

# Comparing the Similarity of Statistical Shape Models Using the Bhattacharya Metric

K.O. Babalola<sup>1</sup>, T.F. Cootes<sup>1</sup>, B. Patenaude<sup>2</sup>, A. Rao<sup>3</sup>, and M. Jenkinson<sup>2</sup>

<sup>1</sup> Division of Imaging Science and Biomedical Engineering, University of Manchester,  
M13 9PT, UK

{kola.babalola/tim.cootes}@manchester.ac.uk

<sup>2</sup> FMRIB Centre, John Radcliffe Hospital, University of Oxford, OX3 9DU, UK

{mark/brian}@fmrib.ox.ac.uk

<sup>3</sup> Visual Information Processing, Department of Computing, Imperial College London, SW7  
2BZ, UK

ar17@doc.ic.ac.uk

**Abstract.** A variety of different methods of finding correspondences across sets of images to build statistical shape models have been proposed, each of which is likely to result in a different model. When dealing with large datasets (particularly in 3D), it is difficult to evaluate the quality of the resulting models. However, if the different methods are successfully modelling the true underlying shape variation, the resulting models should be similar. If two different techniques lead to similar models, it suggests that they are indeed approximating the true shape change. In this paper we explore a method of comparing statistical shape models by evaluating the Bhattacharya overlap between their implied shape distributions. We apply the technique to investigate the similarity of three models of the same 3D dataset constructed using different methods.

## 1 Introduction

Statistical Shape Models (SSMs) are parametric representations of the shape and variation in the shape of a class of objects, learnt using a suitably representative training set. They are suited to medical structures as they allow the variation of the shape of the structure(s) being modelled to be captured. This proves useful for model fitting to unseen structures (segmentation), and investigations into the changes in shapes of anatomical structures due to disease (morphometric analysis).

The construction of statistical shape models requires that a correspondence is established across the training set. To construct SSMs of 3D structures, establishing correspondence manually is time consuming and subject to errors. Therefore, there is a burgeoning field of research into automatic methods for obtaining correspondence e.g. [1],[2]. In the case where the shapes of the structures being modelled are represented as points, the statistical shape model is a point distribution model (PDM) [3]. This is a linear model of the  $n$ -dimensional space in which each member of the training set is represented as a vector. Different methods of establishing correspondence for the same training set, represented by the same number of points, will result in different distributions in the  $n$ -dimensional space. Therefore, given the same training set represented by

the same number of points, the different methods of establishing correspondence give rise to different SSMs.

Each such SSM captures information about the true variation in shape of the underlying dataset, but will also be influenced by the method used to obtain correspondences. It is important to be able to understand the true underlying variation in the data, especially for medical applications (where one may be looking for changes due to disease). It is often impractical to manually annotate enough data to construct models, so methods are required to confirm that the models resulting from automatic methods are reliable. One approach is to use two or more different model building methods - if they all result in similar models, one can begin to have confidence that the models are a true representation of the underlying variation. However, this requires a method of comparing the similarity of models which is the subject of this paper. The fact that two models are similar does not necessarily mean they are right, and methods of measuring model quality are still needed.

SSMs have been compared indirectly by evaluating their compactness and specificity [1],[4], and the quality of registration used in model construction has been evaluated using overlap measures [5]. In this paper we propose a method of quantitatively evaluating the similarity between two models. SSMs give compact representations of the probability distribution functions (PDFs) from which training examples are assumed to be drawn. Thus each SSM implies a PDF, and two models can be compared by evaluating the overlap between their implied PDFs. Such a measure gives insight into the different ways of constructing SSMs.

The Bhattacharya metric [6] has been shown to be suitable for comparing the similarity of two probability distributions. Thacker et. al. [7] have explored its application to model similarity in 1D and for parameter estimation. In the following, we describe the Bhattacharya measure and how we apply it to comparing the similarity of SSMs. Its application to 3D SSMs in their standard form is not possible because of memory requirements. An important contribution of this paper is the application of singular value decomposition to reduce the dimensionalities involved. Details of this are given in the Appendix at the end of the paper.

## 2 Bhattacharya Measure and Applicability

### 2.1 Bhattacharya Measure for Overlap of Two $n$ -variate Gaussian Distributions

The Bhattacharya measure for two distributions is given by

$$B = \int \sqrt{p_1(\mathbf{x})p_2(\mathbf{x})} \, d\mathbf{x} \quad (1)$$

The measure gives the overlap for two distributions  $p_1$  and  $p_2$  as a value between 0 (no overlap) and 1 (exact overlap) and is symmetric i.e.  $B(p_1, p_2) = B(p_2, p_1)$ . If  $p_1(\mathbf{x})$  and  $p_2(\mathbf{x})$  are two  $n$ -variate normal distributions with means  $\boldsymbol{\mu}_1$  and  $\boldsymbol{\mu}_2$ , and covariance matrices  $\mathbf{V}_1$  and  $\mathbf{V}_2$ , then

$$p_i(\mathbf{x}) = |\det(\mathbf{V}_i)|^{-\frac{1}{2}} (2\pi)^{-\frac{N}{2}} \times \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^\top \mathbf{V}_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i)\right) \quad (2)$$

The Bhattacharya overlap  $B(p_1, p_2)$  is then

$$B = 2^{\frac{N}{2}} |\det(\mathbf{V}_1)|^{-\frac{1}{4}} |\det(\mathbf{V}_2)|^{-\frac{1}{4}} |\det(\mathbf{V}_1^{-1} + \mathbf{V}_2^{-1})|^{-\frac{1}{2}} \times \exp\left(-\frac{1}{4}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^\top (\mathbf{V}_1 + \mathbf{V}_2)^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\right) \quad (3)$$

## 2.2 Application to PDMs

A PDM is a linear model of the form (see [3])

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{P}\mathbf{b} + \mathbf{r} \quad (4)$$

Where  $\mathbf{x}$  is a *shape vector* formed by concatenation of the coordinates of the points used to represent an instance of the class of shapes being modelled and  $\boldsymbol{\mu}$  is the mean shape vector.  $\mathbf{P}$  is a matrix whose columns are the eigenvectors of the covariance matrix,  $\mathbf{b}$  is a vector of weights allowing instantiation of a particular set of values for  $\mathbf{x}$ , and  $\mathbf{r}$  is a vector of residuals not explained by the model.

If we assume the original data was drawn from a Gaussian distribution, then each  $b_i$  is distributed as a zero mean Gaussian with variance  $\lambda_i$  (the eigenvalue of the original covariance matrix of the training data).

Given a model constructed by applying PCA to the shape vectors of a training set, we can estimate the covariance matrix from the model as

$$\mathbf{V} = \mathbf{P}^T \Delta \mathbf{P} + \sigma_r^2 \mathbf{I} \quad (5)$$

$\Delta$  is a diagonal matrix giving the variance of each parameter  $b_i$  and  $\mathbf{I}$  is a square identity matrix.  $\sigma_r^2$  represents the *variance of residuals*. It is given by:

$$\sigma_r^2 = \sigma_{\gamma_{min}}^2 + \left( \frac{\sum_{i=t+1}^{t_{max}} \lambda_i}{n - t} \right) \quad (6)$$

Each dimension in each of the distributions being compared has an element of noise associated with it. This is expressed by the first term in equation 6 and its value is determined empirically (see section 4.1). In applying PDMs to image search and morphometric analysis, not all the modes of the model are necessarily used as the first few modes usually contain most of the variation. In the case where not all modes are used, the mean of the variance of the modes left out (second term of equation 6) is included in determining the residual variance. In such cases  $t$  represents a *cutoff* in the number of modes of variation of the model, and  $t_{max}$  is the maximum number of modes of variation in the model. The value of  $\sigma_r^2$  should be less than the lowest value of  $b_i$  of the modes used in the comparison.

## 2.3 Implementation Details

Calculating the Bhattacharya overlap involves computing the inverse and determinants of “large” matrices, which require huge amounts of computer memory and can lead to

numerical instability unless handled with care. The size of the covariance matrices are determined by the number of points used to represent the shapes. For  $n$  points in 3D the covariance matrix is of size  $3n \times 3n$ .

To allow application of the Bhattacharya overlap to 3D SSMs we use *singular value decomposition* (SVD) to move the  $3n \times 3n$  covariance matrix into a smaller  $m \times m$  matrix, where  $m \leq 2 \times$  number of training examples (see appendix for details). The Bhattacharya overlap is obtained using the smaller  $m \times m$  matrix. In the present case we reduced our covariance matrices from  $(76, 776 \times 76, 776)$  to  $(74 \times 74)$ . Results from 2D have shown that the values obtained for the Bhattacharya overlap with and without the SVD stage are equivalent.

### 3 Building the Statistical Shape Models

A training set consisting of magnetic resonance images of 37 subjects was used to build the 3D statistical shape models. The dataset contained T1 structural MR scans of the head of each subject at 1mm x 1.5mm x 1mm resolution. For each subject there was a corresponding image in which a number of cortical and subcortical structures had been labelled.

Three methods were used to build 3D SSMs of 10 subcortical structures (lateral ventricles, caudate, putamen, accumbens, pallidum, thalamus, amygdala, hippocampi, brain stem and the fourth ventricle). The first method was based on B-Spline free form deformations (FFD) [2], the second on composition of grid based diffeomorphic warps [8] and the last used a deformable surface mesh [9].

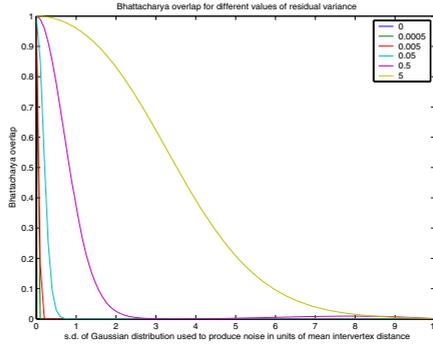
A reference subject was chosen by identifying the member of the training set that was most similar to all other subjects within the set. An initial mesh for each of the labelled structures in the reference subject was obtained using the deformable mesh. This point-set was used to build the SSM for each method. Therefore, there were three SSMs of the same training set built by different methods with the same number of points. The SSM produced by the mesh based method was used to investigate properties of the Bhattacharya metric, then the three models were compared with each other.

## 4 Experiments and Results

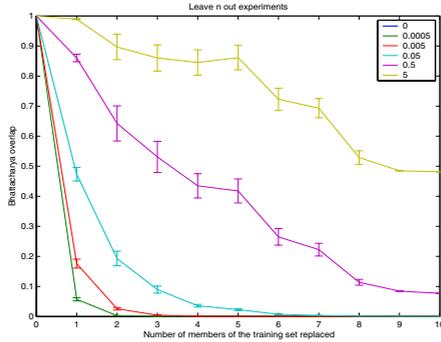
### 4.1 Calibrating the Metric

The Bhattacharya overlap returns a value between 0 and 1. In order to better interpret the values obtained when comparing models, we performed two different experiments to investigate how the metric degrades as model training data is corrupted. The model constructed using the mesh based method was used in these experiments.

In the first experiment random Gaussian noise was added in increasing amounts to vertices of the points of each member of the training set. The random noise was obtained from a Gaussian distribution with standard deviation expressed in fractions of the mean distance between connected vertices. The model constructed from the members of the training set with noise was compared to that constructed from the original training set.



**Fig. 1.** Value of the Bhattacharya overlap as the amount of noise added to a 3D SSM is increased. The amount of noise is expressed in terms of the s.d. of the Gaussian distribution from which the noise values were sampled. The s.d. was in units of the mean distance between the connected vertices of the mesh. The values of  $\sigma_{\gamma_{min}}^2$  for each plot are shown in the legend.



**Fig. 2.** Value of Bhattacharya overlap during leave  $n$  out experiments. The values for  $\sigma_{\gamma_{min}}^2$  are shown in the legend

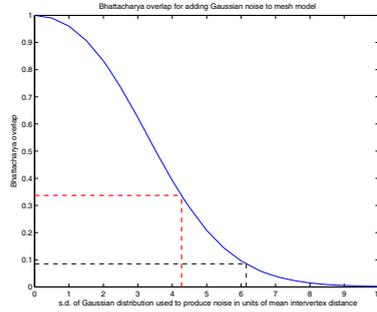
In the second experiment the members of the training set were incrementally replaced (*leave-n-out*). The training set was divided into two parts. 27 members were used to construct a SSM. The members of the 27 member training set were incrementally replaced by randomly picking members from the 10 member training set. The SSM constructed from this new set was compared with that of the original 27 member training set. For each data point 10 random choices from the left out members were used.

Figure 1 shows that the Bhattacharya overlap decreases as the amount of noise added to the model increases. The resulting curve has a sigmoid shape. Figure 2 shows that the Bhattacharya overlap decreases as the number of the 27 member training set replaced increases.

The results from the random noise experiment allows a relationship to be defined between the Bhattacharya overlap and the similarity of two models. For a particular value of the measure, the amount of noise required to achieve a similar difference from the initial distribution can be estimated.

**Table 1.** Values of the Bhattacharya overlap for comparing the first three modes of three SSMs of the same dataset.  $\sigma_{\gamma_{min}}^2$  was set at 5.

Models Compared	Bhattacharya Overlap
FFD vs Compositional Warp	0.166
FFD vs Mesh	0.085
Mesh vs Compositional Warp	0.337



**Fig. 3.** Variation of Bhattacharya overlap with noise for Mesh model. The intercept marked by the red (thick) line shows the overlap with the Compositional warp model is equivalent to adding noise from a Gaussian distribution with s.d. 4.25 times the mean intervertex distance. That for the overlap with the FFD model (6.15 times the intervertex distance) is marked by the thin line.  $\sigma_{\gamma_{min}}^2$  was set at 5.

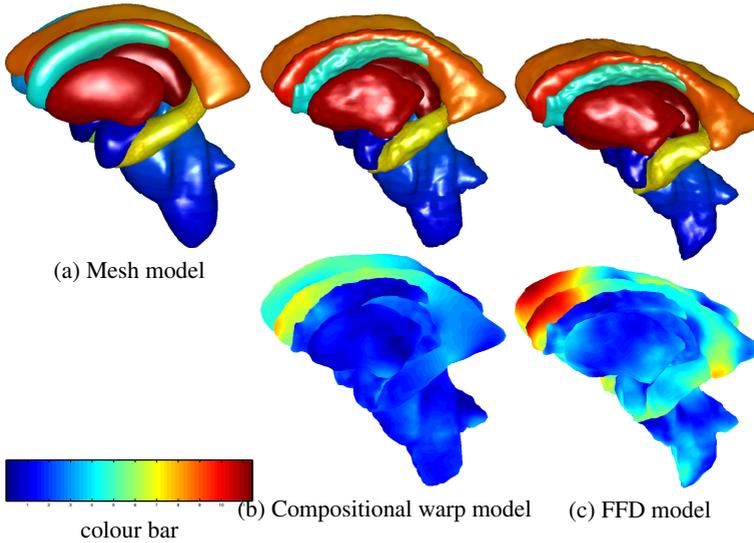
## 4.2 Quantitative Model Similarity Results

Table 1 lists the values obtained for the Bhattacharya overlap when comparing the different SSMs. These values were obtained using  $\sigma_{\gamma_{min}}^2 = 5$ . To give some physical meaning to the values, they are marked on a plot of the Bhattacharya overlap with noise for the Mesh model for  $\sigma_{\gamma_{min}}^2 = 5$  (figure 3).

## 4.3 Qualitative Analysis of Models

The mean shapes of the three models are shown in Figure 4. The figure also shows colour-mapped images of the difference between the mean shapes of the FFD and the compositional warp models relative to the mesh model. This shows that the mean shape of the mesh model is more similar to that of the compositional warp model than it is to the mean shape of the FFD model.

In the PDM equation 4, the eigenvectors contained in  $\mathbf{P}$  each define a *mode of variation*, with eigenvectors corresponding to the largest eigenvalues describing the most significant modes of variation. To visualise the differences captured by the modes, instances of the shape can be generated between limits determined by each  $b_i$ . We instantiated shapes between  $+2.5\sqrt{\lambda_2}$  and  $-2.5\sqrt{\lambda_2}$  of the first three modes of variation for each of the different models. However, we could not draw definite conclusions on which models were more similar by observing these modes of variation.



**Fig. 4.** The mean shapes of the three models. In subfigures (b) and (c) the difference in Euclidean distance between the corresponding points in the Mesh model was used to colour-map the surfaces in the 2<sup>nd</sup> row. The relationship between the colours and the Euclidean distance (in mm) is shown in the colour bar

## 5 Discussion

The results have shown that the Bhattacharya overlap can be used as a measure of the similarity of two SSMs. We have quantified the value of the metric in terms of the amount of Gaussian noise added to a model. The residual variance acts as a regularisation term allowing the sensitivity of the measure to be adjusted.

Results from the permutation experiments suggest that the metric could also be used to evaluate the generalisation ability of a model. This is the ability of the model to adequately represent the training set. We would expect a model that is representative of a class of shapes to be relatively insensitive to the actual members of the training set. However, as the results of figure 2 show this is not the case for the models considered here. This is a common problem in medical applications - basically the numbers in the training set are too small.

We are pursuing several aspects of this work further. Evidently,  $\sigma_{\gamma_{min}}^2$  plays an important role in determining the value obtained for the Bhattacharya overlap. We intend to investigate their relationship further and explore alternative approaches to calibrating the measure. We would also like to investigate decoupling the differences in the mean shapes of the two models from the differences in the modes of variation. Lastly, we intend to further explore the application of the Bhattacharya overlap to measuring the generalisation ability of 3D SSMs.

To conclude, we have presented a novel method of comparing statistical shape models. The method gives us a new approach for evaluating the quality of models built automatically from large datasets, particularly in the medical field.

## Acknowledgements

This work was funded by the EPSRC under the IBIM project. David Kennedy of the Center for Morphometric Analysis, Boston, provided the MR images used.

## References

1. Twining, C.J., et. al.: A unified information-theoretic approach to groupwise NRR and model building. In: LNCS (IPMI 2005 Proceedings). Volume 3565. (2005) 1–14
2. Rueckert, D., Frangi, A.F., Schnabel, J.A.: Automatic construction of 3-D statistical deformation models of the brain using NRR. IEEE TMI **22** (2003) 1014–1025
3. Cootes, T.F., et. al.: ASMs - their training and application. CVIU **61** (1995) 38–59
4. Styner, M., et. al.: Evaluation of 3d correspondence methods for model building. In: LNCS (IPMI 2003 Proceedings). Volume 2732. (2003) 63–75
5. Crum, W.R., Camara, O., Rueckert, D., Bhatia, K.K., Jenkinson, M., Hill, D.L.G.: Generalised overlap measures for assessment of pairwise and groupwise image registration and segmentation. In: (MICCAI proceedings). Volume 3749 of LNCS. (2005) 99–106
6. Bhattacharya, A.: On a measure of divergence between two statistical populations defined by their probability distributions. Bulletin of Calcutta Maths Society **35** (1943) 99–110
7. Thacker, N.A., Aherne, F., Rockett, P.I.: The bhattacharya metric as an absolute similarity measure for frequency coded data. Kybernetika **34** (1997) 363–368
8. Cootes, T.F., Twining, C.J., Taylor, C.J.: Diffeomorphic statistical shape models. In: Proceedings of the 15<sup>th</sup> British Machine Vision Conference, Norwich. (2004) 447–456
9. Patenaude, B., Smith, S., Jenkinson, M.: A bayesian cost function applied to model-based registration of sub-cortical brain structures. In: To appear in LCNS volume on Proceedings of the Workshop on Biomedical Image Registration (WBIR) 2006. (2006)

## Appendix - Using SVD to Decrease Matrix Size

Given two models of training sets consisting of  $s$  training examples each with  $n$  landmarks defined in three dimensions :

$$\mathbf{x}_1 = \boldsymbol{\mu}_1 + \mathbf{P}_1 \mathbf{b}_1 + \mathbf{r} \quad \text{and} \quad \mathbf{x}_2 = \boldsymbol{\mu}_2 + \mathbf{P}_2 \mathbf{b}_2 + \mathbf{r} \quad (7)$$

$\mathbf{x}$  and  $\mathbf{r}$  are column vectors of length  $3n$ ,  $\mathbf{P}$  is of size  $(3n \times s - 1)$ , and  $\mathbf{b}$  is a row vector of length  $(s - 1)$ .

We create a  $(3n \times 2s)$  union space of vectors and perform an SVD:

$$\mathbf{U} \mathbf{W} \mathbf{V}^T = SVD(\boldsymbol{\mu}_1 | \boldsymbol{\mu}_2 | \mathbf{P}_1 | \mathbf{P}_2) \quad (8)$$

$\mathbf{U}$  is a  $(3n \times 2t)$  matrix whose columns form basis vectors in the union space. The components of the models in the union space can be obtained by projecting using  $\mathbf{U}$  :

$$\mathbf{U}^T \mathbf{x}_i = \mathbf{U}^T \boldsymbol{\mu}_i + \mathbf{U}^T \mathbf{P}_i \mathbf{b}_i + \mathbf{U}^T \mathbf{r} \quad \text{and} \quad p_i(\mathbf{x}) = p_i^{(U)}(\mathbf{U}^T \mathbf{x}) \times p_i^{(\bar{U})}(\mathbf{C}^T \mathbf{x}) \quad (9)$$

where  $\mathbf{C}$  is the complement of  $\mathbf{U}$ . Then

$$\begin{aligned}
 B(p_1, p_2) &= \int \sqrt{p_1^{(U)} p_1^{(\bar{U})} p_2^{(U)} p_2^{(\bar{U})}} \, d\mathbf{x} \\
 &= \int \sqrt{p_1^{(U)} p_2^{(U)}} \cdot \sqrt{p_1^{(\bar{U})} p_2^{(\bar{U})}} \, d\mathbf{x} \\
 &= B(p_1^{(U)}, p_2^{(U)}) B(p_1^{(\bar{U})}, p_2^{(\bar{U})})
 \end{aligned} \tag{10}$$

If we assume that the variance of the noise in both models is the same, then the distributions in the null space of  $\mathbf{U}$ ,  $p_1^{(\bar{U})}(\cdot) = p_2^{(\bar{U})}(\cdot)$  and  $B(p_1^{(\bar{U})}, p_2^{(\bar{U})}) = 1$ . Therefore,

$$B(p_1, p_2) = B(p_1^{(U)}(\mathbf{U}^T \mathbf{x}_1), p_2^{(U)}(\mathbf{U}^T \mathbf{x}_2)) \tag{11}$$

and the  $(2s \times s - 1)$  matrix  $\mathbf{U}^T \mathbf{P}$  can be used to estimate a  $(2s \times 2s)$  covariance matrix, instead of using the  $(3n \times s - 1)$  matrix  $\mathbf{P}$  to estimate a  $(3n \times 3n)$  covariance matrix.