

# Atlas Stratification\*

Daniel J. Blezek and James V. Miller

GE Research, Niskayuna, NY 12309, USA  
{blezek, millerjv}@research.ge.com

**Abstract.** The process of constructing an atlas typically involves selecting one individual from a sample on which to base or root the atlas. If the individual selected is far from the population mean, then the resulting atlas is biased towards this individual. This, in turn, can bias any inferences made with the atlas. Unbiased atlas construction addresses this issue by either basing the atlas on the individual which is the median of the sample or by an iterative technique whereby the atlas converges to the unknown population mean. In this paper, we explore the question of whether a single atlas is appropriate for a given sample or whether there is sufficient image based evidence from which we can infer multiple atlases, each constructed from a subset of the data. We refer to this process as *atlas stratification*. Essentially, we determine whether the sample, and hence the population, is multi-modal and is best represented by an atlas per mode. We use the mean shift algorithm to identify the modes of the sample and multidimensional scaling to visualize the clustering process.

## 1 Introduction

Atlas-based techniques have many applications in medical image analysis. Atlases take on many forms, ranging from an intensity image of the *average* subject to more detailed shape, intensity and functional models of specific structures. Atlases are used in basic research on population analysis, as guides in gross segmentation and seed point selection, as context in navigation tasks, and as models to overcome signal limitations and indistinct boundaries. Atlases may be based on a single individual or on a sample of a population. Atlases can be deterministic, where each region of space is associated with a single structure, or atlases can be probabilistic, where each region of space is assigned a likelihood of belonging to a variety of structures.

When atlases are constructed from a sample of a population, the imagery for the subjects in the sample are transformed into a common coordinate frame prior to consolidating their information. This step of *rooting* the atlas is common to both deterministic and probabilistic atlas construction. Establishing this common coordinate frame is a critical step that impacts the quality of the resulting atlas. A common practice is to select one subject from the sample on

---

\* This work is part of the National Alliance for Medical Image Computing (NAMIC), funded by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant U54 EB005149. Information on the National Centers for Biomedical Computing can be obtained from <http://nihroadmap.nih.gov/bioinformatics>.

which to base the atlas. But if the selected subject is far from the population mean, the resulting atlas will be biased towards this individual. This, in turn, can bias any inferences made with the atlas. This issue has led to recent interest in *unbiased* atlas construction. Unbiased atlases can be constructed by searching for the *subject* closest to the population mean [1, 2] and rooting the atlas on that subject, or by searching for the *common coordinate frame* in the center of the population [3, 4, 5, 6, 7] and rooting the atlas on that coordinate frame.

Current atlas construction techniques are based on an assumption that the population is best described by a single atlas, treating the population as unimodal after transformation to the common coordinate frame. While this transformation may be non-rigid, and may therefore normalize away a portion of the inter-subject variability, substantial inter-subject variability may remain. Studying this remaining variability is the subject of population analysis. However, this same variability may render an atlas ineffective when used as a prior to combat signal limitations and indistinct boundaries. For these applications, variations beyond unimodal variations are particularly troubling.

In this paper, we explore the question of whether a population is best described by a single atlas or whether there is sufficient evidence to infer multiple atlases, each constructed from a subset of the data. We refer to this process as *atlas stratification*. We identify the modes in the population using a mean shift algorithm [8]. Each mode represents a subspace of the population which requires a unique atlas. In the process of identifying the modes, we determine which subjects should be used in constructing the atlas for each mode. The stratification process iterates between performing pairwise registrations of subjects and constructing atlases from subsets of subjects.

## 2 Mean Shift

Fukunaga and Hostetler introduced the mean shift algorithm [8] to estimate the gradient of a probability density function given a set of samples from the distribution. Using a hill climbing algorithm, this gradient estimate can be used to identify the modes of the underlying distribution. The mean shift algorithm has been used for clustering [8, 9], segmentation [10], and tracking [11].

Following the notation and derivation in [8], let  $X_1, X_2, \dots, X_N$  be a set of  $N$  iid.  $n$ -dimensional random vectors. The kernel density estimate of the underlying distribution is

$$\hat{p}_N(X) = (Nh^n)^{-1} \sum_{j=1..N} k\left(\frac{1}{h}(X - X_j)\right) \quad (1)$$

where  $k(X)$  is a scalar function satisfying the requirements for a kernel [12] and  $h$  is a parameter often referred to as the bandwidth [12]. If  $k(X)$  is a differentiable function, the gradient of the density estimate is

$$\hat{\nabla}_x p_N(X) = (Nh^n)^{-1} \sum_{j=1..N} \nabla_x k\left(\frac{1}{h}(X - X_j)\right) \quad (2)$$

where  $\nabla_x$  is the gradient operator with respect to  $x_1, x_2, \dots, x_n$ . A simple kernel of the form

$$k(X) = \begin{cases} c(1 - X^T X) & \text{if } X^T X \leq 1 \\ 0 & \text{if } X^T X > 1, \end{cases} \tag{3}$$

where  $c$  is a normalizing constant chosen to make the kernel integrate to one, satisfies the conditions for the density estimate to be asymptotically unbiased, consistent, and uniformly consistent [8]. Substituting this kernel into (2) yields

$$\widehat{\nabla}_x p_N(X) = (Nh^{n+2})^{-1} 2c \sum_{X_i \in S_h(X)} (X_i - X) \tag{4}$$

where  $S_h(X)$  is a neighborhood with a radius equal to the bandwidth,  $h$ .

When  $S_h(X)$  is small,  $p_N(x)$  over the restricted domain of  $S_h(X)$  is approximately uniform. The terms prior to the summation in (4) can be shown to be proportional to the density of an  $n$ -dimensional uniform distribution over  $S_h(X)$ . Therefore, we can approximate the normalized gradient (see [8] for details)

$$\frac{\widehat{\nabla}_x p_N(X)}{p_N(x)} \approx \frac{n+2}{h^2} M_h(X) \tag{5}$$

where

$$M_h(X) = \frac{1}{k} \sum_{X_i \in S_h(X)} (X_i - X)$$

is referred to as the sample mean shift, or simply the mean shift, and  $k$  is the number of samples in  $S_h(X)$ .

We can use this approximation to the normalized gradient to cluster samples  $X_j, j = 1, 2, \dots, N$ , using the update equations

$$X_j^0 = X_j, \tag{6}$$

$$X_j^{t+1} = X_j^t + a \frac{\nabla_x p_N(X_j^t)}{p_N(X_j^t)}. \tag{7}$$

Using (5) and setting  $a = \frac{h^2}{n+2}$ , yields a simplified update equation

$$X_j^{t+1} = X_j^t + M_h(X_j). \tag{8}$$

This derivation of the mean shift is a  $k$ -nearest neighbor formulation, where the distance to the  $k^{th}$  nearest neighbor defines the bandwidth  $h$ .

### 3 Atlas Stratification

To apply the mean shift algorithm to the problem of atlas stratification, we consider the images for a subject as one sample. As such, each sample sits in a very high dimensional space (rows  $\times$  cols  $\times$  slices). We drop the standard

$L_2$  norm, used by the original mean shift algorithm to establish the nearest neighbor set, in favor of an information theoretic measure, specifically mutual information [13, 14, 15]. Other metrics and measures for comparing subjects for atlas stratification are possible. For instance, overlap metrics or shape similarity metrics applied to presegmented structures could be used in the mean shift framework.

Labeling the images for a subject as  $I_j$  and treating the  $N$  subjects as samples, let

$$I_j^0 = I_j \quad j = 1, 2, \dots, N.$$

At each iteration  $t$ , we align each pair of subjects using a Mattes mutual information registration algorithm [16]. The transformations that align each pair of subjects is  $T_{ij}^t$ . The mutual information values for each pair of subjects is  $MI(T_{ij}^t \circ I_i^t, I_j^t)$ . The nearest neighbor set for the mean shift iteration is the set

$$S_{d_k}(I_j^t) = \{I_i^t : MI(T_{ij}^t \circ I_i^t, I_j^t) \leq d_k\}$$

where  $d_k$  is  $k^{th}$  largest mutual information value to the image  $I_j^t$ . The mean shift is defined here as

$$M_h(I_j^t) = \frac{1}{k} \sum_{T_{ij}^t \circ I_i^t \in S_h(I_j^t)} (T_{ij}^t \circ I_i^t - I_j^t). \quad (9)$$

The samples are updated

$$I_j^{t+1} = I_j^t + M_h(I_j^t) \quad (10)$$

$$= I_j^t + \frac{1}{k} \sum_{T_{ij}^t \circ I_i^t \in S_h(I_j^t)} (T_{ij}^t \circ I_i^t - I_j^t) \quad (11)$$

where after each mean shift iteration,  $t = t + 1$ , the pairwise registrations are repeated using  $I_i^t$  and  $I_j^t$  from the previous iteration, producing new transformations  $T_{ij}^t$  and mutual information metric values  $MI(T_{ij}^t \circ I_i^t, I_j^t)$ . These transformations and metric values define the nearest neighbor sets for the next mean shift iteration. As the iterations of registrations and mean shift progress, the samples converge to the modes of the population.

## 4 Data

Our data was a random selection of 222 MR scans from the High Field MRI Studies of Neurodegenerative Disease conducted at the Albany Medical College's Neuroimaging Center. The scans were acquired on a 3T scanner (GE Medical Systems, Milwaukee WI). Mean age of the subjects was 74 years and ranged from 55-90 years. The scans were SPGR T1 weighted acquisitions with 15 deg flip angle, 12.1/5.2 TR/TE, 22 cm FOV, 2 mm slice thickness. In each scan, 96 coronal slices were acquired.

## 5 Results

The atlas stratification algorithm was implemented using the Insight Segmentation and Registration Toolkit [17]. Twenty iterations of atlas stratification were performed using mean shift bandwidths ( $k$ ) of 7, 15, and 30 neighbors. At each iteration, 222x221 or 49062 volume registrations were performed followed by 222 averages of  $k$  volumes. The registrations were limited to estimating affine transformations. In total, the processing comprised over 3 million registrations and was distributed over a cluster of 500 computers.

### 5.1 Multidimensional Scaling

Multidimensional scaling is a cluster analysis technique that constructs a low-dimensional representation of a set of high dimensional samples given just the pairwise intersample distances [18]. Multidimensional scaling has previously been used in atlas construction by Park *et al* [2] to identify a subject close to the geometric mean of the population and rooting their atlas on that subject. Here, we use multi-dimensional scaling as a qualitative tool to visualize the mean shift clustering process. Figure 1 shows the results of multidimensional scaling applied to the intersubject distances after mean shift iterations 1, 2, 4, and 20 and for mean shift bandwidths of 7, 15, and 30 neighbors. Subjects in the same cluster at the last iteration are labeled with the same symbol. These symbol assignments are propagated back to earlier iterations to illustrate the clustering process. Note that multidimensional scaling produces a representation unique up to rotation/flip/permutation. Therefore, the visualizations across the iterations or down the bandwidth may vary in configuration by a rotation/flip/permutation.

To apply multidimensional scaling to the results of the mean shift algorithm, the pairwise mutual information metric values are converted to distances

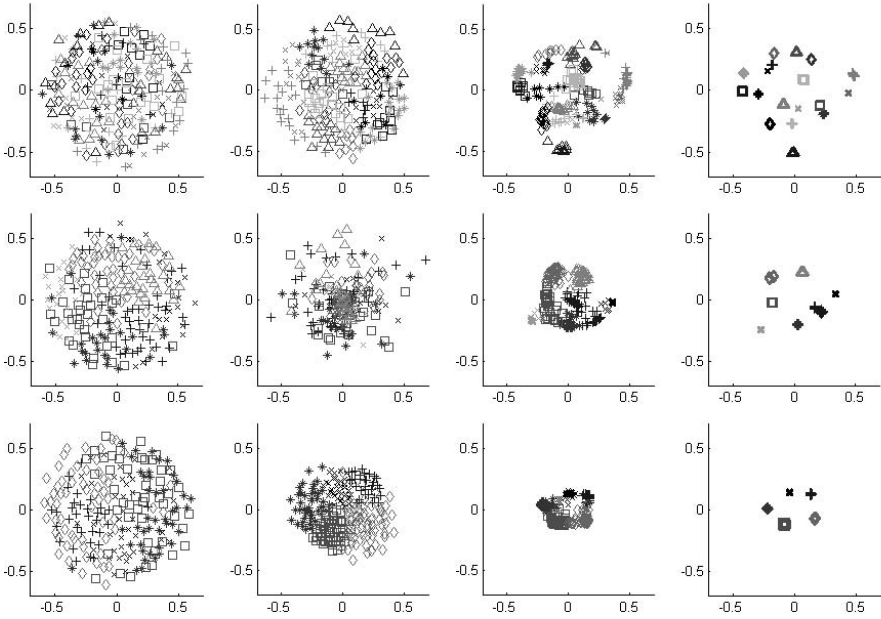
$$D(I_i, I_j) = \alpha \left( \max_{I_k \in \mathcal{S}} MI(I_k, I_j) \right) - MI(I_i, I_j) \quad (12)$$

where  $\alpha$  is a small scale factor near unity to keep the distances distinct from zero. Since the mutual information values are the result of image registration, and these registrations are based on random sampling and gradient descent optimization, the mutual information values are not necessarily commutative. To combat this, the distance matrix  $D(I_i, I_j)$  is made symmetric simply by

$$D(I_i, I_j) = (D(I_i, I_j) + D(I_j, I_i)) / 2.0 \quad (13)$$

and by prescribing the diagonal to be zero. These approximations are reasonable since the multidimensional scaling is only used to provide a qualitative assessment of the results of the mean shift algorithm.

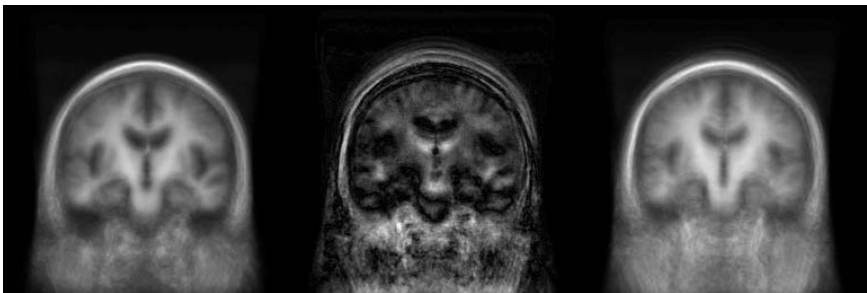
Figure 1 shows the mean shift algorithm identifies multiple modes in the population. The number of modes being a function of the mean shift bandwidth. For a bandwidth of 30 neighbors, the mean shift algorithm produces 5 clusters containing 33, 39, 43, 47, and 60 subjects.



**Fig. 1.** Multidimensional scaling is an effective visualization tool for illustrating the progress of the mean shift clustering algorithm applied to the entire image database. The columns show multidimensional scaling applied to the output of mean shift iterations 1, 2, 4, and 20. The rows illustrate the effect of the mean shift bandwidth for nearest neighbor sizes 7, 15, and 30.

## 5.2 Atlas Modes

Figure 2 compares the atlases from two of the modes for a bandwidth of 30 neighbors. The atlas on the left is from a mode describing 43 subjects. The atlas on the right is from a mode describing 60 subjects. In each case, the atlas was



**Fig. 2.** Atlases from two modes (left, right) identified using a bandwidth of 30 neighbors and the difference between these two atlases (center)

constructed using only the 30 neighbors in the mean shift bandwidth. There are two interesting features to note. First, the atlas on the left has larger ventricles. However, the stratification process employed no explicit information on brain morphometry. Second, the atlas on the left is crisper. From this we infer subjects within this mode have less inter-subject variability than the subjects in the mode described by the atlas on the right (recall the same bandwidth, 30, was used to construct each atlas).

## 6 Conclusions

In this paper, we investigate *atlas stratification*, questioning whether a single atlas is appropriate for a given sample or whether there is evidence from which we can infer multiple atlases, each constructed from a subset of the data. We use the mean shift algorithm to search for modes in the population. If a population has multiple modes, the population should be described by multiple atlases to minimize bias. Our algorithm alternates between performing pairwise registrations of subjects and constructing atlases from subsets of the subjects. As the iterations of registrations and mean shift progress, the subjects converge to the modes of the population.

While the approach taken here is direct, it is not the only possible construction. For instance, the distance metric does not have to be based on mutual information. Overlap metrics or shape similarity metrics on presegmented structures could also be used in this mean shift framework. While our studies were based on an affine transform between subjects, higher order transformations and deformable registrations could be used. Mean shift formulations other than the nearest neighbor variant could also be used.

Aside from the above refinements, we've identified two areas of future research for atlas stratification. The first is a study of the algorithm itself, quantifying the differences between the atlases produced by atlas stratification. The second is a study of the algorithm in context, quantify an improvement in an atlas-based technique when multiple atlases are available. The latter will require a method to select the most appropriate atlas for a particular subject.

## Acknowledgments

We would like to note our appreciation for the considerable efforts in collecting and managing the imagery used in this paper. Dr. John Schenck from GE Research and Dr. Earl Zimmerman from the Albany Medical College co-directed the collaboration on High Field MRI Studies of Neurodegenerative Disease as well as recruited volunteers for the study. Dave Henderson from GE Research provided data management and John Cowan from the Albany Medical College performed the subject scans.

## References

1. S. Marsland, C. Twining, and C. Taylor. Groupwise non-rigid registration using polyharmonic clamped-plate splines. In *MICCAI*, pages 771–779, 2003.
2. H. Park, P.H. Bland, A.O. Hero, and C.R. Meyer. Least biased target selection in probabilistic atlas construction. In *MICCAI*, pages 419–426, 2005.
3. C. Studholme and V. Cardenas. A template free approach to volumetric spatial normalization of brain anatomy. *Pattern Recogn. Lett.*, 25(10):1191–1202, 2004.
4. K.K. Bhatia, J.V. Hajnal, B.K. Puri, A.D. Edwards, and D. Rueckert. Consistent groupwise non-rigid registration for atlas construction. In *ISBI*, pages 908–911, 2004.
5. M. De Craene, A. du Bois d’Aische, B. Macq, and S.K. Warfield. Multi-subject registration for unbiased statistical atlas construction. In *MICCAI*, pages 655–662, 2004.
6. P. Lorenzen, B. Davis, and S.C. Joshi. Unbiased atlas formation via large deformations metric mapping. In *MICCAI*, pages 411–418, 2005.
7. L. Zollei, E. Learned-Miller, E. Grimson, and W. Wells. Efficient population registration of 3d data. In *ICCV 2005*, 2005.
8. K. Fukunaga and L.D. Hostetler. The estimation of the gradient of a density function with applications in pattern recognition. *T-IT*, 21(1):32–40, 1975.
9. Y. Cheng. Mean shift, mode seeking and clustering. *T-PAMI*, 17(8):790–799, 1995.
10. D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *T-PAMI*, 24(5):603–619, 2002.
11. D. Comaniciu, V. Ramesh, and P. Meer. Real-time tracking of non-rigid objects using mean shift. In *CVPR 2000*, volume 2, pages 142–149, 2000.
12. B.W. Silverman. *Density estimation for statistics and data analysis*. Chapman and Hill, 1992.
13. W. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis. Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1(1):35–51, 1996.
14. A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal. Automated multimodality medical image registration using information theory. In *Proc. 14th Int. Conf. Information Processing in Medical Imaging; Computational Imaging and Vision 3*, pages 263–274, 1995.
15. F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187–198, Apr. 1997.
16. D. Mattes, D.R. Haynor, H. Vesselle, T.K. Lewellen, and W. Eubank. Pet-ct image registration in the chest using free-form deformations. *IEEE Transactions on Medical Imaging*, 22(1):120–128, Jan. 2003.
17. L. Ibanez, W. Schroeder, L. Ng, and J. Cates. *The ITK Software Guide*. Kitware, Inc. ISBN 1-930934-15-7, <http://www.itk.org/ItkSoftwareGuide.pdf>, second edition, 2005.
18. B.S. Everitt, S. Landau, and M. Leese. *Cluster Analysis*. Arnold, 2001.