Local Residual Statistics for Segmentations Using Deformable Shape Model

Xiaoxiao Liu, Ja-Yeon Jeong , Joshua H. Levy,
Rohit R. Saboo, Edward L. Chaney , Stephen M. Pizer

Abstract

Both posterior optimization of deformable shape models and multiscale approaches have been potent directions for segmentation. In this work these two concepts are united. It is shown that the combination has measurable advantages both from the point of view of estimation robustness of the shape prior and from the point of view of segmentation accuracy. A probabilistic multiscale framework using m-rep (medial) structures is proposed, focusing on applying the local residual statistics to refine the segmentation results from the global scale. A measurement on the predictability of the statistics on a simulated data set of 1000 warped ellipsoids shows improved robustness via the multiscale training. Finally, the computed tomography image experiments for the bladder segmentation in day-to-day adaptive radiation therapy demonstrate that the proposed approach can bring the segmentation to subvoxel accuracy and thereby makes the results clinically adequate.

1. Introduction

3D medical image segmentation via deformable models is challenging; adoption occurs only if its accuracy is competitive with manual segmentations. For instance, in CT adaptive radiation therapy for male pelvic organs, we usually have 1mm by 1mm by 3mm voxels. Subvoxel precision of the segmented model is important for clinical use, but also a difficult task for automated segmentation.

Deformable models with various representations have been proven effective in capturing variations of a population of geometric entities [9]. An off-line learning process can build up the variation space for deformation [16] [3] [11]. Considering the various degrees of locality for geometric features, an efficient way of representing the shape variance is to decompose them into different scales.

The multiscale idea has been widely adopted in various segmentation methods. Several of them apply the scales to image intensity features; others apply them to model features. A coarse-to-fine multi-resolution protocol on landmarks is used by Christensen et al [2] for diffeomorphic mapping of brain anatomy. Krissian et al [8] developed multiscale detection of the vessel centerlines based on a cylindrical model, with the radius of the cross sections providing the scaling feature. The Active Shape Model (ASM) developed by Cootes et al [4] uses multi-resolution image profiles around each landmark for boundary refinement. Joshi et al. developed a multiscale m-rep segmentation [8] where the global scale stage optimized the posterior probability of the shape with trained statistics, but the local scale stage was a deterministic optimization on the full space of atom deformations. Without shape prior information, the local scale deformation can go wild if the intensity patterns are weak or the ad hoc weighting penalties are not sensitive for certain configurations.

In this paper, we describe a probabilistic multiscale approach for 3D grey-scale image segmentation via robust estimates of probability distributions for geometry. We bring the multiscale strategy into the posterior optimization framework by utilizing statistics on the residuals left from a global scale. Sharing the essence of wavelets, we apply the residual statistics locally to match the remaining spatial variability there.

Our probabilistic multiscale framework is applied on the posterior optimization segmentation using the medial structure called *m-reps* (Fig. 1) [16]. The previous m-rep segmentation method [15] has shown its strength in 3D CT images by the reported competent segmentation results compared with the manual segmentation . And in terms of average closest point distance and Dice similarity coefficient, the m-rep segmentation is among one of the best in literature of CT segmentation for male pelvic organs [5][13][7] [12][17]. However, the global m-rep segmentation results are not accurate enough due to occasional localized errors, which often are several voxels off the manually segmented boundaries. The proposed method is developed to improve the accuracy to subvoxel level.

The rest of the paper is organized as follows: Section 2 introduces the multiscale optimization methods used for m-rep segmentation. The methodology for applying local residual statistics to the fine scale stage is presented in Sec-
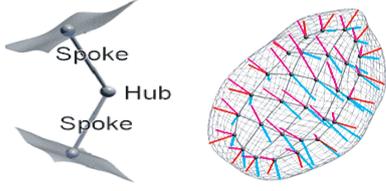
Figure 1. A m-rep model: An interior medial atom (left) with two spokes and an object (right). The object is composed of multiple medial atoms. Interior atoms have two spokes each. Exterior atoms (on the crest region) have three spokes each.

Figure 2. The probabilistic multiscale segmentation scheme for single object m-rep model: The object and atom stage shape statistics are collected from training samples and then used during the optimization process in the corresponding stages respectively.

tion 3, in which the robustness of the multiscale statistics against training sample size is also demonstrated. In Section 4, segmentation results on about 80 clinical CT image data of the male pelvic area are discussed in detail. Section 5 summarizes and discusses the method and results.

2. Probabilistic multiscale segmentation

2.1. M-rep Segmentation framework

In simple m-rep models, an object is a sheet of medial atoms represented by a quadrilateral mesh (Fig.1), which carries geometric properties such as the widening, bending and tapering, with the locality scale given according to the grid spacing. Each atom controls a local boundary region implied by its spokes' ending points. The object as a whole gives these properties in a way reflecting global interrelations of primitives / atoms, whereas each atom, and its relation to its immediate neighbors, provides more local features.

Based on the multiscale shape description ability of the m-rep, a coarse-to-fine strategy can deform the model hierarchically, as shown in Fig. 2. In each stage, a Bayes' posterior optimization scheme is applied using the off-line trained statistics in the corresponding scale. The global scale segmentation results provides a good starting point for the local refinement stage, where the model deforms locally towards more "correct" configurations.

2.2. Two-stage optimization

In the object stage the entire sheet of medial atoms deform together, restricted by the variations learned from the multiple sheets of atoms of the training samples. The generalized version of PCA for nonlinear data, Principal Geodesic Analysis (PGA) [6], is used for learning. We compute a Fréchet mean and the the principal eigenmodes from globally aligned training m-rep models. The first several modes that cover nearly 90% of the total variance of the training samples are used for optimization. PGA thereby reduces the dimension of the variation space from several hundred to a number less than 10. However, with the often limited number of training cases, the eigenmodes with little
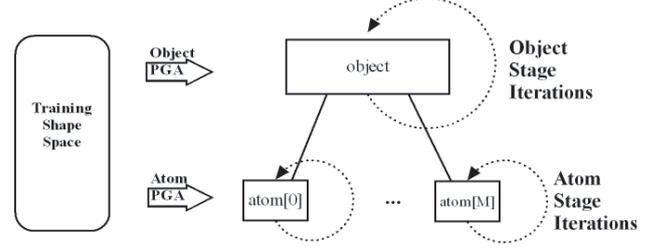
variance are not robust against the large dimensions of the object representation, which necessitates limiting the number of eigenmodes used and training the rest statistics on the residuals.

Besides the shape training, the image intensity patterns are learned by PCA on Regional Intensity Quantile Functions (RIQFs) of local regions corresponding to the spoke ends of the atoms, which probabilistically represent the appearance of the local region in the image. The details for the image appearance model can be found in [1]. With the training statistics, at each of the two scales the model deforms by optimizing the sum of two penalty terms: the log geometric prior and the image match measurement computed by the log probability distribution on the RIQFs of the implied boundary of the m-rep, since by Bayes' rule $\arg\max_{\mathbf{m}}(\log p(\mathbf{m}|\mathbf{I})) = \arg\max_{\mathbf{m}}[\log p(\mathbf{I}|\mathbf{m}) + \log p(\mathbf{m})]$. The optimization is over the principal geodesic coefficients, thus restricting the result to the shape space spanned by the principal geodesic directions.

At the object stage, the log prior is given by the global PGA and the image appearance statistics are collected across the whole object surface. While at the residual atom stage, each log prior is given by the PGA on the respective atom residual, and appearance statistics are collected in the local regions. The local scale posterior optimization is described by the formula:

$$\arg\max_{\Delta\mathbf{A}_i}(\log p((\mathbf{A}_i \oplus \Delta\mathbf{A}_i)|\mathbf{I}_i)$$
$$= \arg\max_{\Delta\mathbf{A}_i}[\log p(\mathbf{I}_i|\mathbf{A}_i \oplus \Delta\mathbf{A}_i) + \log p(\Delta\mathbf{A}_i)] \quad (1)$$

where $\Delta\mathbf{A}_i = \sum(x_j v_j^{atom})$ is the residual term that we will expand in detail in the next section. The optimization is over the coefficient vector x for atom i. Starting from the object stage resulting model, we randomly loop over all the atoms and update each atom by adding the residual deformation that gives the best log posterior probability value. During the iterations if one atom gets updated, all its neighbors will be updated later due to the new local configurations. Usu-

ally it takes 2-5 iterations for the optimization to converge.

3. Probability distributions on local residuals

3.1. Local residual statistics calculation

An atom in a m-rep model lives in a curved space, as does an atom shape residual. The atom resides in the symmetric space $\mathbf{G} = \mathbb{R}^3 \times \mathbb{R}^+ \times \mathbb{S}^2 \times \mathbb{S}^2$. Given two medial atoms $\mathbf{A}_i$, $\mathbf{A}_i' \in \mathbf{G}$, the residual $\Delta\mathbf{A}_i$, from $\mathbf{A}_i'$ to $\mathbf{A}_i$, is calculated by the subtraction operator.

$$\ominus : \mathbf{G} \times \mathbf{G} \to \mathbf{G}, \qquad (2)$$
$$\Delta\mathbf{A}_i \doteq \mathbf{A}_i \ominus \mathbf{A}_i'$$
$$\doteq (\Delta\mathbf{p}_i, \Delta\mathbf{r}_i, \Delta\mathbf{u}_i, \Delta\mathbf{v}_i).$$
$$\doteq (\mathbf{p}_i - \mathbf{p}_i', (\mathbf{r}_i - \mathbf{r}_i')/\mathbf{r}_i', \mathbf{R}_{\mathbf{u}_i'}(\mathbf{u}_i), \mathbf{R}_{\mathbf{v}_i'}(\mathbf{v}_i)),$$

where $\mathbf{A}_i = (\mathbf{p}_i, \mathbf{r}_i, \mathbf{u}_i, \mathbf{v}_i)$ with the hub position $\mathbf{p}_i \in \mathbf{R}^3$, the spoke length $\mathbf{r}_i \in \mathbb{R}^+$, and two unit spoke directions $\mathbf{u}, \mathbf{v} \in \mathbf{S}^2$. $\mathbf{R}_{\mathbf{w}}$ represents the rotation along the geodesics in $\mathbb{S}^2$ that moves a point $\mathbf{w} \in \mathbb{S}^2$ to the North Pole $(0, 0, 1)$ $\in \mathbb{S}^2$. The addition operator is correspondingly defined as

$$\oplus : \mathbf{G} \times \mathbf{G} \to \mathbf{G}, \qquad (3)$$
$$\mathbf{A}_i \doteq \mathbf{A}_i' \oplus \Delta\mathbf{A}_i$$
$$\doteq (\mathbf{p}_i' + \Delta\mathbf{p}_i, \mathbf{r}_i' \cdot (\Delta\mathbf{r}_i + 1), \mathbf{R}_{\mathbf{u}_i'}^{-1}(\Delta\mathbf{u}_i), \mathbf{R}_{\mathbf{v}_i'}^{-1}(\Delta\mathbf{v}_i)).$$

The shape residual can be calculated as

$$\Delta\mathbf{A}_i = \mathbf{A}_i^{train} \ominus \mathbf{A}_i^{obj}, \qquad (4)$$

where $\mathbf{A}_i^{train}$ is the training m-rep, which fits the corresponding hand-segmented binary image, and $\mathbf{A}_i^{obj}$ is the object-stage segmentation resulting model. We can get the approximated object stage model by projecting the ground truth model into the trained object shape space. The dot product of the model variance vector with the N eigenvectors produces the coefficients for the projection model:

$$x_j = < \mathbf{A}_i^{train} \ominus \mu_i^{train}, v_j^{obj} >$$
$$\hat{\mathbf{A}}_i^{obj} = \mathbf{A}_i^{(proj,obj)}$$
$$= \mu_i^{train} \oplus \sum_{j=1}^{N} x_j v_j^{obj}, \qquad (5)$$

where μ_i^{train} is the Fréchet mean of the training sample.

As for training the shape residual statistics, we apply PGA to get the major principal modes v_j^{atom} for each atom. The atom parametrization is 9 dimensional. The first K (typically 3 to 5) principal modes that cover more than 90% total variance are used in the atom stage optimization. Similar to the global scale projection, we project the residual into its trained residual shape space. Together with the global

scale projection, we calculate the multiscale projection as follows:

$$\mathbf{x}_j = < \mathbf{A}_i^{train} \ominus \mathbf{A}_i^{(proj,obj)}, v_j^{atom} >$$
$$\hat{\mathbf{A}}_i^{MultiStage} = \mathbf{A}_i^{(proj,obj+atom)}$$
$$= \mathbf{A}_i^{(proj,obj)} \oplus \sum_{j=1}^{K} x_j v_j^{atom}. \qquad (6)$$

To date we have used the assumption for the atom stage training that the probabilistic distribution of local residual $\Delta\mathbf{A}_i$ on atom i is independent of its neighbor atoms. Namely, we assume that the large range correlation among the atoms are fully captured by the global scale statistics so that the residual shape variation space of each atom is isolated. The independence assumption might be too strong if the global scale statistics are not tight enough or the geometric correspondence of the local regions are not good. There are ways to tighten the statistics via local alignment or computing the residual probability distribution conditioned on the neighbor configurations, which are beyond the focus of this paper.

3.2. Robustness test for the local residual prior

In addition to testing the multiscale strategy on the segmentation applications, we evaluate the robustness of the statistics on a simulated data set of 1000 warped ellipsoids. Binary images of warped ellipsoids were produced by successively applying independent bending, tapering, and twisting of a base ellipsoid. Each transformation was sampled from a zero-mean Gaussian with its respective variance. Training models were produced by optimizing the deformable m-rep models using object boundary, local interpenetration, and irregularity penalties. The resulting training data set of the 1000 m-rep models is quite satisfying both in terms of the quality of the boundary fits and high signal-to-noise ratio in the statistics, which might be better than the real world data, where we have manual segmentations from noisy gray scale intensity patterns and more complicated shape variations. An example is given in Fig.3, which shows the distribution of the shape variations among the trained eigenmodes from 40 randomly selected m-rep models. We found that the first three eigenmodes obtained by PGA separately capture each of the three types of variations that we put into the images. Consistently, the first three eigenmodes cover more than 95% of the total variations.

According to Muller [10] the robustness of estimation of the probability distribution for a given training sample size can be measured by the fit of a test population to the probability distribution in the trained shape space, over a variety of training samples of the specified size. The fitness is evaluated by the squared correlation ρ^2 between test cases
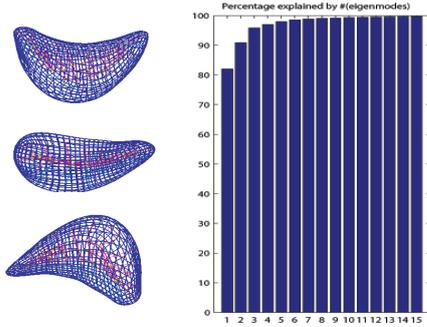
Figure 3. Left: 3 typical training m-reps that fit to corresponding warped ellipsoids binary images. Right: the accumulative variation percentage plot for the principal geodesic eigenmodes on randomly selected 40 m-reps from 1000 training m-reps.

Figure 4. Prediction measurement improvement by using the local residual statistics.

and their projections in the space estimated from training cases, which reduces to the following formula:

$$\rho^2 = \frac{\sum_{i=1}^{N} d(\widehat{m}_{i,test}, \overline{m}_{train})^2}{\sum_{i=1}^{N} d(m_{i,test}, \overline{m}_{train})^2}, \qquad (7)$$

where $\hat{m}_{i,test}$ is the projection of the test model $m_{i,test}$ on the shape space and $\overline{m}_{train}$ is the Fréchet mean of the training sample. The projected model indicates the closest model to the test model in the trained shape space, according to geodesic distances d in the feature space. The closer ρ^2 is to 1, the better predictability provided by the training.

We carried out three experiments to evaluate how well the residual statistics help to improve the predictability of the shape space on the warped ellipsoids data set.

In the first test we compared a) multiscale probability estimation using three global eigenmodes, corresponding to the modes we put in to the images, followed by another 4 residual eigenmodes that capture most of the variations for each atom, with b) global probability estimation with only the three global modes. The projections introduced in section 3.1 were used to estimate the closest model in the training space. We can see the improvement for the predictability of the multiscale shape space from Fig. 4.

Second, we compared the multiscale approach with the single global scale, using the same total number of eigenvectors for each of the atoms. That is, the raw dimension size (number of variables in all the eigenvectors taken together) for the entire model in the global scale analysis is the same as for the multiscale analysis. To be specific, in the example of ellipsoid m-rep, we have the 189-dimensional vector representing an single object, and a 9-dimensional vector for each of the 21 atoms. Thus we were comparing the predicted space of 7 global modes, each of which is 189-dimensional, with the multiscale space composed of 3 modes of 189 dimensions and an extra 4 modes of 9 dimensions for each of the 21 atoms. Fig. 5 shows that the
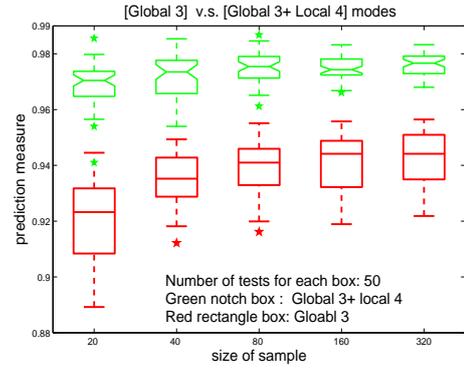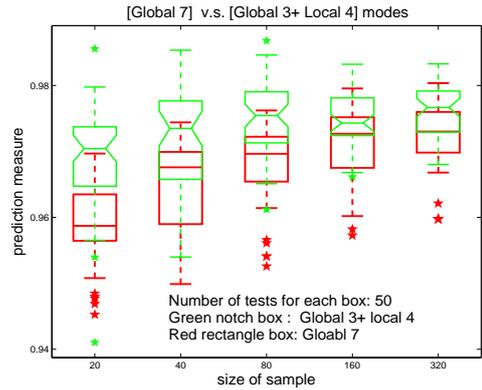
Figure 5. Prediction measurement comparison between the global scale and the multiscale approach, using the same total number of eigenvectors for each atom.

global-scale strategy requires 160 training samples to produce the same quality of estimation as provided by the multiscale strategy with 40 samples. We suspect that for other High Dimensional Low Sample Size (HDLSS) problems of shape analysis, the probabilistic multiscale framework will also be a better choice.

In our third experiment we compared the aforementioned multiscale analysis to a global analysis with the same total number of inherent dimensions, i.e., the same number of coefficients to be determined for each model. There are 87 such coefficients. The ρ^2 values for the global analysis were larger for each training sample size. However, optimizing over 87 coefficients at once is far less efficient and more likely to yield local optima than the successive optimization over far fewer dimensions at each stage provided by the multiscale approach.

4. Segmentation Results and Discussion

The real world data has more complicated shape variations and intensity patterns. We tested the methodology on a data set of CT images from 5 patients, and on-average 16 daily CT scans of the male pelvic area taken for each patient during the image-guided radiation therapy. The image has an in-plane resolution of 512×512 with pixel dimension of 0.98 mm $\times$ 0.98 mm and an inter-slice distance of 3 mm. The manual segmentation by an expert was provided.

For each patient, we first carried out a leave-one-day-out study on the bladder. Training was done on all other days when segmenting the target day. The results were compared with the manual segmentations and measured in terms of average closest surface point distance. Fig. 6 depicts the segmentation results . The graph is arranged by sorting each case based on the measurement for the object stage result. The three data points for each case show the metrics for the atom stage result, the object stage result and the training model as the reference respectively. As we can see from the figure, almost every case is more or less improved by the atom stage. The observed improvements in average surface distance, on the order of 0.1 millimeter, were often the result of significant improvements, i.e. of several millimeters, localized to just a few regions of the model. And the atom scale improvement can be seen across the population, which on average brings the accuracy into the subvoxel level. We also found that the improvements in the bladder segmentation are most noticeable at parts of the boundary where the contrast is high.

In addition, we applied the statistical atom stage to a leave-one-patient-out [14] study. Among the 5 patients, one was chosen as the target patient and the other 4 became the training sample. We pooled the trained variations from different patients together after a pelvic bone based alignment [14]. During the segmentation time, we took the previous days' mean model as the initialization for the next day's image segmentation, with the first day's segmentation being provided beforehand. Compared to the leave-one-day-out study, although the training sample size is about 3 times larger, this experiment was more challenge in terms of the more varied shape space across different patients and the less strong correspondences for the local regions across patients. Nevertheless, the refinement results can be seen from Fig. 6, where we can see the encouraging refinement.

An example of typical bladder segmentation results is shown to demonstrate how much the difference made by the atom stage visually on the boundaries. Fig.7 shows the 3D surface of the m-rep models of the object stage result (left) and the atom stage result (right), and Fig. 8 shows the three orthogonal views through the models with the corresponding slices of gray scale CT image as the background. In this case the average surface distance is reduced from 1.38 mm at the object stage to 1.15 mm after the atom stage, and
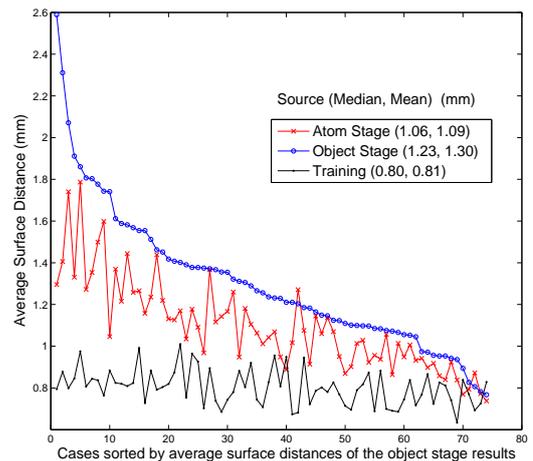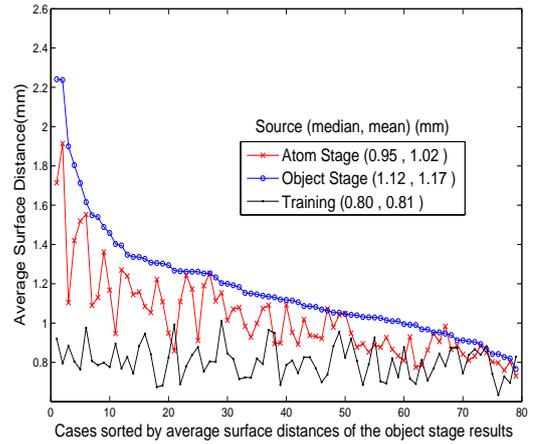
Figure 6. Leave-one-day-out(upper) and leave-one-patient-out (lower) bladder segmentation results for the 5 patients, measured by average surface distance. Cases are sorted by the object stage results. Two curves represent the atom stage and object stage results separately; The other training curve measures the fit of the training/ground-truth m-reps and thus represent the best possible segmentations.
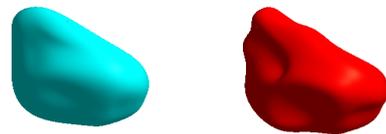
Figure 7. Comparison of object stage result (left) and the atom stage result (right): rendered surfaces of a particular bladder case.

volume overlap is increased from 92.3% to 93.8%.

From the CT image segmentation, we conclude that the local residual statistics work effectively. For the low contrast regions where even human have difficulties to locate
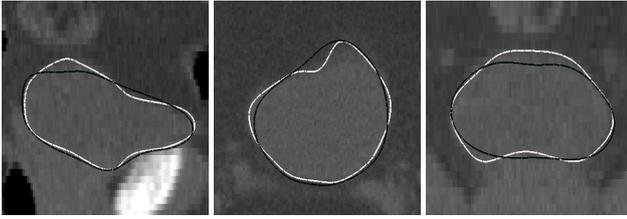
Figure 8. Comparison of object stage contours (black) and the atom stage contours (white) in 2D slices from sagittal (left), axial (middle) and coronal (right) views for one typical bladder case.

the boundary, the atom stage tends to be neither helpful nor harmful.

5. Conclusion

In this paper, we have presented a novel method to apply posterior optimization segmentation in a multiscale fashion by using the residual statistics locally. Our study suggests that the shape residual statistics are effective to restrict results in the trained shape space and to penalize the local deformation towards better posterior estimates, combined with the image appearance likelihood. We showed the improved robustness of multiscale statistics by measuring the squared correlation metric within a large simulated data set with random selected test groups and training groups various sample sizes. Also segmentation experiments on 79 CT bladder images demonstrated improved segmentation accuracy by the local scale refinement.

Besides m-reps, the usage of the local residual statistics is applicable to all sorts of representations with a notion of locality. Further investigation on the robustness of the method against noisy real world data via ρ^2 test needs to be carried out on a large real world data set.

References

[1] R. E. Broadhurst, S. M. Stough, Joshua V.and Pizer, and E. Chaney. A statistical appearance model based on intensity quantiles. *IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 422–425, 2006. 2

[2] G. Christensen, S. Joshi, and M. Miller. Volumetric transformation of brain anatomy. *IEEE Transactions on Medical Imaging*, 16(6):864–877, 1997. 1

[3] T. Cootes, D. H. Taylor, C. J.and Cooper, and J. Graham. Active shape models-their training and application. *CVGIP: Image Understand*, 61(1):38–59, 1994. 1

[4] T. F. Cootes, C. Taylor, and A. Lanitis. Active shape models: Evaluation of a multiresolution method for improving image search. *5 th British Machine Vision Conference*, pages 327–336, 1994. 1

[5] M. Costa, H. Delingette, and N. Ayache. Automatic segmentation of the bladder using deformable models. *Proc. of IEEE Int.Symposium on Boimedical Imaging(ISBI)*, pages 904–907, 2007. 1

[6] P. Fletcher and S. Joshi. Principal geodesic analysis on symmetric spaces: statistics of diffusiontensors. *ECCV 2004 Workshop on Computer Vision Approaches to Medical Image Analysis*, 2004. 2

[7] D. Gibou, F nd levy. Partial differential equationscbased segmentation for radiotherapy treatment planning. *IMathematical biosciences and engineering*, 2(2):209–226, 2005. 1

[8] K. Krissian, J. Ellsmere, K. Vosburgh, R. Kikinis, and C.-F. Westin. Multiscale segmentation of the aorta in 3d ultrasound images. *International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, pages 638–641, 2003. 1

[9] T. McInerney and D. Terzopoulos. Deformable models in medical image analysis: A survey. *Medical Image Analysis*, 1(2):91–108, 1996. 1

[10] K. E. Muller, S. Ray, and Y.-Y. Chi. Goodness of prediction for principal components, including high dimension, low sample size. *In prepertation*. 3

[11] C. Nikou, F. Heitz, J. Armspach, G. Bueno, and D. Vernon. A physically-based statistical deformable model for brain image analysis. *Proceedings of the 6th European Conference on Computer Vision-Part II*, 1843:528–542, 2000. 1

[12] T. Pasquier, D.and Lacornerie, R. J. Vermandel, M., E. Lartigau, and N. Betrouni. Automatic segmentation of pelvic structures from magnetic resonance images for prostate cancer radiotherapy. *Int. J. Rad Onc. Biol. Phys.*, 68(2):592–600, 2007. 1

[13] V. Pekar, T. McNutt, and M. Kaus. Automated model-based organ delineation for radiotherapy planning in prostatic region. *Int. J. Rad Onc. Biol. Phys.*, 60(3):973–980, 2004. 1

[14] S. M. Pizer, R. E. Broadhurst, J.-Y. Jeong, Q. Han, R. Saboo, J. V. Stough, G. Tracton, and E. L. Chaney. Intra-patient anatomic statistical models for adaptive radiotherapy. *MICCAI Conference: Workshop on From Statistical Atlases to Personalized Models: Understanding Complex Diseases in Populations and Individuals*, pages 43–46, 2006. 5

[15] S. M. Pizer, E. L. Chaney, and R. E. e. a. Broadhurst. Segmentation of kidneys and pelvic organs from ct by posterior optimization of m-reps. *techreport*, 2006. 1

[16] S. M. Pizer, P. T. Fletcher, and S. C. e. a. Joshi. Deformable m-reps for 3d medical image segmentation. *International Journal of Computer Vision*, 55(2-3):85–106, 2003. 1

[17] M. Rousson, A. Khamene, and M. e. a. Diallo. Constrained surface evolutions for prostate and bladder segmentation in ct images. *CVBIA*, pages 251–260, 2005. 1