Hierarchical Bayesian Modeling, Estimation, and Sampling for Multigroup Shape Analysis

Yen-Yun Yu[1], P. Thomas Fletcher[1], and Suyash P. Awate[1,2]

[1] Scientific Computing and Imaging (SCI) Institute, School of Computing, University of Utah
[2] Computer Science and Engineering Department, Indian Institute of Technology (IIT), Bombay

Abstract. This paper proposes a novel method for the analysis of anatomical shapes present in biomedical image data. Motivated by the natural organization of population data into multiple groups, this paper presents a novel *hierarchical generative* statistical model on shapes. The proposed method represents shapes using pointsets and defines a joint distribution on the population's (i) shape variables and (ii) object-boundary data. The proposed method solves for optimal (i) point locations, (ii) correspondences, and (iii) model-parameter values as a *single* optimization problem. The optimization uses expectation maximization relying on a novel Markov-chain Monte-Carlo algorithm for *sampling* in Kendall shape space. Results on clinical brain images demonstrate advantages over the state of the art.

Keywords: Shape analysis, hierarchical Bayes, sampling in shape space.

1 Introduction and Related Work

Shape analysis [6,9] entails the inference of shape models from population data and associated statistical analyses, e.g., hypothesis testing for comparing groups. The natural organization of biomedical data into groups, and possibly subgroups, calls for a *hierarchical* modeling strategy. Previous works on hierarchical shape modeling typically concern (i) multi-resolution models, e.g., a face model at fine-to-coarse resolutions, or (ii) multi-part models, e.g., a car decomposed into body, tires, and trunk. In contrast, the proposed framework deals with population data comprising multiple groups, e.g., the Alzheimer's disease (AD) population comprising people with (i) dementia due to AD, (ii) mild cognitive impairment due to AD, and (iii) preclinical AD.

Figure 1 outlines the proposed *generative* model, where (i) top-level variables capture the shape properties across the population (e.g., all individuals with and without medical conditions), (ii) variables at a level below capture the shape distribution in different groups within the population (e.g., clinical cohorts based on gender or type of disease within a spectrum disorder), and (iii) variables at the next lower level capture individual shapes, which finally relate to (iv) individual image data at the lowest level. Moreover, the top-level population variables provide a common reference frame for the group shape models, which is necessary to enable comparison between the groups.

This paper makes several contributions. **(I)** It proposes a novel hierarchical generative model for population shape data. It represents a shape as an equivalence class of pointsets modulo translation, rotation, and isotropic scaling [6]. This model tightly couples each individual's shape (unknown) to the observed image data by designing their

P. Golland et al. (Eds.): MICCAI 2014, Part III, LNCS 8675, pp. 9–16, 2014.

joint probability density function (PDF) using current distance or kernel distance [8,17]. The current distance makes the logarithm of the joint PDF a nonlinear function of the point locations. Subsequently, the proposed method solves a *single unified model-fitting optimization problem to estimate optimal point locations, correspondences, and parameter values*. **(II)** The proposed model fitting relies on expectation maximization (EM), treating the individual-shape and group-shape variables as hidden random variables, thereby integrating them out while estimating parameters (e.g., the population shape mean and covariance). In this way, the proposed EM algorithm improves over typical methods that use mode approximation for shape variables. **(III)** The EM algorithm entails evaluating an expectation over the posterior PDF of the shape variables. For instance, the posterior PDF for individual-shape variables involves the (i) likelihood PDF designed using the current distance and (ii) prior PDF conditioned on the group shape model. To compute the expectation, the proposed EM algorithm relies on a novel adaptation of Hamiltonian Monte Carlo (HMC) [5] *sampling in Kendall shape space*. **(IV)** The results show that the hierarchical model leads to more compact model fits and improved detection of subtle shape variations between groups.

Early approaches [2,6] to statistical shape modeling rely on manually defined homologous landmarks. Later approaches optimize point positions or correspondences using statistical compactness criteria such as the (i) logarithm of the determinant of the model covariance matrix [10], (ii) minimum description length [4,16], or (iii) minimum entropy [1]. However, these approaches (i) do *not* incorporate a generative statistical model, (ii) introduce adhoc terms in the objective function to obtain correspondences, and (iii) do *not* estimate shape-model parameters within the aforementioned optimization. Some generative models for shape analysis do exist [3,7,12,14], but these models rely on a pre-determined template shape with manually placed landmarks.

2 Hierarchical Bayesian Shape Model

We first describe the proposed hierarchical model for multigroup shape data (Figure 1).

Data: Consider a group of I vector random variables $X := \{X_i\}_{i=1}^{I}$, where X_i is a vector random variable denoting a given set of points on the boundary of an anatomical structure in the i-th individual's image data. That is, $X_i := \{X_i(n)\}_{n=1}^{N_i}$ where $X_i(n) \in \mathbb{R}^D$ is the D-dimensional spatial coordinate of the n-th point in the pointset. Such points can be obtained from a given segmentation or delineation of the anatomical structure. In this paper, $D = 3$. In any individual's image data, the number of boundary points N_i can be arbitrary. Similarly, consider other groups of data, e.g., data $Y := \{Y_j\}_{j=1}^{J}$ derived from a group of J individuals, data $\{Z_k\}_{k=1}^{K}$, etc.

Individual Shape Variables: For the first group (corresponding to data X), consider a group of I *hidden* random variables $U := \{U_i\}_{i=1}^{I}$, where U_i is a vector random variable representing the shape of the anatomical structure of the i-th individual. That is, $U_i := \{U_i(t)\}_{t=1}^{T}$ where $U_i(t) \in \mathbb{R}^D$ is the D-dimensional coordinate of the t-th point in the shape representation of the i-th individual's structure. We assume the observations X_i to be derived from the individual shape U_i. Similarly, we consider hidden random variables, i.e., V, W, etc., representing shapes for the other groups. To enable intra-group and inter-group statistical analysis, we ensure that all shape models lie in the same space by enforcing the same number of points T in all shape models.

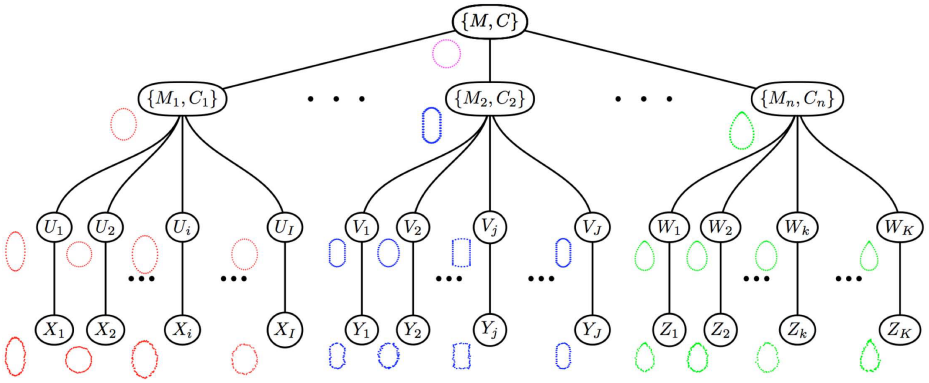

Fig. 1. Proposed Hierarchical Generative Statistical Model for Multigroup Shape Data

Group Shape Variables: Consider the first group of shapes U to be derived from a shape PDF having a mean shape M_1 and a shape covariance C_1. Consider other groups of shapes modeled analogously, i.e., V derived from a group with shape mean and covariance (M_2, C_2), W derived from a group with shape mean and covariance (M_n, C_n), etc. This paper treats the group means, i.e., M_1, M_2,$\cdots$,M_n, as hidden random variables and the group covariances, i.e., C_1, C_2,$\cdots$,C_n, as parameters. The proposed method can be generalized to treat the group covariances as random variables.

Population Shape Variables: Consider all group shape means, i.e., M_1, M_2,$\cdots$, M_n, to be derived from a single population of shapes with mean M and covariance C. In this paper, without loss of generality, we only consider two groups ($n = 2$) for simplicity.

Joint PDF: We model the joint PDF with (i) parameters M, C, C_1, C_2, (ii) group shape variables M_1, M_2, (iii) individual shape variables U, V, and (iv) data X, Y as:

$$P(M_1, M_2, U, V, X, Y|M, C, C_1, C_2) := \tag{1}$$
$$P(M_1|M, C)P(M_2|M, C)\Pi_{i=1}^{I}P(U_i|M_1, C_1)P(X_i|U_i)\Pi_{j=1}^{J}P(V_j|M_2, C_2)P(Y_j|V_j).$$

PDF of Individual Data given Individual Shape: We model $P(X_i|U_i)$, $P(Y_j|V_j)$ using current distance. Between pointsets $A := \{a_i\}_{i=1}^{I}$ and $B := \{b_j\}_{j=1}^{J}$, the squared current distance is $d_K^2(A, B) := \sum_{i=1}^{I}\sum_{i'=1}^{I} K(a_i, a_{i'}) + \sum_{j=1}^{J}\sum_{j'=1}^{J} K(b_j, b_{j'}) - 2\sum_{i=1}^{I}\sum_{j=1}^{J} K(a_i, b_j)$, where $K(\cdot, \cdot)$ is a Mercer kernel. In this paper, $K(\cdot, \cdot)$ is the Gaussian kernel with isotropic covariance $\sigma^2\mathbf{I_D}$. We use the current distance to define $P(X_i|U_i) := (1/\gamma)\exp\left(-d_K^2(X_i, U_i)\right)$, over finite support, where γ is the normalization constant. The current-distance model allows the number of points in the shape models U_i to be different from the number of boundary points in the data X_i.

Group Shape PDF: We model $P(U_i|M_1, C_1)$ as Gaussian with mean M_1 and covariance C_1 and $P(V_j|M_2, C_2)$ as Gaussian with mean M_2 and covariance C_2.

PDF of Group Variables given Population Parameters: We model $P(M_1|M, C)$ and $P(M_2|M, C)$ as Gaussian with mean M and covariance C; we choose the Gaussian

(i) to be maximally non-committal during model design and (ii) as the conjugate prior for the Gaussian means M_1, M_2. Under the Gaussian model, strange-looking shapes can be avoided by preventing over-regularization of the covariance estimate and preventing very large deviations from the mean (which are rare events under the Gaussian). More importantly, the hierarchical model alleviates this issue by producing covariance estimates that are more compact and restrict variation over fewer modes (see Figure 3).

3 Fitting the Shape Model to Data Using Monte-Carlo EM

This section presents the EM algorithm for the model-fitting optimization problem. The parameters in our model are: (i) the population mean M and covariance C and (ii) the group covariances C_1, C_2. Denoting $\theta := \{M, C, C_1, C_2\}$, the optimal model fit is:

$$\arg\max_{\theta} P(x, y|\theta) = \arg\max_{\theta} \int P(u, v, m_1, m_2, x, y|\theta) du dv dm_1 dm_2. \quad (2)$$

3.1 E Step: Sampling in Shape Space by Adapting the Hamiltonian Monte Carlo

In the i-th iteration, with parameter estimate $\widehat{\theta^i}$, the E step constructs the Q function as

$$Q(\theta|\widehat{\theta^i}) := E_{P(U,V,M_1,M_2|x,y,\widehat{\theta^i})} \log P(U, V, M_1, M_2, x, y|\theta). \quad (3)$$

Because of the analytical intractability of this expectation, we approximate $Q(\theta|\widehat{\theta^i}) \doteq \widehat{Q}(\theta|\widehat{\theta^i}) := \sum_{s=1}^{S} (1/S) \log P(u^s, v^s, m_1^s, m_2^s, x, y|\theta)$ using Monte-Carlo simulation. To sample the set of individual shapes u^s, v^s and the group-mean shapes m_1^s, m_2^s from $P(U, V, M_1, M_2|x, y, \widehat{\theta^i})$, we propose Gibbs sampling coupled with a novel adaptation of the HMC sampler [5]. Before describing the adapted HMC sampler, we outline the proposed shape-sampling algorithm for generating a sample of size S:

1. Set the sample index variable s to 0. Initialize the sampling algorithm with the sample point $s = 0$ denoted by $u^0 := \{u_i^0\}_{i=1}^I, v^0 := \{v_j^0\}_{j=1}^J, m_1^0, m_2^0$.
   Given sample point s, sample the $(s+1)$-th sample point as follows.
2. Initialized with u_i^s, $\forall i$ sample $u_i^{s+1} \sim P(U_i|v^s, m_1^s, m_2^s, x, y, \widehat{\theta^i})$.
3. Initialized with v_j^s, $\forall j$ sample $v_j^{s+1} \sim P(V_j|u^{s+1}, m_1^s, m_2^s, x, y, \widehat{\theta^i})$.
4. Initialized with m_1^s, sample $m_1^{s+1} \sim P(M_1|u^{s+1}, v^{s+1}, m_2^s, x, y, \widehat{\theta^i})$.
5. Initialized with m_2^s, sample $m_2^{s+1} \sim P(M_2|u^{s+1}, v^{s+1}, m_1^{s+1}, x, y, \widehat{\theta^i})$.
6. If $s + 1 = S$, then stop; otherwise increment s by 1 and repeat the previous 4 steps.

We ensure the independence of samples between Gibbs iteration s and the next $s + 1$, by running the HMC algorithm sufficiently long and discarding the first few samples s.

HMC is a Markov-chain Monte-Carlo sampling algorithm. HMC exploits the gradient of the log PDF for fast exploration of the space of the random variables. The HMC approach first augments the original random variables with auxiliary momentum variables, then defines a Hamiltonian function combining the original and auxiliary variables, and, subsequently, alternates between simple updates for the auxiliary variables

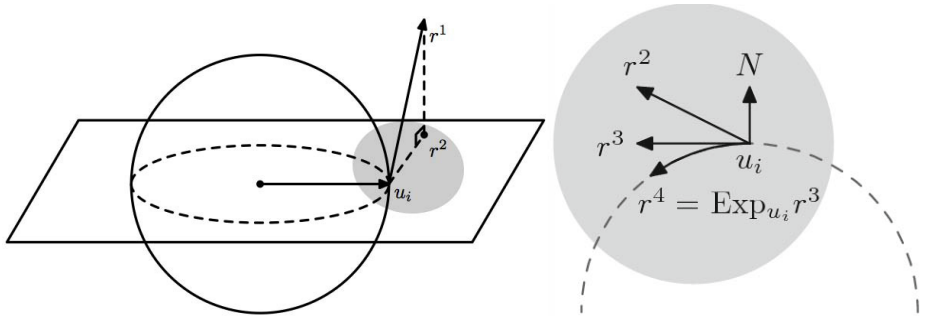

Fig. 2. Gradient Projection within HMC for Sampling in Shape Space **Left:** shows Kendall's pre-shape space [9] (dotted hypersphere) that is the intersection of the (bold) hypersphere of fixed radius ρ (i.e., $\sum_t \|u_i(t)\|^2 = \rho^2$; fixes scale) and the hyperplane through the origin (i.e., $\sum_t u_i(t) = \mathbf{0}$; fixes translation). For a pointset u_i, log-posterior gradients r^1 are projected onto the hyperplane to produce r^2 that eliminates translation. **Right:** To remove changes in scale, the resulting projection r^2 is then projected onto the tangent space at u_i, tangent to the pre-shape space, and the resulting tangent-space projection r^3 is mapped to the pre-shape space via the manifold exponential map to give r^4. The text describes the last part of the projection.

and Metropolis updates for the original variables. HMC proposes new states by computing a trajectory according to the Hamiltonian dynamics implemented with a leapfrog method and guarantees the new proposal states to be accepted with high probability. In our case, HMC requires gradients of $\log P(U, V, M_1, M_2 | x, y, \widehat{\theta^i})$ with respect to the hidden variables $\{U_i\}_{i=1}^I, \{V_j\}_{j=1}^J, M_1, M_2$.

Using HMC naively leads to pointset updates that can change the location, scale, and pose of the pointset, thereby making the sampler very inefficient. For this problem, we propose to modify HMC by replacing the gradient of the log posterior by a *projected gradient* that restricts the updated shape to Kendall shape space. As shown in Figure 2, starting with pointset u_i, the log-posterior gradient r^1 is first projected onto the preshape space to produce r^4 that has the same centroid and scale as u_i. Then, to remove rotation effects, the resulting pre-shape r^4 is rotationally aligned with the u_i, yielding r^5 (not shown in figure). These steps project the log-posterior gradient at u_i, within HMC, to generate an updated shape r^5 as part of the trajectory within HMC.

3.2 M Step: Parameter Estimation

In iteration i of the EM optimization, the M step maximizes $\widehat{Q}(\theta|\widehat{\theta^i})$ over θ and sets $\widehat{\theta}^{i+1} \leftarrow \arg\max_\theta \sum_{s=1}^S \log P(u^s, v^s, m_1^s, m_2^s, x, y|\theta)$. Equating the gradient of this objective function to zero gives closed-form optimal estimates for all parameters.

4 Group Comparison Using Permutation Testing

After the model is fit to the data, we can perform hypothesis testing to compare any pair of groups; the null hypothesis is that the two groups of data were drawn from the same

PDF. Since the shape PDF in each group is modeled using Mahalanobis distances based on means M_1, M_2 and covariances C_1, C_2, we use Hotelling's two-sample T^2 statistic to measure dissimilarity between any pair of groups. However, in 3D medical image data, the dimensionality TD can be very high compared to the number of individuals. Low sample sizes can render the F-distribution unusable. Simulating shapes with sample sizes higher than the dimensionality TD can be computationally expensive. Thus, we propose to employ distribution-free hypothesis testing, namely, permutation testing, using Hotelling's T^2 as the test statistic. Permutation testing is conservative in rejecting the null hypothesis and enhances robustness to specific modeling choices, e.g., the cardinality of the shape-model pointsets and internal model free parameters.

5 Results and Discussion

This section shows results on simulated and real data. We assume the input images, undergoing shape analysis, to be binary or soft masks, having intensities in the range $[-1, 1]$, that segment the image into the object of interest and the background. For hypothesis testing, it is less interesting to compare the performances of methods when the two groups are (i) exactly similar or (ii) differ extremely. The real challenge is in being able to reject the null hypothesis when the two groups differ in subtle ways.

For the proposed hierarchical model, we initialize the pointsets that model shape as follows. First, we solve a groupwise registration problem on the mask images, using a similarity transform, to (i) register the images, representing shape, to a common space and (ii) find an average (mask) image in that space. We assume the data to be the set of voxels on the zero crossing of the mask images warped to the common space. Then, we (i) threshold the average mask to get an object boundary, (ii) embed it as the zero level set of a signed distance-transform image, and (iii) generate a 3D triangular mesh for the zero level set using [13]. Finally, we use this mesh-vertex pointset as the *initial* value for M, m_1^0, m_2^0, $\{u_i^0\}_{i=1}^I$, and $\{v_j^0\}_{j=1}^J$. We set C, C_1, C_2 to (scaled) identity. We set the σ for the Gaussian kernel, underlying the current distance, to be the average edge length in the mesh. With this initialization, we compare the proposed method with a state-of-the-art algorithm [1], implemented in the open-source software ShapeWorks [15].

5.1 Validation on Simulated 3D Shapes

We simulate 2 groups of ellipsoidal shapes (ellipsoids in canonical form; 20 pointsets per group), where the groups are subtly different from each other. Two of the axes have length 1. The lengths of the third axis for the (i) 1st group are drawn from a Gaussian with mean 0.9 and variance 0.01 and (ii) 2nd group are drawn from a Gaussian with mean 1.1 and variance 0.01. The pointsets are then rescaled to constant norm.

The proposed method as well as ShapeWorks (i) both employ $T = 64$ points per pointset for shape modeling and (ii) both take as input equivalent information, i.e., while ShapeWorks takes as input a signed-distance-transform image (Figure 3) representing the ellipsoids implicitly, the proposed method takes as input the corresponding zero-crossing image. With $T = 64$, the average distance between a point and its nearest neighbor, in the shape pointset, is around 10 voxels. For both methods, the covariance

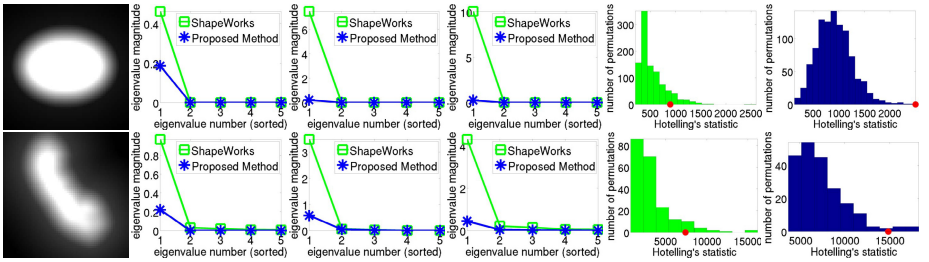

Fig. 3. Results. Top Row: *Ellipsoidal Shapes in Simulated Data.* Bottom Row: *Hippocampal Shapes in Clinical Brain MR images.* Left to Right: Distance transform image data (2D slice); Eigenspectra of the population covariance C and the group covariances C_1, C_2; Permutation distribution of Hotelling's T^2 test statistic for ShapeWorks (green) and the proposed method (blue); the red circle shows the value of the test statistic for the unpermuted group labeling.

estimates are regularized by addition of a scaled identity matrix δI, where δ is a free parameter; the experiments explore the robustness of both approaches to changes in δ.

Figure 3 (Top Row) shows the results from the proposed method compared to ShapeWorks, for the regularization parameter δ set to 10^{-4}. The proposed method leads to a fitted model that has smaller variances, at the group level as well as the population level. This indicates that the proposed method leads to a model that is more compact and fits the data better; this stems from improvements in optimal point placement and estimation of correspondences/parametrization. For the permutation distribution of the Hotelling's T^2 statistic, the p value for ShapeWorks is 0.05 and that for the proposed method is 0.001. Varying δ over $10^{-3}, 10^{-4}, \cdots, 10^{-10}$, we find that the p value for the proposed method stays at 0.001, but the p value of ShapeWorks varies and is never lower than 0.05. These results were unchanged when value of the current-distance parameter σ was multiplied by factors $\in [0.5, 2]$. This indicates that, compared to ShapeWorks, the proposed method was more robust to changes in δ and consistently produces a p value that tends to (correctly) reject the null hypothesis significantly more strongly.

5.2 Results on Clinical Brain MR Images: Hippocampal Shapes in Dementia

This section employs clinical brain magnetic resonance (MR) images from the OASIS [11] dataset. We use 10 randomly selected OASIS brains that uniformly sample the age span, including 4 cases with very-mild to mild Alzheimer's dementia and 6 controls, having hippocampus segmentations manually performed by a radiologist.

The proposed method and ShapeWorks, both, employ $T = 128$ points per pointset; the average distance between a point and its nearest neighbor is around 5 voxels. Figure 3 (Bottom Row) shows the results using $\delta = 10^{-4}$. These results were unchanged when value of the current-distance parameter σ was multiplied by factors $\in [0.5, 2]$. The proposed method leads to a fitted model that has smaller variances, indicating a compact better-fitting model. The p value for ShapeWorks is 0.07. The p value for the proposed method is 0.03 that indicates a relatively stronger rejection of the null hypothesis.

Discussion: The results show that the proposed hierarchical model and unified-optimization approach leads to compact-fitting shape models that can differentiate

subtle variations in hippocampal shapes (open-access data) better than the state of the art (open-source software). The main originality in the paper is in being able to solve the three problems of point placement, correspondence, and model-parameter estimation (given data from one or more groups) as a single optimization problem. Another key originality is in being able to sample in Kendall shape space, using a novel adaptation of HMC sampling using restricted gradients. The proposed framework can benefit from more accurate and efficient schemes for modeling and estimation.

Acknowledgments. We thank support from NIH-NCRR Center for Integrative Biomedical Computing P41-RR12553, NIH-NCBC NAMIC U54-EB005149, Royal Academy of Engineering 1314RECI076, NIH 5R01EB007688, and NSF CAREER 1054057.

References

1. Cates, J., Fletcher, P.T., Styner, M., Hazlett, H.C., Whitaker, R.: Particle-based shape analysis of multi-object complexes. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 477–485. Springer, Heidelberg (2008)
2. Cootes, T., Taylor, C., Cooper, D., Graham, J.: Active shape models - their training and application. Comp. Vis. Image Understanding 61(1), 38–59 (1995)
3. Coughlan, J., Ferreira, S.: Finding deformable shapes using loopy belief propagation. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002, Part III. LNCS, vol. 2352, pp. 453–468. Springer, Heidelberg (2002)
4. Davies, R., Twining, C., Cootes, T., Taylor, C.: A minimum description length approach to statistical shape modeling. IEEE Trans. Med. Imag. 21(5), 525–537 (2002)
5. Duane, S., Kennedy, A., Pendleton, B., Roweth, D.: Hybrid Monte Carlo. Physics Letters 195, 216–222 (1987)
6. Goodall, C.: Procrustes methods in the statistical analysis of shape. J. Royal Stat. Soc. 53(2), 285–339 (1991)
7. Gu, L., Kanade, T.: A generative shape regularization model for robust face alignment. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part I. LNCS, vol. 5302, pp. 413–426. Springer, Heidelberg (2008)
8. Joshi, S., Kommaraji, R., Phillips, J., Venkatasubramanian, S.: Comparing distributions and shapes using the kernel distance. In: ACM Symp. Comp. Geom., pp. 47–56 (2011)
9. Kendall, D.: Shape manifolds, Procrustean metrics, and complex projective spaces. Bull. London Math. Soc. 16(2), 81–121 (1984)
10. Kotcheff, A., Taylor, C.: Automatic construction of eigen shape models by direct optimization. Medical Image Analysis 2(4), 303–314 (1998)
11. Marcus, D., Wang, T., Parker, J., Csernansky, J., Morris, J., Buckner, R.: Open access series of imaging studies (OASIS): Cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. J. Cogn. Neuro. 19(9), 1498–1507 (2007)
12. Neumann, A.: Graphical Gaussian shape models and their application to image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 25(3), 316–329 (2003)
13. Persson, P., Strang, G.: A simple mesh generator in Matlab. SIAM Rev. 46(2), 329–345 (2004)
14. Rangarajan, A., Coughlan, J., Yuille, A.: A Bayesian network framework for relational shape matching. In: Int. Conf. Comp. Vis., pp. 671–678 (2003)
15. SCI Institute: ShapeWorks: An open-source tool for constructing compact statistical point-based models of ensembles of similar shapes that does not rely on specific surface parameterization (2013), `http://www.sci.utah.edu/software/shapeworks.html`
16. Thodberg, H.: MDL shape and appearance models. Info. Proc. Med. Imag. 2, 251–260 (2003)
17. Vaillant, M., Glaunes, J.: Surface matching via currents. Info. Proc. Med. Imag. 19, 381–392 (2005)