Medical Image Synthesis via Monte Carlo Simulation

James Z. Chen, Stephen M. Pizer, Edward L. Chaney, and Sarang Joshi

Medical Image Display & Analysis Group, University of North Carolina, Chapel Hill, USA
{chenz,pizer}@cs.unc.edu, {chaney,sarang}@radonc.unc.edu

Abstract. A large number of test images and their “ground truth” segmentations are needed for performance characterization of the many image segmentation methods. In this work we developed a methodology to form a probability distribution of the diffeomorphism between a segmented template image and those from a population, and consequently we sample from these probability distributions to produce test images. This method will be illustrated by producing simulated 3D CT images of the abdomen for testing the segmentation of the human right kidney.

1 Introduction

This work explores a methodology for generating realistic, synthetic medical images for characterizing the performance of segmentation methods. It is intended to allow more effective validation and inter-comparison of algorithms that are designed to perform objective, reproducible segmentations of anatomical objects from medical images.

Test images intended for performance characterization should be able to represent the modality of interest and contain the “ground truth” segmentations, against which a particular segmentation method can be evaluated. Also, these images should represent statistical variations in the shape of target object across a population and span the range of image qualities found in a clinical setting. These properties mandate a large set of test images. In practice, however, it is a highly demanding task to define the ground truth segmentations for so many test images because 1) the manual segmentation by medical experts is a laborious and expensive process; 2) it has been shown that there exists a large variation even among the experts regarding the “correct” segmentations [1] due to subjective bias and the dependence on objectives.

A promising approach that could partially replace the costly performance evaluation procedures is to make use of synthetic images generated from simulations. Several statistical models of shape and appearance variability have been investigated and applied to perform various medical image analysis tasks. Examples are Active Shape Models (ASM) [2] and Statistical Deformation Models (SDM) [3]. These methods can also be adapted to generate synthetic images for performance characterization by sampling the distribution of variability. But in building an ASM, a set of segmentations of the shape of interest is required as well as a set of landmarks that can be unambiguously defined in each sample shape. The manual identification of the point correspondences is a time-consuming task and is prone to the subjective bias. In the SDM method, this difficulty is removed by applying an automated nonlinear image

registration. However, since the sample points in this model are distributed in the entire image volume, the computational complexity is non-trivial. In our method, the shape of the target object is represented via a limited number of automatically determined fiducial points. The computational efficiency is thus dramatically improved. Similar to the SDM method, the point correspondence across a population is established by a nonlinear, diffeomorphic image registration.

This work focuses on simulating the shape variability of the target anatomical object as observed across a population. The methodology will be laid out first, which includes the scheme of fiducial-point shape representation, the training process, the shape analysis and the subsequent image synthesis. Then, this method will be applied to synthesize the CT images of the human abdomen as a demonstration. Although the illustration here is limited to a single-organ system, this method should be applicable for generating synthetic images of any anatomical structure in principle, including multi-organ complexes.

2 Methodology

Our method for generating realistic, synthetic medical images proceeds in two steps: training and sampling. In the training phase, the shape of the target object is studied, from which a distribution of its variation is derived. In the sampling phase, a large number of synthetic images can be produced by randomly sampling from the probability distribution.

For training purposes a set of clinical images containing the target object need to be collected, so as to capture the shape variations of the anatomical object of interest across the population. These images will be referred to as the training images $\{I_t\}$. One of $\{I_t\}$ will be chosen as the template image I_0 , and other images will be registered to I_0 via a nonlinear, diffeomorphic warp function $H_t: I_t \rightarrow H_t(I_0)$. Given a predefined segmentation S_0 of the target object in I_0 , $\{H_t(S_0)\}$ can be analyzed to find its probability distribution (referred to as the “normal shape domain”, or NSD). Artificial warp functions, and consequently the synthetic images and their segmentations, can then be generated by randomly sampling in the NSD. Details of this new method will be developed in the following sections.

2.1 Training Image Segmentation

Training defines the degrees of freedom that will characterize a typical object through observed shape variations across a population. The target object in a training image I_t can be recognized through an image segmentation S_t , which can be achieved via an automated image registration of I_t to the pre-labeled template image I_0 . The large-scale fluid deformation method [4] is used in this work and I_0 is registered to I_t through a diffeomorphic function H_t :

$$I_t \rightarrow H_t(I_0) . \quad (1)$$

It is assumed here that the correspondence between these images established by $\{H_t\}$ is the “true correspondence”. With the target object in I_0 having been labeled *a priori*

(denoted as S_0), that in I_i (denoted as S_i) can thus be defined by applying the same warp function H_i on S_0 ,

$$S_i = H_i(S_0) . \quad (2)$$

Here the “ground truth” S_0 is defined by the consensus of an expert panel.

2.2 Fiducial-Point Shape Representation

The shape of an object is defined as those aspects of its geometric conformation that are invariant under similarity transformations, for example, its topology. A fiducial-point shape representation made from a limited number of points on the surface will be developed next to represent an object’s shape according to its variation in the training set. The shape analysis then becomes the study of the displacement distributions of these points.

Given a predefined template shape S_0 , and a set $\{H_i\}$ that registers I_0 to $\{I_i\}$, the various shapes of the object of interest in $\{I_i\}$ can be represented by $(S_0, \{H_i\})$. Assume that each H_i can be encoded into the displacements of a limited number of points, from which another deformation field H'_i can be constructed to reproduce the deformation around H_i (S_0). Then the representation of the various shapes is further simplified to the locations of these individual points and their corresponding displacements. The algorithm employed here to construct H'_i from a set of points and their displacements has been developed under the fluid deformation model and has been implemented as the “landmark fluid deformation” function in the fluid deformation program [5]. This function takes the initial and terminal locations of a set of points as the input, and generates a diffeomorphic deformation vector field that drives each point from its initial location to its destination.

This set of points on the surface of S_0 is named the “fiducial points”, denoted by $\{F_m\}$, ($m \in [1, M]$, M being the total number of fiducial points in the representation). In the extreme case of including all points on the surface of S_0 , we would have an enumerative representation of the surface. However, since the locations of these points are highly correlated due to the biological nature of the target object, some of these points are redundant in this representation and can be eliminated without losing significant information. The fiducial points predominantly reside at salient geometric and/or intensity features of an image.

An iterative procedure has been developed to find an optimal set of fiducial points in representing the shape of an anatomical object adequately and efficiently such that, on the average over all training samples, the difference between H_i and H'_i

$$\frac{1}{T \square N_0} \sum_{t=1}^T \sum_{x \in S_0} |H_i(x) \square H'_i(x)| \quad (3)$$

is reduced below a predefined threshold on the surface of S_0 , where N_0 is the number of voxels on the surface of S_0 and T is the total number of training samples. A surface curvature based geometric screening can be employed to initialize the seed point set – assuming large surface curvature implies important shape conformation, the first few points with the highest surface curvature are selected to define the initial fiducial

point set $\{F_m\}$. Then the following algorithm can be applied to each I_t ($t \in [1, T]$) to find the optimal fiducial point selection.

1. Apply the training warp function H_t on $\{F_m\}$ to get the warped fiducial points: $F_{m,t} = H_t(F_m)$;
2. Reconstruct the diffeomorphic warp field H'_t for the entire image volume based on the displacements $\{F_{m,t} - F_m\}$;
3. Locate the point p_t on the surface of S_0 that introduces the largest discrepancy between H_t and H'_t : $H_t(p_t) - H'_t(p_t) = \max\{ |H_t(x) - H'_t(x)| \}$, where $p_t, x \in$ the surface of S_0 ;
4. In the point set $\{p_t\}$ established from all training images, find the point p that introduces the largest discrepancy in $\{H_t(p) - H'_t(p)\}$ and add it to the fiducial point set: $\{\{F_m\}, p\} \forall \{F_m\}$;
5. Loop back to step 1 until an optimization criterion is reached.

The optimization criterion can take various forms and one example will be demonstrated in Sect. 3.

2.3 The Normal Shape Domain

Denote $\vec{f}_0 = ((x_{f_1}, y_{f_1}, z_{f_1}), (x_{f_2}, y_{f_2}, z_{f_2}), \dots, (x_{f_M}, y_{f_M}, z_{f_M}))$ the collective coordinates of the fiducial point set $\{F_m\}$, and $\vec{f}_t$ the displaced fiducial points $\{F_{m,t}\}$ ($t = 1, 2, \dots, T$). So

$$\vec{f}_t = H_t(\vec{f}_0). \quad (4)$$

The object's shape in I_t can be represented by $\vec{f}_t$ and the subsequent shape analysis will be on $\{\vec{f}_t\}$.

$\vec{f}_t$ is a single point in a $3M$ dimensional space (M being the total number of fiducial points in the shape representation). A set of T training shapes gives a cloud of T points in this $3M$ -D space. Assume that these points lie within some region of the space, which is named the "normal shape domain" (NSD). Every $3M$ -D point within this domain will give a shape definition $\vec{f}$ that complies with those in the original training set. Thus new shapes can be generated systematically by sampling in this NSD. Also assuming that the NSD is approximately ellipsoidal, we can calculate its center and its major axes to give an analytical definition.

For a set of T training samples represented by $\{\vec{f}_t\}$, the mean shape is $\langle \vec{f} \rangle$. The principal axes of the ellipsoid can be calculated by applying a principal component analysis to the deviation in the data from the sample mean. Each principal axis gives a significant mode of variation, in which the fiducial points tend to move collectively as the shape varies. These principal axes are described by the unit eigenvector $\{\vec{p}_k\}$ ($k = 1, 2, \dots, 3M$) of the covariance matrix S formed from the deviations:

$$S\vec{p}_k = \sqsubset_k \vec{p}_k . \quad (5)$$

Most of the variation can usually be explained by a small number of modes n , and the original ellipsoid has a relatively small width beyond these n dimensions.

2.4 Realistic Synthetic Image Generation

Each point in the NSD can be reached by taking the mean shape and adding a linear combination of the eigenvectors. Therefore, a new shape complying with those training samples is

$$\vec{f}_s = \langle \vec{f} \rangle + \sum_{k=1}^n b_k \sqsubset_k \vec{p}_k , \quad \sqsubset_k \in [3\sqrt{\lambda_k}, 3\sqrt{\lambda_k}] \quad (6)$$

where b_k is the component coefficient of $\vec{p}_k$ and suitable limits are typically of the order of three standard deviations. From the original and the displaced fiducial points $(\vec{f}_0, \vec{f}_s)$, a deformation field (H_s) can be obtained by applying the landmark fluid deformation function. A synthetic image I_s and its “ground truth” segmentation will then be defined by

$$\begin{aligned} I_s &= H_s(I_0) \\ S_s &= H_s(S_0) \end{aligned} \quad (7)$$

A large number of realistic, synthetic images with known ground truth segmentations can be thus produced and used for characterizing the performance of the image segmentation methods.

3 Application: Synthetic Images of the Human Right Kidney

A pool of 36 clinical CT images $\{I_i\}$ was prepared and one of them was selected as the template I_0 , which has clear kidney region, low noise, minimal breathing artifacts and is close to the mean geometric conformation of all the training samples. The right kidney in I_0 was then carefully segmented by medical experts and was used as the ground truth S_0 . For each training image I_i , its intensity was normalized to that of the template I_0 , and its resolution was resampled to 2mm·2mm·2mm via trilinear interpolation. To improve the convergence of registration, the MIRIT program [6] was applied first to find the similarity transformation between each (I_0, I_i) pair. As image quality varies, parameters in the fluid deformation program were fine-tuned in each case to give the best registration between I_0 and I_i , as judged by visual inspection.

3.1 Fiducial Point Shape Representation

The algorithm developed in *Sect. 2.2* was applied to find the optimal point set for this data. The optimization criterion was taken from a previous independent study [7], in which a comparison was made between two human raters’ manual segmentations on a

subset of the human kidney CT images used in this work. It was found that the average volume overlap between these two manual segmentations was 94.0%, and the average surface distance was 1.2mm apart, as measured by VALMET [1]. Thus the iterative procedure was terminated when, on the average over all training samples, the similarity between $H_i(S_0)$ and $H'_i(S_0)$ surpassed the statistics from the human raters' comparison. In the fiducial point set initialization, a geometric screening selected the top 32 points with the highest surface curvature. The progress of optimization is graphed in Fig. 1.

Fig. 1. Fiducial-point set optimization for the human right kidney

In this chart, each of the three statistical indices has been averaged over all training samples. Both surface distance measurements are the averaged magnitude over the surface voxels. The curve labeled “ $\langle H_i - H'_i \text{ Surface Distance} \rangle$ ” shows the average point-wise (diffeomorphisms) error between H_i and H'_i on the surface of kidney. The curve labeled “ $\langle \text{Closest Surface Distance} \rangle$ ” shows the averaged nearest surface distance from every point on surface of $H_i(S_0)$ to that of $H'_i(S_0)$, so it is necessarily lower than the previous one. The curve in black gives the volume overlap between these two warped segmentations. The last two indices are calculated by the VALMET program, which compares and evaluates a pair of segmentations. Both distance measurements are in voxel units, and their ordinate is on the left. The volume overlap is measured in percentage, and its ordinate is on the right. It is seen that as more points are added into the fiducial-point set, the average volume overlap increases monotonically and the distance measurements decrease correspondingly.

In representing the human right kidney, the volume overlap index has already exceeded 90% when only about 30 fiducial points are selected. Intuitively, this few points can hardly represent a shape in detail even for a simple 3D object. In segmentation comparison, the volume overlap measurement alone is sometimes misleading and should be used with great caution. We therefore report the average surface distance index in combination with this volume overlap index, and we sometimes even include the quartile statistics for more detailed analysis.

Applying the optimization criteria, the fiducial point shape representation for the human right kidney was established to comprise 88 points on the surface of S_0 , with $\langle \text{Volume Overlap} \rangle = 94.2\%$, $\langle \text{Closest Surface Distance} \rangle = 0.591$, and $\langle H_i - H'_i \text{ surface distance} \rangle = 0.743$.

3.2 Synthetic Image Generation

MATLAB was used for the principal component analysis. Although the covariance matrix in this case was a 264-D ($3 \cdot 88=264$) square matrix, only the first 35 modes were non-trivial and the NSD in this model case was defined by the first seven modes. They covered more than 88% of the total variance observed in the training samples. The validity of the NSD can be assessed by cross-validation, which shows that this NSD is able to explain the shape variations observed in these samples. Due to the page limit, the technique and the result will be reported in a separate publication.

An arbitrary large number of synthetic human kidney CT images can now be produced from the current system. One such image is shown in *Fig. 2*, in comparison with the template. Note that the kidney object in the synthetic image takes a different shape from that of the template, but their background conformations (structures like liver, lung, spinal cord, and soft tissues) are very similar. Also, the intensity contrast and the noise level in these images are almost identical.

Fig. 2. The template (top) and one synthetic image (bottom)

4 Discussion & Conclusion

In the model system of the human right kidney, only the kidney organ is modeled as an object (defined by S_0) whereas all the other structures in the image are treated as the background intensity variation without any shape description. In order to produce truly realistic synthetic images for a large, multi-organ complex (for example, the entire abdominal region), the shape of each object in the complex needs to be studied to define an ensemble NSD of the system.

Also, because one single image I_0 is used as the template in this example, all the synthetic images will inevitably share a similar background and intensity profiles. To alleviate this situation, a template transformation procedure can be employed – in generating a synthetic image I_s , a transformed warp function $H_s H_i^{-1}$ can be used in place of H_s , with I_i as the template image instead of I_0 . The characterization of this

process is currently under investigation, and the result will be reported in a future publication.

In summary, a methodology has been developed to generate realistic, synthetic medical images via the Monte Carlo simulation for characterizing the performance of the image segmentation methods. As a demonstration, it has been applied to produce the synthetic CT images of the human right kidney. A test of the m-rep deformable model segmentation method [7] is currently in progress and its result will be reported in the next publication.

Acknowledgements

We sincerely appreciate the help and input from our colleagues during the development of this research work, with our special thanks to Dr. Keith Muller for the many insightful discussions on the statistical analysis and to Gregg Tracton for his superior technical support. The research reported here was carried out under the support of National Cancer Institute Grant P01 CA47982.

References

1. Gerig, G., M. Jomier, M. Chakos (2001). "Valmet: A new validation tool for assessing and improving 3D object segmentation." Proc. MICCAI 2001, Springer LNCS 2208: 516-523.
2. Cootes, T. F., A. Hill, C.J. Taylor, J. Haslam (1994). "The Use of Active Shape Models for Locating Structures in Medical Images." Image and Vision Computing **12**(6): 355-366.
3. Rueckert, D., A.F. Frangi, and J.A. Schnabel (2001). "Automatic Construction of 3D Statistical Deformation Models Using Non-rigid Registration." MICCAI 2001, Springer LNCS 2208: 77-84.
4. Christensen, G. E., S.C. Joshi and M.I. Miller (1997). "Volumetric Transformation of Brain Anatomy." IEEE Transactions on Medical Imaging **16**: 864-877.
5. Joshi, S., M.I. Miller (2000). "Landmark Matching Via Large Deformation Diffeomorphisms." IEEE Transactions on Image Processing.
6. Maes, F., A. Collignon, D. Vandermeulen, G. Marchal, P. Suetens (1997). "Multi-Modality Image Registration by Maximization of Mutual Information." IEEE-TMI **16**: 187-198.
7. Pizer, S.M., J.Z. Chen, T. Fletcher, Y. Fridman, D.S. Fritsch, G. Gash, J. Glotzer, S. Joshi, A. Thall, G. Tracton, P. Yushkevich, and E. Chaney (2001). "Deformable M-Reps for 3D Medical Image Segmentation." IJCV, submitted.