# A Minimum Entropy Approach for Automatic Statistical Model Building

Zihua Su, Tryphon Lambrou *MIEEE*, Andrew Todd-Pokropek *MIEEE*

***Abstract*— Statistical shape models have wide applications in medical image analysis both for image segmentation and morphometry. In this paper, inspired by Minimum Description Length (MDL), we developed a novel algorithm for automatic landmark building using an Entropy-based cost-function. The results are tested on four different datasets (metacarpal bones, heads of femur, silhouettes of facial profiles, and hand outlines), and compared with the original MDL constructed using the "one fixed master shape" model approach. It can be seen from our preliminary results that, the new Minimum Entropy Model (MEM) conveys better than the MDL technique on the measures of Generalization Ability, Specificity and Similar Compactness. It also shows good potential in solving the so-called "run away" problem in MDL.**

## I. Introduction

**GEOMETIRC** shape information plays a key role in many computer vision and image processing applications, especially in medical image analysis where many anatomic structures and related functions can be identified and classified in terms of their unique shape. In many applications, we need to analyze the shape of the same structure or object across a group of individuals in order to construct a deformable or dynamic shape model [1]. Therefore, finding a basis of homologous points is a fundamental issue on which relevant work has been published [2][3]. Unfortunately, there is no generally accepted definition for anatomically meaningful correspondence. It's difficult to judge the correctness of an established correspondence. Recently, some inspiring work has been done by Brechbühler [4], Kotcheff [5] and Davies [6]. Especially in [6], the authors proposed an information theory objective function, and established the correct correspondence into the learning process. They also established three objective criteria that can be used to compare models built using different methods. Inspired by that, we propose another framework by using entropy based information theory in order to compose the objective function. From our results on datasets of femur, metacarpal, hand shapes and silhouettes, it can be seen that our method achieves better scores on the Generalization Ability, Specificity and Similar Compactness compared to the original MDL. Further analysis of the silhouette datasets shows that our proposed method, using the proper number of components in the cost function, can, to some extent, solve the so-called "run away" MDL problem, without adding an external term such as curvature or node penalty as suggested in [7].

## II. Related work

The principle approach for finding correspondence is to treat it as an optimization problem, choosing correspondences to optimize an explicit function. This allows for models with well defined properties to be created. Three important developments in automatic correspondence in model building are as follows:

*1)* Hill and Taylor [8] use the trace of the model covariance matrix plus a correction term that penalizes points for moving off the model boundary.
*2)* Kotcheff and Taylor [5] use the determinant of the model covariance matrix as the objective function.
*3)* Davies *et.al.* [6] propose and use information theory to find the correspondence by minimizing the description length.

In particular, Davies *et.al.* work which proposes that the correspondence problem as one of finding the optimal model building and provide a principal framework for statistical models. They also suggest three criteria for modeling assessment that can be used for comparing different models. Although MDL has shown many good properties, it has its own drawbacks. One is that the landmarks can pile up in some areas and therefore can not describe the rest of the shapes, in addition when this happens the algorithm reaches a small but meaningless description length. One way to avoid this "run-away" problem is to select a single shape as master example (as introduced in [6]) on which the marks are not allowed to move. Those landmarks can be placed by manual annotation by an expert. In some cases, as shown in Figure 8, a single fixed master example is not sufficient to keep the whole set in place. For example the free endpoints of open curves can drift systematically to one side or the other, neglecting the master. This is because the statistical weight of the majority can outweigh the single master and the gain of run-away exceeds the cost of a single outlier. More precise details are given in Figure 8. Hans [7] has shown that adding an external term like node penalty or curvature energy can avoid this,

Manuscript received Jun 30, 2006.

Z. Su is a PhD student with the Department of Medical Physics & Bioengineering, University College London, Gower Street, London WC1E 6BT, UK, (phone: +44-020-7679-0287; e-mail: z.su@ medphys.ucl.ac.uk).

T. Lambrou is a Research Fellow with the Department of Medical Physics & Bioengineering, University College London, Gower Street, London WC1E 6BT, UK, (e-mail: tlambrou@medphys.ucl.ac.uk).

A. Todd-Pokropek is a Professor and Head of the Department of Medical Physics & Bioengineering, University College London, Gower Street, London WC1E 6BT, UK, (e-mail: atoddpok@medphys.ucl.ac.uk).

however weighting this term in different datasets could be a problem. Therefore we propose an entropy model, which will achieve similar or better properties on those three evaluation criteria compared with original MDL and also, to some extent, solve the run away problem without using an external term.

## III. Method

### A. Information theoretical techniques

It can be useful to think of finding correspondence as trying to maximize the amount of shared information in all images in data sets. In a qualitative sense, we may say that if two images with correct correspondence are correctly aligned then the corresponding structures will overlap. On the other hand if their correspondence is poor, the images will be out of alignment, in which case, we will have duplicate versions of information from image *A* and *B*.

Using this concept, finding correspondence can be thought of as reducing the amount of information in the combined image, which suggests the use of a measure of information as criteria. The most commonly used measure of information in signal and image processing is the Shannon-Wiener entropy measure *H* [9]

$$H = -\sum_{i=1}^{n} p_i \log p_i \qquad (1)$$

*H* is the average information supplied by a set of *n* symbols whose probabilities are given by $p_1 p_2 p_3, \ldots p_n$. One of the desired properties of Entropy is that it will have a maximum value if all symbols have equal probability of occurring, which is the case when a stack of points pile up in to one location. Though trivial, this observation can, to some extent, solve the so-called run away problem inherently.

### B. Joint Entropy

In finding correspondence, we have two images *A* and *B* to align. We therefore have probabilities from this training set. Joint entropy measures the amount of information we have in the two combined images [9]. The concept of joint entropy can be visualized using the assumption that the probability distribution for every weighting component in Active Shape Model (ASM) [10] is zero centered Gaussian distribution. So, for $i_{th}$ weight on $j_{th}$ component

$$p_i = \frac{1}{\sqrt{2\pi\lambda_j}} e^{-b_i^2 / 2\lambda_j} \qquad (2)$$

### C. Cost function

We propose a method of composing our cost function in a combination of entropy with different assigned weights.

$$Costfunction = \sum_{j=1}^{t} \lambda_j H_j \qquad (3)$$

Where $\lambda_i$ is the $j_{th}$ eigenvalue, *t* is the number of weighting components used in shape model.

### D. Shape representation

We are seeking a set of $2^L + 1$ marks on each curve, where *L* is an integer. For closed shapes, the start and end points are identical. The mark location is specified in a hierarchical manner on *L* levels. For closed curves with 65 marks, we specify on the first level of marks 0 and 32 by their absolute arc length position. On the second level, mark 16 and 48 are specified by parameter between 0 and 1. On the third level the marks 8, 24, 40 and 56 are specified in between already fixed marks. Along this way, until level 6 are finished. For details, we recommend you see [6].

The benefit for doing this is that we can optimize our cost function into any level and set the extra levels to be equaled spaced.

### E. Iterative optimization

The optimization strategy used is as follows. We first initialize all landmarks to be equaled spaced, then realign all nodes, say 8, into a level ascending order and move these nodes according to this order. Every node is given an initial step length 0.01 and moved to probe the cost function descending direction till stabilized. The total number of circles in most of the experiment is 40.

## IV. Experiments results and discussion

For validating this algorithm, experiments were performed on four different datasets, examining different purposes. Data are: contours of metacarpals (Closed curve), femurs (open curve with free ends), silhouettes (Open curve with free ends and run away analysis) and hands (Open curve with fixed ends). The first four experiments show results after applying MEM to these datasets. In each result, 5 levels of marks were placed on the contours and the effects of moving the first three components of the model within the interval of $-3\sqrt{\lambda_i}$ and $3\sqrt{\lambda_i}$ are shown. Quantitative comparisons were performed in experiments five to seven on metacarpals, femurs and hand shapes. The three classic measures of compactness, generalization ability and specificity were estimated using MEM and MDL and their comparisons are shown.

Briefly, generalization ability of a model measures its capability to represent unseen instances of the class of the object modeled. The generalization ability $(G)$, is measured from training sets using leave-one-out reconstruction. The reconstruction error for each model is $\varepsilon_i^2(M)$, and *M* is the number of the components used in the model.

$$G(M) = \frac{1}{n_s} \sum_{i=1}^{n_s} \varepsilon_i^2(M)$$
$$\sigma_{G(M)} = \frac{\sigma}{\sqrt{n_s - 1}} \quad \textbf{(4)}$$

Where $n_s$ is the number of training sets, $\sigma$ is the sample standard deviation of *G(M)*.

Specificity $(S)$ is the ability to measure if the model can generate instances of the object that are close to those in the training set.

$$S(M) = \frac{1}{N} \sum_{j=1}^{N} \left| x_j(M) - x'_j \right|$$
$$\sigma_{S(M)} = \frac{\sigma}{\sqrt{n_s - 1}} \quad \textbf{(5)}$$

Where $x_j$ are shape examples generated by the model (by choosing the value of the weighting components randomly in the range over the training set), $x'_j$ is the closest member in the training set to $x_j$, $\sigma$ is the sample standard deviation of *S(M)* and *N* is the number of samples, in our case , *N* is 100000.

A compact $(C)$ model is the one that can use fewer components to represent the same variation.

$$C(M) = \sum_{m=1}^{M} \lambda^m$$
$$\sigma_{C(M)} = \sum_{m=1}^{M} \sqrt{\frac{2}{n_s}} \lambda^m \quad \textbf{(6)}$$

So, we can conclude, if model A is better in compactness, generalization ability or specificity than model B, it will achieve lower value on that ability measurement. For more details you can read [11].

Furthermore, we did an extra comparison, in experiment eight, between MDL and MEM. In total, 22 sets of facial silhouettes were used to evaluate the pile up problem, since silhouettes offer high local curvature information and complicated landforms on their shapes.

### A. First Experiment

Data sets: 24 contours of metacarpals with 64 marks, 8 nodes and a master example.
Goal: Closed curve

Fig.1.1. From top to bottom, they are mean shape plus $3\sqrt{\lambda}$ , mean shape and mean shape plus $-3\sqrt{\lambda}$ .

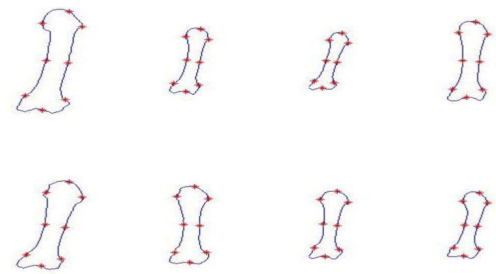

Fig.1.2. Eight results with the final optimal nodes positions.

### B. Second Experiment

Data sets: 32 contours of femurs with 65 marks, 9 nodes,
Goal: Open curve, free ends

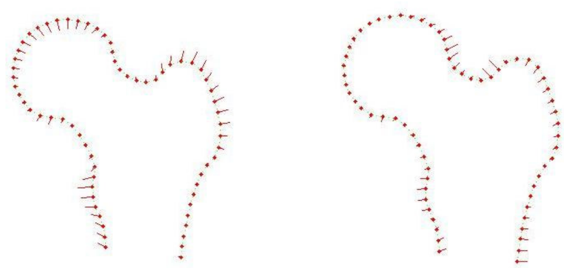

Fig.2.1. Shown is the mean shape with red marks, the whiskers emanating from the marks indicate three standard deviations of the first two principal components.

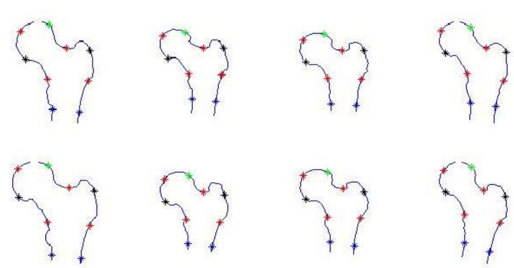

Fig. 2.2 Result of MEM analysis of femur contours. Here 8 of the 32 examples are shown with optimized node positions. It can be seen that they appear to be placed in a corresponding manner and the free end-points have been placed in different portions of the shafts.

*C. Third Experiment*

Data sets: 22 silhouettes of heads with 65 marks, 9 nodes
Goal: Open curve with free ends

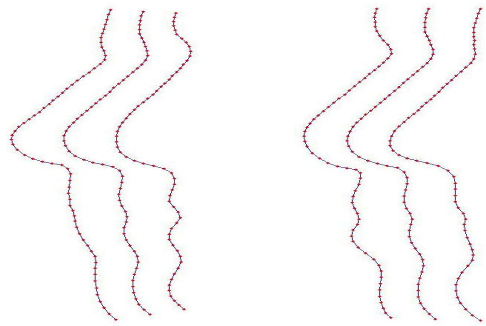

Fig.3.1. after MEM, the model shows the effect of moving the first and second components in the range over the training set.

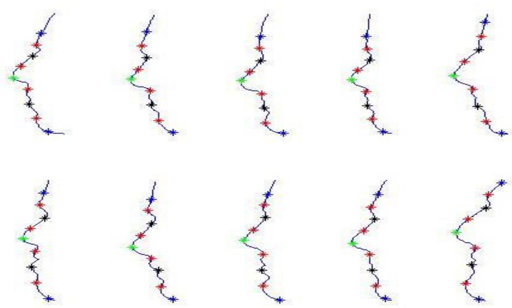

Fig. 3.2 Result of MEM analysis of silhouettes contours. Here 10 of the 22 examples are shown with optimized node positions (Blue is level one, green is level two, black is level three and red is level four). It can be seen that they appear to be placed in a corresponding manner and the free end-points (in both ends) have been placed in different portions of the shafts.

*D. Forth Experiment*

Data sets: 10 hand shapes with
Goal: Open curve, fixed ends

Fig. 4.1 Shown is the mean shape with red marks, the whiskers emanating from the marks indicate three standard deviations of the first three principal components.

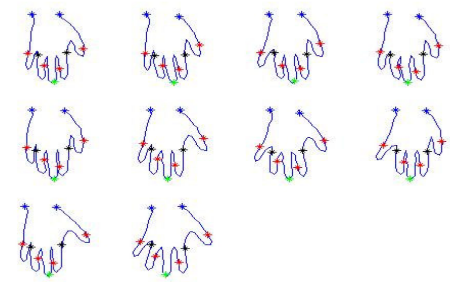

Fig. 4.2 Result of MEM analysis of hand contours. Here 10 examples are shown with optimized node positions (Blue is level one, green is level two, black is level three and red is level four). It can be seen that they appear to be placed in a corresponding manner and the free end-points (in both ends) are fixed

*E. Fifth Experiment*

The comparisons are done between MDL (Han's [4] with one fixed shape) with MEM (our proposed method) on three evaluation properties: Compactness, generalization ability and specificity**.**

Three properties on metacarpals:

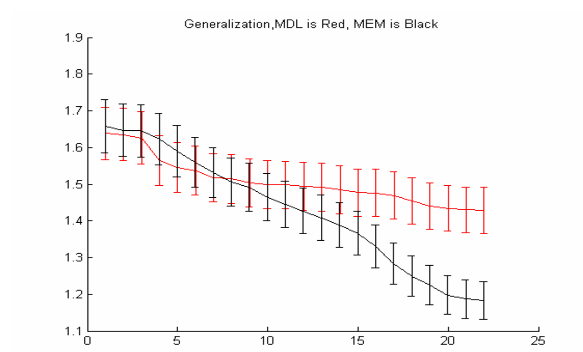


Fig.5.1. Generalization comparison between MDL and MEM on metacarpals.

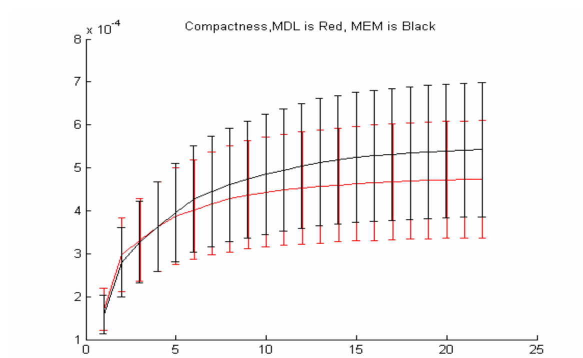


Fig.5.2. Compactness comparison between MDL and MEM on metacarpals.

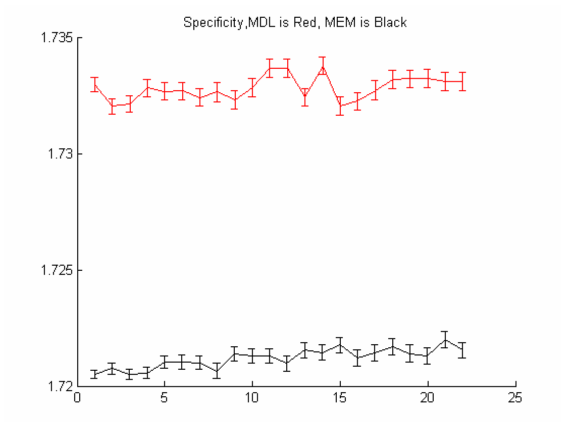


Fig.5.3. Specificity comparison between MDL and MEM on metacarpals.

### F. *Sixth Experiment*

Three properties on femurs:

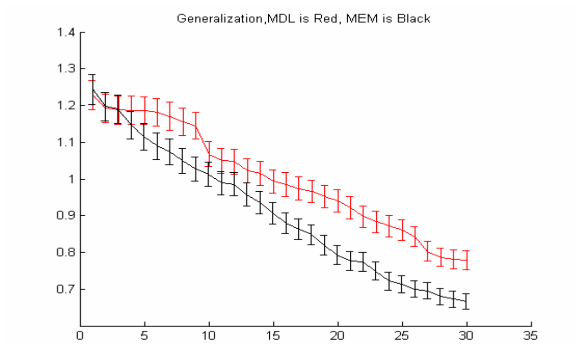


Fig. 6.1 Generalization comparison between MDL and MEM on femurs

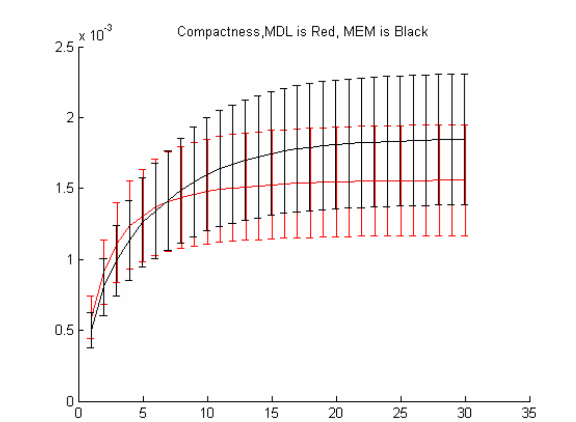


Fig. 6.2 Compactness comparison between MDL and MEM on femurs.

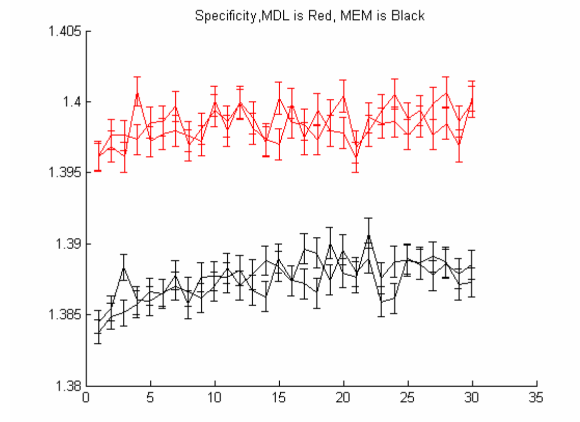


Fig. 6.3 Specificity comparison between MDL and MEM on femurs.

### G. *Seventh Experiment*

Three properties on hand shapes:

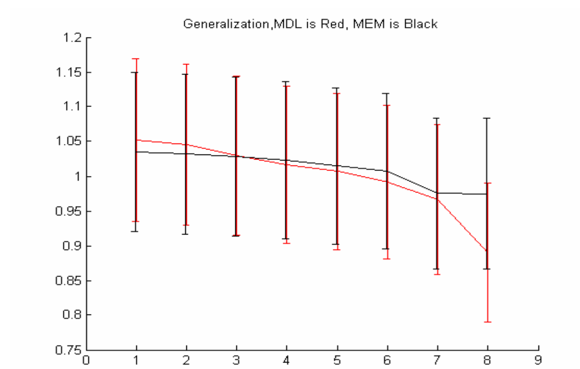


Fig. 7.1 Generalization comparison between MDL and MEM on hands.

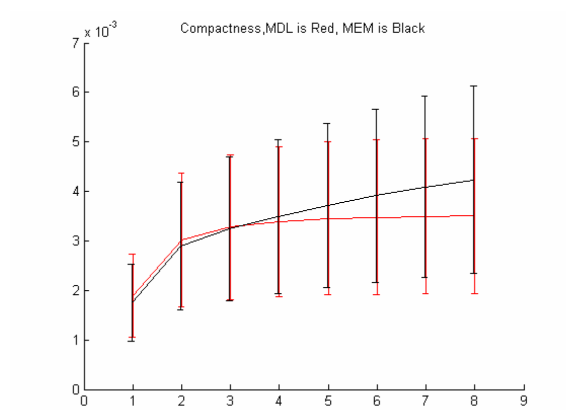


Fig. 7.2 Compactness comparison between MDL and MEM on hands.

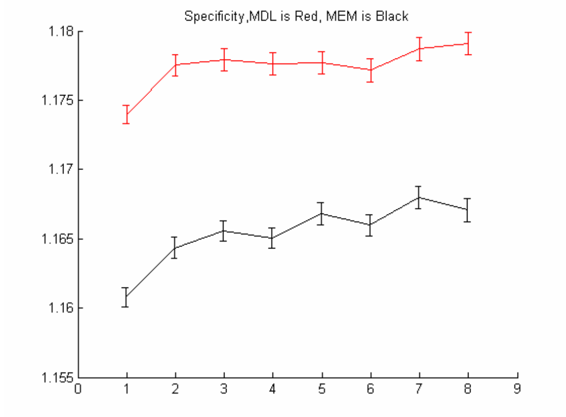


Fig.7.3. Specificity comparison between MDL and MEM on hands.

### *H. Eighth Experiment - improved control of "run away" in silhouettes*

When applying the MDL technique, one may encounter the so-called "run away" problem. The problem is that during optimization, points can possibly move into one location that will actually attain a global minimum avoiding descripting other parts of the shape. Davis [6] has tried to use a fixed shape to control these pile up and Hans [7] argues that one fixed master example shape would not always hold right. Furthermore he added an external term such as curvature or a node penalty to bound points in a reasonable area. We argue that our proposed algorithm does not suffer from the "pile up" problem without using an external term. Results are given below.

Here, in Figure 8 (One step before two points overlap together), MDL meets the problem of pile up. A level four point (red, bottom) and a level one point (blue, bottom) pile up. However, it can be seen at Figure 3.2 that our algorithm can conquer this problem.

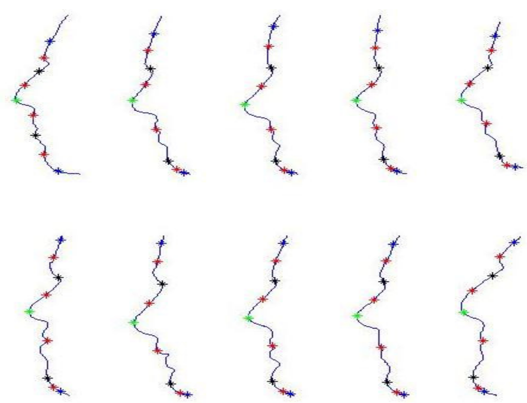

Fig. 8 Result of MDL analysis of silhouettes contours. Here 10 examples are shown, they are one step between MDL finally converged (Blue is level one, green is level two, black is level three and red is level four). It can be seen that the points at the bottom tried to pile up though one fixed master example has been used (first one). Compared with Fig 3.2, MEM shows reasonable better results.

## V. CONLUSIONS AND FUTURE WORK

The main original contributions in this paper are:

*1)* We propose a novel method for finding point correspondences for shape model building applications.

*2)* We provide quantitative comparisons between MEM and the original MDL using the criteria of: compactness, generalization ability and specificity.

*3)* Finally, we provide a comparison between MEM and MDL when dealing with the "run away" or "pile up" problem. Out proposed technique avoids this well known pitfall.

The work in this paper could be extended in several ways. Firstly we are planning in testing the proposed method on 3D datasets; in that case, a new and/or more efficient optimization strategy has to be used. Secondly we will like to compare our proposed method against models built using the PCA with EM algorithm on larger datasets. In addition we are hoping to employ and test our model building method is more geometrically complicated abdominal organs.


## ACKNOWLEDGMENT

Thanks to Hans Thodberg for his support and data on MDL.


## REFERENCES


[1] G. Gerig, M. Styner, "Shape verse Size: Improved Understanding of the Morphology of Brain Structures ," *MICCAI.*2001 pp.24-32

[2] Y. Wang and L.H. Staib, "Boundary finding with correspondence using statistical shape models" in *Proceedings of IEEE, Conference on computer vision and pattern recognition* , Santa Barbara, California 1998, pp. 338–345.

[3] H, Chui and A.Rangarajan. "A new algorithm for non-rigid point matching" *Procing, IEEE conference, computer vision pattern Recognition* 2000, pp. 44–51.

[4] C. Brechbühler, G. Gerig, O. Kübler, "Parameterization of Closed Sureface for 3-D Shape Description" *Computer Vision and Image Understanding* 1995 pp.154-170

[5] A.C.W.Kotcheff and C.J.Taylor, "Automatic construction of eigenshape models by direct optimization" *Medical Image Analysis.* 1998, pp.303-314.

[6] R.H. Davies, T.F. Cottes, and C.J. Taylor, "A minimum description length Approach to statistical shape modeling " *IEEE transaction medical Imaging,* 2002, Vol 21,pp. 525-537.

[7] H.H. Thodberg, " A Minimum Description length Approach to statistical shape modeling " *Lecture note in computer science* 2003, Vol 2732, pp.525-537

[8] A. Hill and C.J. Taylor. "Automatic landmark generation for point distribution models," *5th British Machine Vision Conference,* 1994, pp.429-438

[9] C.E. Shannon, "The mathematical theory of communication (parts 1 and 2)," bell syst Tech. J. 27 1948 pp.379–423, 623–56

[10] T.F. Cootes, A. Hill, C.J. Taylor, J. Haslam, "The Use of Active shape Models for Locating Structuring in Medical Imaging " *Image Vision Computation* 1994 pp.355-366

[11] Martin A. Styner, Kumar T. Rajamani, Lutz-Peter Nolte, Gabriel Zsemlye, Gabor Szekely, Chris J. Taylor and Rhodri H. Davies. "Evaluation of 3D Correspondence Methods for Model Building" *Information Processing In Medical Imaging, Proceedings Leture Notes In Computer Science* 2003 pp.63-75