/Metropolis sampling algorithm to explore its marginal moments.

2 The Model

Facets are located in the template image on several grids of varying densities that are connected to form a tree structure. The children of each facet in the tree are those on the next lowest level located closest to the parent in the template image.

The highest resolution (bottom) level contains a sufficient number of facets to represent salient features of the image on its original scale. To achieve this, the template and observed image Laplacians are convolved with a Gaussian kernel of standard deviation slightly larger than the facet spacing on the bottom level of the tree. The convolution is a low-pass filter with threshold frequency approximately equal to the inverse of its standard deviation. Therefore, the spatial sampling rate on the bottom level of the tree is above the Nyquist rate [8].

Facet template locations are fixed in the template image. The model is a probability density function on all facet locations in observed images from the class. Let the observed location of facet i on level j of the tree be denoted x_j^i, the vector of locations on level j be x_j, and the locations of all facets be x. Template facet locations are denoted μ. The unnormalized density on facet locations is denoted by $p(x|k,\tau^2)$, and is the product of two densities, $p_S(x|k)$ and $p_I(x|\tau^2)$.

The "shape" density, $p_S(x|k)$, is a hierarchical normal distribution made up of components conveying that a facet's observed location is normally distributed with mean equal to its template location plus its parent's displacement from the parent's template position. Let d be the spacing of the highest resolution facet grid in the template image. For $T+1$ levels in the facet tree $(0,\ldots,T)$, the density p_S can be expressed

$$\begin{aligned} p_S(x_0|k) &= \mathrm{N}(\mu_0, k(2^T d)^2) \\ p_S(x_1^i|x_0;k) &= \text{ind } \mathrm{N}(\mu_1^i + x_0 - \mu_0, k(2^{T-1}d)^2) \\ p_S(x_2^i|x_1;k) &= \text{ind } \mathrm{N}(\mu_2^i + x_1^{p_2^i} - \mu_1^{p_2^i}, k(2^{T-2}d)^2) \\ &\vdots \\ p_S(x_T^i|x_{T-1};k) &= \text{ind } \mathrm{N}(\mu_T^i + x_{T-1}^{p_T^i} - \mu_{T-1}^{p_T^i}, kd^2) \end{aligned} \tag{1}$$

where p_j^i is the index of the parent of facet $(\cdot)_j^i$.

The kernel of the "image" density, $p_I(x|\tau^2)$, is a product of independent normal densities on the observed image Laplacian of the lowest level facets with mean equal to each facet's template image Laplacian. (The image Laplacian is used in this term rather than intensity to remove intensity trends of order less than two, and the observed and template Laplacians are rescaled so that their 10th and 90th quantiles match.) For $T+1$ levels $(0,\ldots,T)$, the image contribution to the density on facet locations is

$$\begin{aligned} p_I(x|\tau^2) &= p_I(x_T|\tau^2) \\ &\propto \exp\left\{-\frac{1}{2\tau^2}\sum \left(L_x(x_T^i) - L_\mu(\mu_T^i)\right)^2\right\} \end{aligned} \tag{2}$$

where $L_x(x_T^i)$ and $L_\mu(\mu_T^i)$ are the Laplacians of the observed and template images at positions x_T^i and μ_T^i, respectively. This distribution is regarded as a function of x, and therefore we neglect an unknown normalizing constant not involving x.

The full conditional distribution of facet locations in the observed image is therefore

$$\begin{aligned} &p(x|k,\tau^2;\mu,L_\mu,L_x) \\ &\quad = \frac{1}{Z} p_S(x|k;\mu) p_I(x|\tau^2;\mu,L_\mu,L_x) \end{aligned} \tag{3}$$

where $Z = Z(k,\tau^2,\mu,L_\mu,L_x)$ is a normalizing constant independent of x.

3 Density Maximization and Simulation

For given k and τ^2, the conditional density $p(x|k,\tau^2)$ is maximized by proceeding through the facet tree from top to bottom. At each facet the full conditional distribution on the sub-tree below it is maximized under the constraint that the sub-tree retains its expected inner shape relationships under $p_S(x|k)$. In other words, during numerical maximization the sub-tree is perturbed as a whole unit in three-dimensions. Upon perturbation, the change in

$p(x|k,\tau^2)$ depends on two values: (1) the sub-tree's top facet's spatial deviation from its parent, and (2) the sum of squared Laplacian differences over the bottom level facets in the sub-tree.

When a level of sub-tree tops is complete, the full conditional distribution for every facet above the current level can be maximized analytically since these distributions are all normal. A small number of ICM [3] cycles through the conditional distributions of the current level and those above are undertaken before including the next level of facets in the ICM cycle.

As noted above, numerical maximization is required for the lowest level of facets included in an ICM cycle. Conditional on the locations of facets above this level, the (constrained) distributions on sub-trees whose top facets lie on this level are independent. The joint density on all sub-trees with top facets on this level can therefore be maximized by considering each sub-tree's individual conditional distribution separately. The Nelder-Mead simplex method [6] is used to maximize these trivariate densities.

This estimate of the mode of the density can be used as a starting point for iterative sampling. Gibbs sampling is easily implemented on facet locations above the bottom level of the tree since these full conditional distributions are normal. For each facet on the bottom level the log full conditional distribution on its position contains a squared difference between observed and template image Laplacians. This results in a non-standard form which must be simulated using a Metropolis step.

4 Results

We now illustrate the model on a pair of three-dimensional T1 weighted human brain magnetic resonance images. One image was treated as the template for this class and the other was treated as observed from the class. The field of view was 20cm×20cm×20cm and the image volume was recorded as 60 slices of 256×256 pixel images. The voxel dimensions were therefore .78mm×.78mm×3mm. Both images were convolved with a three-dimensional spherical Gaussian kernel with standard deviation 3.9 mm, and L_x and L_μ were set to the Laplacians of the convolved images.

The facet tree used in this example had 4 levels with each parent facet having 64 children ($4 \times 4 \times 4$) for a total of 266,305 facets. Facets were evenly dispersed throughout the entire image cube so that bottom level facets were located at 3.1 mm intervals. This sampling rate is slightly above the spatial Nyquist frequency implied by the convolved Laplacian images.

The ratio τ^2/k was adjusted to reflect our beliefs about the relative importance of the shape constraint enforced by the shape density (eq. 1) with respect to the shape change driven by the image distribution (eq. 2). Heuristically, the ratio was set to a value that caused, during maximization, a relative weighting between shape log density variation and image log density variation of approximately 1:20. Three ICM cycles were performed above each level in the tree as described in Section 3.

A subset of the facets is shown in figure 1. The figure shows slice 30 from the template image and slices 29 and 31 from the observed image. The facets in three regions are highlighted to showcase the model's three-dimensional capabilities. It has been noted by visual inspection that the head of caudate in the observed image is positioned roughly 1 slice higher when compared with its position in the template image, and the model does predict that observed "head of caudate" facets should be in that slice. Note also that the template grid structure allows us to see the more subtle shape variation modeled by the probability density. In another part of the brain the highlighted sulcus appears 1 slice lower in the observed image than in the template image. The model predicts this shape deformation accurately.

Upon completion of estimating the maximum density facet positions, a Gibbs/Metropolis algorithm was applied to sample from the distribution $p(x|k,\tau^2)$. In this preliminary experiment the parameters (k,τ^2) were set by inspection of the resulting marginal distributions on several facets' predicted locations in the observed image. A set of 1000 iterations were discarded to reduce the effect of the starting point and 1000 samples were recorded. Figure 2 shows a summary of the sampling behavior of 8 facets chosen to be in template slice 24. The top panel shows their fixed template positions, and the middle and bottom panels display normal kernel density estimates of the marginal distributions of the predicted facet positions in the observed image. Note that these marginal density estimates are three-dimensional; those displayed are marginalized again to lie in the x/y plane and are displayed in the highest probability slice. Half maximum density contours are shown. Note that any deviation from normality in these marginal distributions is attributable to the contribution of the image distribution $p_I(x|\tau^2)$ so that regions inside the contours have Laplacian values similar to each respective facet's template value. Contours are therefore elongated along Laplacian isocurves.

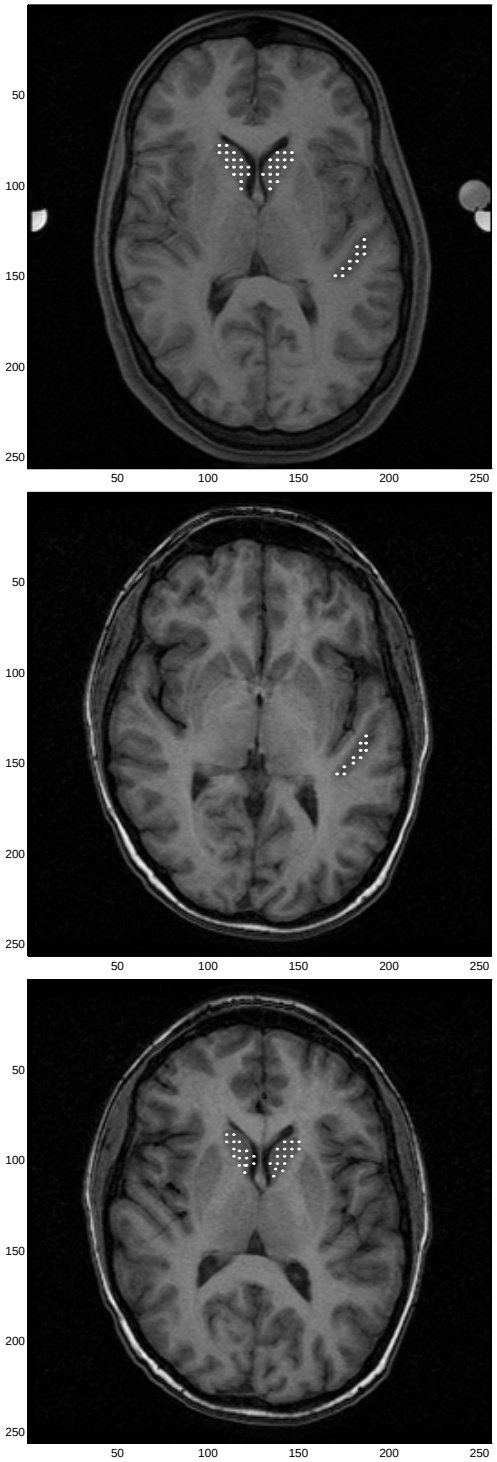

Figure 1: A subset of facets whose template positions are in slice 30. The top panel is slice 30 of the template and the middle and bottom panels are slices 29 and 31 of the observed image, respectively. Facets in the middle and bottom panels are shown at their maximum density predicted locations.

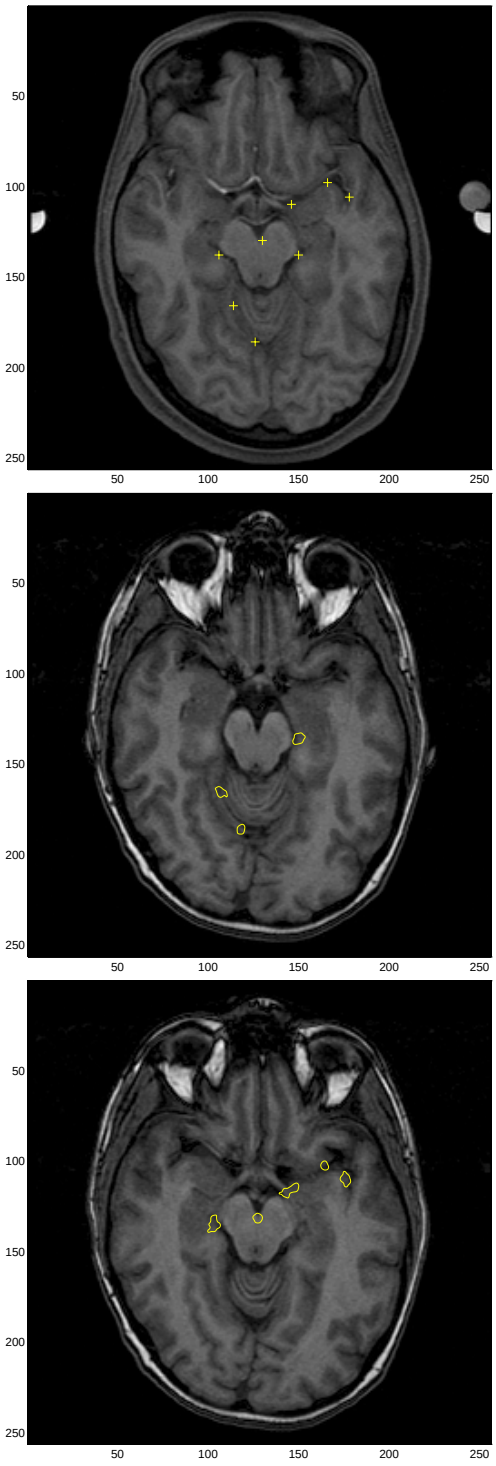

Figure 2: A subset of facets whose template positions are in slice 24. The top panel is slice 24 of the template and the middle and bottom panels are slices 24 and 25 of the observed image, respectively. Monte Carlo samples were used to predicted kernel density estimated regions of highest probability in the marginal distributions on each facet's location.

Density maximization required approximately 1/2 hour of CPU time on a mid-price UNIX workstation and simulation required approximately 30 seconds per sample.

5 Conclusions

We have presented a model for feature identification in images from a class using a template image for the class. The simple form of the density and extensive conditional independence allow for fast maximization and enable straightforward sampling for the investigation of its moments.

In this paper the distribution was maximized to give point estimates of feature identifications and sampled to give estimated uncertainties in predicted facet locations in a three-dimensional human brain MRI. Excellent results were seen with empirically chosen hyperparameters.

Further research into this model will be toward a coherent theory of parameter estimation. If data were available in the form of manual point feature identifications in a template and several observed images from a class, then parameter estimation for that class would be reasonably straightforward [5]. However, in the clinical setting, data is more likely to be found in the form of region segmentations. This will require another level of modeling to address the variation in a particular region's shape given the predicted locations of facets in that region. Mathematical morphology is currently being investigated as a tool for implementing this part of the model. Through this addition to the model we are taking steps toward the automation of clinically useful high-level image analysis tasks such as image segmentation and volume calculation.

Future research will also address several generalizations of the model. First, template structure of the facet tree need not be generated by the method of Section 2. New trees will be constructed using the template image itself to explore different shape constraints on predicted facet locations [4]. Also, the structure of $p_S(x)$ will be generalized to have a conditional independence structure representable by a graph rather than a tree. By having more than one parent for each facet, spatially extended correlation structures between facets can be explored while keeping a high level of conditional independence in the model. Finally, the form of $p_I(x)$ will be generalized to include several other image functions rather than the Laplacian, for instance the boundariness [7] and the correlation measure used in Collins [2].

References

[1] F. Bookstein. *Morphometric Tools for Landmark Data*. Cambridge University Press, Cambridge, 1991.

[2] D. Collins, C. Holmes, T. Peters, and A. Evans. Automatic 3-d model-based neuroanatomical segmentation. *Human Brain Mapping*, 3:190–208, 1995.

[3] S. Geman and D. McClure. Methods for tomographic image reconstruction. In *Proceedings of the 46th Session of the International Statistics Institute Bulletin ISI 52*, 1987.

[4] J. K. Laading, C. McCulloch, and V. Johnson. A hierarchical object deformation model applied to the digital chest radiograph. In *The American Statistical Association Proceedings of the Section on Bayesian Statistical Science*, Anaheim, California, 1997.

[5] C. McCulloch, J. Laading, A. Wilson, and V. Johnson. A shape-based framework for automated image segmentation. In *The American Statistical Association Proceedings of the Section on Bayesian Statistical Science*, pages 1–6, Chicago, Illinois, 1996.

[6] J. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7:308–313, 1965.

[7] B. M. terHaar Romeny, L. Florack, J. Koenderink, and M. Viergever. Scale-space: its natural operators and differential invariants. In *Lecture Notes in Computer Science*, number 511, pages 239–255, Berlin, Germany, 1991. Springer-Verlag.

[8] A. M. Yaglom. *An Introduction to the theory of Stationary Random Functions*. Prentice-Hall, Englewood Cliffs, New Jersey, 1962.