A Validation Benchmark for Assessment of Medial Surface Quality for Medical Applications

Agnés Borràs, Debora Gil, Sergio Vera, and Miguel Angel González

Computer Vision Center, Comp. Science Dep. UAB, Spain
Alma I.T. Systems, Barcelona, Spain
`{agnesba,debora}@cvc.uab.es,`
`{sergio.vera,miguel.gonzalez}@alma3d.com`

Abstract. Confident use of medial surfaces in medical decision support systems requires evaluating their quality for detecting pathological deformations and describing anatomical volumes. Validation in the medical imaging field is a challenging task mainly due to the difficulties for getting consensual ground truth. In this paper we propose a validation benchmark for assessing medial surfaces in the context of medical applications. Our benchmark includes a home-made database of synthetic medial surfaces and volumes and specific scores for evaluating surface accuracy, its stability against volume deformations and its capabilities for accurate reconstruction of anatomical volumes.

Keywords: Medial Surfaces, Shape Representation, Medical Applications, Performance Evaluation.

1 Introduction

Medial manifolds (skeletons in 1D, medial surfaces in 2D) are powerful descriptors of shapes that have been used in many computer vision areas such as object recognition, computer graphics, animation or shape modelling [23,10,2]. Their associated medial representation [1] defines a coordinate system for volumetric shapes that allows easy localization of internal points from the volume boundary surface. Given the usefulness of such a coordinate system, in recent years several medial representations have been adapted to medical imaging tasks:

- *Localization of injured tissue.* The radial coordinate of medial representations allows parameterizing [7,16] the (possibly diseased) parenchyma of organs, as well as their internal vascular system, powerful sources of information in organ functionality, analysis and diagnosis
- *Segmentation of medical images.* Techniques such as M-Reps [4] and CM-Reps [25] have shown the potential to describe complex shapes in a versatile manner. Using information of a medial surface for medical imaging segmentation has proven to improve segmentation results [12,19]. It follows that deformable medial modelling has been used in a variety of medical imaging analysis applications, including computational neuroanatomy [27,17], 3D cardiac modelling [18] or cancer treatment planning [15,3].

– *Anatomy Modelling.* In shape analysis, medial representations can model not only the shape but also the interior variations [26]. Medial manifolds of organs have proved robust and accurate to study group differences in internal structures of the brain [17,16]. They also provide more intuitive and easily interpretable representations of complex organs [24] and their relative positions [11].

In order to be suitable for the above listed tasks, anatomical medial manifolds should satisfy some requirements. First, they should be simple enough to allow an easy generation of the radial coordinate, as well as, provide a good representation of organs positions and shapes. A main requirement for a confident representation of shapes valid for clinical applications is that medial manifolds present branches that correspond to changes in object boundary convexity [5]. In this manner, they could be useful for detecting pathological components [22]. Finally, medial surfaces should reach a compromise between simplicity and a satisfactory reconstruction of the whole volume. In particular, they should ensure that finest details on the organ boundary are preserved in order to allow early identification of pathological deformations.

Although there are plenty of approaches addressing computation of medial surfaces, existing methods often generate useless spikes or loose connectivity at main branches. A main concern is that validation of accuracy of medial surfaces for medical applications lacks of a solid benchmark. Validation in the medical imaging field is a delicate issue due to the difficulties for generating ground truth data and quantitative scores valid for reliable application to clinical practice.

We propose a benchmark for evaluating medial surface quality in the context of medical applications. The benchmark is divided in three tests. The first test evaluates the quality of the medial surface generated, the second one checks medial branch stability for detection of volume deformations and the third one explores the capabilities of the generated surfaces to recover the original volume and describing anatomical structures. We have applied our benchmark to two representative algorithms for medial surface computation in order to illustrate its performance.

2   Validation Benchmark

In order to address the representation of organs for medical use, medial representations should achieve a good reconstruction of the full anatomy and guarantee that the boundaries of the organ are reached from the medial surface. Given that small differences in algorithm criteria can generate different surfaces, we are interested in evaluating the quality of the generated manifold as a tool to recover the original shape. In this context, a validation benchmark should cover three topics:

1. **Medial Surface Accuracy.** Representations of the original anatomical geometry are accurate as far as the extracted medial manifold satisfies three main properties [13]: 1) preservation of the topology of the original object

(homotopy); 2) one-pixel thin structures (thinness); 3) structures equidistant to object boundaries (medialness). A first test should evaluate to what extent medial surfaces satisfy these 3 quality requirements.

2. **Medial Surface Stability.** A main requirement for a confident representation of shapes valid for clinical applications is the stability of medial manifolds under perturbations of the object boundary [14]. In order to fully describe anatomical shapes, medial manifolds should present branches that correspond to changes in object boundary convexity [5]. A second test should evaluate the stability of the medial surface branches for known volumes undergoing a controlled deformation.

3. **Medial Surface Reconstruction Power.** For a confident application in medical applications, medial representation have to achieve a good reconstruction of the full anatomy and guarantee that the boundaries of the organ are reached from the medial surface [22]. Therefore, a third test should assess the capabilities of the generated surfaces to recover the original volume and describing anatomical structures.

In order to illustrate our validation benchmark, we have applied it to two representative methods of current approaches to medial surface computation. The ridge based method (labelled $GSM2$) described in [21] and the morphological thinning approach (labelled $Th26P$) described in [13].

2.1   Medial Surface Accuracy

Surface quality tests start from known medial surfaces, that will be considered as ground truth. From these surfaces, volumetric objects can be generated by placing spheres of different radii at each point of the surface. The newly created object is the input to medial surface algorithms, which output is compared to the surfaces used to generate synthetic volumes.

The test set of synthetic volumes / surfaces aims to cover different key aspects of medial surface generation (see first row in Fig.1). The first batch of surfaces (labelled 'Simple') includes objects generated with a single medial surface. A second batch of surfaces is generated using two intersecting medial surfaces (labelled 'Multiple'), while a last batch of objects (labelled 'Homotopy') covers shapes with different number of holes. We have used a 3D modelling software called 3D Studio Max to produce the ground truth medial sufaces containig several deformations and holes. Surfaces are exported to voxel format in Matlab where the volumes are computed.

The volumetric object obtained from a surface can be generated either using spheres of uniform radii (identified as 'UnifDist') or spheres of varying radii (identified as 'VarDist'). Volumes are constructed by assigning a radial coordinate to each medial point. In the case of UnifDist, all medial points have the same radial value, while for VarDist they are assigned a value in the range $[r_1, r_2]$ using a polynomial. The values of the radial coordinate must be in a range ensuring that volumes will not present self intersections. Therefore, the maximum range and procedure this radius is assigned depends on the medial topology:

- *Simple.* In this case, there are no restrictions on the radial range.
- *Multiple.* For branching medial surfaces, especial care must be taken at surface self-intersecting points. At these locations, radii have to be below the maximum value that ensures the medial representation defines a local coordinate change [8]. This maximum value depends on the principal curvatures of the intersecting surfaces [8] and it is computed for each surface. Let X be the medial surface, Z denote the self-intersection points and $d(Z)$ the distance map to Z. The radial coordinate is assigned as follows:

$$R(X) = min(R(X), max(r_Z, d(Z)))$$

  for $R(X)$ the value of the polynomial function and r_Z the maximum value allowed at self-intersections. In this manner, we obtain a smooth distribution of the radii ensuring volume integrity.
- *Homotopy.* In order to be consistent with the third main property of medial surfaces [13], volumes must preserve all holes of medial surfaces. In order to do so, the maximum radius r_2 is set to be under the minimum of all surface holes radii. In the case that the medial surface contains multiple branches, r_2 is also set to be under the radius of the self-intersection points r_Z.

Our database, which is publicly available [6], has a total number of 120 samples, distributed in 6 families (20 samples each) covering the 3 medial topologies and the 2 volume distance types. Figure 1 shows an example of the synthetic volumes in the first row (labelled GT). Columns exemplify the different families of volumes generated: one (Simple in 1st and 2nd columns) and two (Multiple in 3rd and 4th columns) foil surfaces, as well as, surfaces with holes (Homotopy in 5th and 6th columns). For each kind of topology we show a volume generated with constant (1st, 3rd and 5th columns) and variable distance (2nd, 4th and last columns). We show medial surfaces in solid meshes and the synthetic volume in semi-transparent color.

The quality of medial surfaces has been assessed by comparing them to ground truth surfaces in terms of surface distance [9]. The distance of a voxel y to a surface X is given by: $d_X(y) = \min_{x \in X} \|y - x\|$, for $\| \cdot \|$ the Euclidean norm. If we denote by X the reference surface and Y the computed one, the scores considered are:

1. *Standard Surface Distances*:

$$AD = \frac{1}{\#Y} \sum_{y \in Y} d_X(y) \quad MD = \max_{y \in Y} (d_X(y))$$

   for $\#$ the number of elements of a set.
2. *Symmetric Surface Distances*:

$$ASD = \frac{1}{\#X + \#Y} \left(\sum_{x \in X} d_Y(x) + \sum_{y \in Y} d_X(y) \right)$$

	Simple		Multiple		Homotopy	
	UnifDist	VarDist	UnifDist	VarDist	UnifDist	VarDist
GT						
Th26P						
AD, MD	0.6, 5.5	3.3, 16.2	0.7, 5.6	1.4, 10.7	0.6, 5.5	1.3, 12.1
ASD, MSD	0.5, 5.5	1.9, 16.1	0.5, 5.6	1.1, 11.1	0.4, 5.7	0.8, 11.0
GSM2						
AD, MD	0.3, 3.0	0.3, 4.6	0.4, 3.6	0.4, 4.8	0.4, 3.4	0.3, 3.7
ASD, MSD	0.3, 3.1	0.3, 4.1	0.4, 4.1	0.4, 4.8	0.3, 3.4	0.3, 3.7

Fig. 1. Representative examples of the data base medial surfaces and the medial accuracy validation

$$MSD = \max\left(\max_{x \in X}(d_Y(x)), \max_{y \in Y}(d_X(y))\right)$$

All distance scores are in the range $[0, \inf)$, being 0 the best matching.

Standard distances measure deviation from medialness, while differences between standard and symmetric distances indicate the presence of homotopy artifacts and presence of unnecessary medial segments. Figure 1 shows an example of the computed medial surfaces using $GSM2$ and $Th26P$, as well as, their quality scores for the shown surfaces. The visual quality of the morphological $Th26P$ is worse by the presence of multiple spikes. We note that extra spikes in surfaces are detected by higher distance scores.

2.2 Medial Surface Stability

Stability of medial surfaces is checked by assessing that their branches correspond to changes in object boundary convexity for known volumes undergoing a controlled deformation. The volumes generated for assessment of medial surface accuracy have been deformed in order to generate branches at specific sites. These sites are points selected among a triangular mesh of the volume boundary. For each point, we displace its position, $\overrightarrow{P}$, a given distance, δ_P, along the boundary normal direction at the point, $\overrightarrow{N}_P$:

$$\overrightarrow{P} \rightarrow \overrightarrow{P} + \delta_P \overrightarrow{N}_P$$
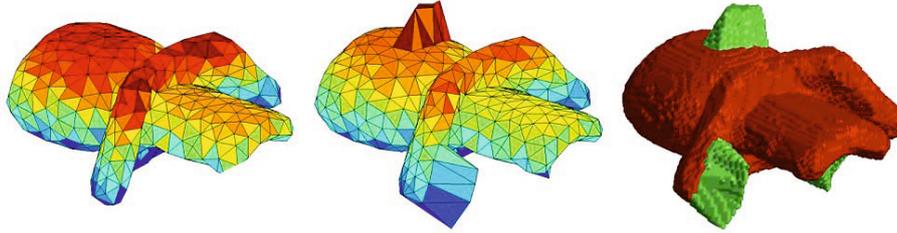
Fig. 2. Generation of volume spikes from original ground truth volumes

for $\delta_P \in [0, D_{max}]$. It should be clear that volume spikes should generate a medial branch if their height is large enough to introduce a significant change in volume boundary curvature. Therefore, normal distances are gradually increased from 0 to a maximum distance D_{max} in order to check the critical deformation that generates a new branch. Given that for $\delta_P = 0$, we have the original volumes, the connected components of the difference between volumes for $\delta_P = 0$ and $\delta_P > 0$ is the collection of volume spikes, namely $\mathcal{VS} = \{VS_i\}_{i=1}^{N_{VS}}$ generated by the deformation process. Figure 2 shows an example of the volume deformation process. The left mesh shows an original ground truth volume which has been deformed to the mesh shown in the middle. The most-right image shows the deformed volume with its spikes in green and the volume for $\delta_P = 0$ in red.

Medial surfaces for $\delta_P = 0$ give the baseline accuracy by comparison to the database ground truth surfaces [20]. For $\delta_P > 0$, computed medial surfaces should generate new branches for each volume spike if the deformation size δ_P is large enough to introduce a significant change in volume curvature. Branches not arising from volumetric spikes changing boundary convexity profile are useless and should as least as possible.

The quality of medial branching arising from volumetric spikes has been assessed in terms of spike detection and its accurate localization. Branches arising from the volume deformation are given by the connected components of the difference between medial surfaces for $\delta_P = 0$ and $\delta_P > 0$. We will note them by $\{B_j\}_{j=1}^{N_B}$. Spike detection rate has been measured in terms of medial branch false and true positives A branch is considered a true positive if it intersects any of the volume spikes VS_i. In order to measure the impact of false branches arising during volume deformation (i.e. detection instability), we have also considered the percentage of area that false positives represent over all medial branches:
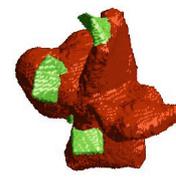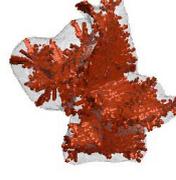
1. *Detection Rates*:
$$TB = \frac{\#\{VS_i \text{ s. t. } \exists B_j, \; B_j \bigcap VS_i \neq \emptyset\}}{N_{VS}}$$

2. *Detection Instability*:
$$DIA = 100 \frac{\sum_{B_j \bigcap VS_i = \emptyset} \|B_j\|}{\sum_{j=1}^{N_B} \|B_j\|}$$

for $\|\cdot\|$ denoting the area of a surface.

GT				
Th26P				
TB, DIA	1, 15%	1, 3.5%	0.75, 10.1%	1, 7.6%
ADL, MDL	1.8, 17.2	0.9, 5.1	1.6, 14.3	1.0, 14.5
GSM2				
TB, DIA	1, 2.0%	1, 2.4%	1, 3.6%	0.75, 5.7%
ADL, MDL	0.9, 8.6	0.9, 5.1	1.4, 8.8	0.8, 6.0

Fig. 3. Assessment of medial stability. First row shows the deformed synthetic volumes, 2nd and 3rd computed medial surfaces and stability scores.

TB is the range $[0, 1]$, with best performance achieved for $TB = 1$ and DIA is in the range $[0, 100]$ with best value for 0%.

We define spike localization in terms of the distance to the volume spikes, namely d_{VS}, and the ground truth medial surfaces, namely d_X. For each point in computed medial surfaces $y \in Y$, we have that the minimum between $d_X(y)$ and $d_{VS}(y)$ reflects a compromise between medial branches size and its proximity to a volume spike. Let $DL(y) := min(d_X(y), d_{VS}(y))$ denote such minimum. Then, our localization scores are given by the average and maximum values of DL over the computed surface:

1. *Spike Localization*:

$$ADL = \frac{1}{\#Y} \sum_{y \in Y} DL(y) \quad MDL = \max_{y \in Y} DL(y)$$

Spikes are best localized when ADL and MDL take value 0.

Figure 3 illustrates assessment of medial branch stability for $GSM2$ and $Th26P$. The detection rate TB drops as a main volume spike is lost as it clearly illustrates the last surface computed using $GSM2$. Meanwhile, DIA scores perfectly agree with the visual quality of surfaces and increase in the presence of

$Th26P$ extra spikes. Finally, distance scores also detect extra structures not associated to a volume spike.

2.3 Reconstruction Power for Clinical Applications

In medical imaging applications the aim is to generate the simplest medial surface that allows recovering the original volume without losing significant voxels. Volumes recovered from surfaces generated with the different methods are compared with ground truth volumes. Volumes are reconstructed by computing the medial representation [1] with radius given by the values of the distance map on the computed medial surfaces.

Let A, B be, respectively, the original and reconstructed volumes and ∂A, ∂B, their boundary surface. Completeness of reconstructed volumes is assessed using the following volumetric and distance measures:

1. *Volume Overlap Error:*

$$VOE(A, B) = 100 \times \left(1 - 2\frac{\|A \cap B\|}{\|A\| + \|B\|} \right)$$

2. *Maximum Volume Boundary Difference:*

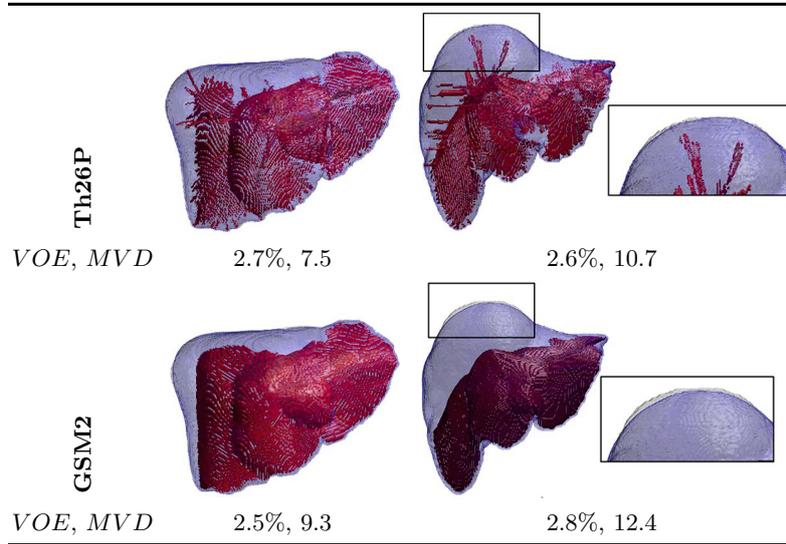$$MVD = \max \left(\max_{x \in \delta A}(d_{\delta B}(x)), \max_{y \in \delta B}(d_{\delta A}(y)) \right)$$

Fig. 4. Reconstruction Power for Clinical Applications

The minimum error in the reconstruction is observed when VOE and MVD are 0, where $VOE \in [0, 100]$ and $MVD \in [0, \inf)$. We would like to note that reconstruction scores do not require a ground truth for medial surfaces, only volumes. Therefore they can be computed over any database.

Figure 4 illustrates assessment of reconstruction power for clinical applications. In order to provide a real scenario for the reconstruction tests we have used livers from the SLIVER07 challenge [9] as a source of anatomical volumes. Volumes reconstructed using the computed medial surfaces (colored in red) are shown in transparent blue over true anatomical volumes shown in transparent gray. Difference between reconstructions and original volumes is better appreciated in the right image close-ups of the liver lobe. Gross differences between volumes are detected by VOE and, in spite of the right liver lobe, none of the cases seem to be significantly better. In medical applications, restoring local deformations can be important for early diagnosis. In this context, the surface distance score MVD is suitable for detection of local differences.

3   Conclusions

We have presented a complete benchmark for evaluating medial surface generation in the context of shape description. Our benchmark includes a battery of synthetic medial surfaces and volumes that cover different medial topologies and volume deformations. We have also defined several scores for measuring 3 different quality aspects: accuracy, stability and reconstruction power. Our benchmark has been applied to 2 representative methods for medial computation. Results show that the proposed scores and methodologies agree with the visual quality of surfaces and, thus, they are valid for quantitative systematic evaluation.

References

1. Blum, H.: A transformation for extracting descriptors of shape. MIT Press (1967)
2. Chang, M., Leymarie, F., Kimia, B.: 3d shape registration using regularized medial scaffolds. In: Proc. 3DPVT, pp. 987–994 (2004)
3. Crouch, J.R., Pizer, S.M., Chaney, E.L., Hu, Y.-C., Mageras, G.S., Zaider, M.: Automated Finite-Element analysis for deformable registration of prostate images. IEEE Transactions on Medical Imaging 26(10), 1379–1390 (2007)
4. Fletcher, P.T., Lu, C., et al.: Principal Geodesic Analysis for the study of nonlinear statistics of shape. IEEE Trans. Med. Imag. 23(8), 995–1005 (2004)
5. Giblin, P.J., Kimia, B.B., Pollitt, A.J.: Transitions of the 3d medial axis under a one-parameter family of deformations. PAMI 31(5), 900–918 (2009)
6. Gil, D., Borràs, A.: Medial surfaces database, http://iam.cvc.uab.es/downloads/medial-surfaces-database (accessed: April 25, 2013)
7. Gil, D., Garcia-Barnes, J., Hernandez, A.A.: Manifold parametrization of the left ventricle for a statistical modelling of its complete anatomy. In: SPIE 2001. LNCS, pp. 304–314 (2010)

8. Gray, A.: Tubes. Birkhäuser (2004)
9. Heimann, T., van Ginneken, B., Styner, M.A., Arzhaeva, Y., Aurich, V.: Comparison and evaluation of methods for liver segmentation from CT datasets. IEEE Trans. Med. Imag. 28(8), 1251–1265 (2009)
10. Keim, D., Panse, C., North, S.C.: Medial-axis-based cartograms. IEEE Comput. Graph. Appl. 25, 60–68 (2005)
11. Liu, X., Linguraru, M.G., Yao, J., Summers, R.M.: Organ pose distribution model and an MAP framework for automated abdominal multi-organ localization. In: Liao, H., "Eddie" Edwards, P.J., Pan, X., Fan, Y., Yang, G.-Z. (eds.) MIAR 2010. LNCS, vol. 6326, pp. 393–402. Springer, Heidelberg (2010)
12. Pizer, S.M., Fletcher, P.T., et al.: A method and software for segmentation of anatomic object ensembles by deformable M-Reps. Medical Physics 32(5), 1335–1345 (2005)
13. Pudney, C.: Distance-ordered homotopic thinning: A skeletonization algorithm for 3D digital images. Comp. Vis. Imag. Underst. 72(2), 404–413 (1998)
14. Siddiqi, K., Zhang, J., Macrini, D., et al.: Retrieving articulated 3-d models using medial surfaces. Mach. Vis. Appl. 19(4), 261–275 (2008)
15. Stough, J.V., Broadhurst, R.E., Pizer, S.M., Chaney, E.L.: Regional appearance in deformable model segmentation. In: Karssemeijer, N., Lelieveldt, B. (eds.) IPMI 2007. LNCS, vol. 4584, pp. 532–543. Springer, Heidelberg (2007)
16. Styner, M., Gerig, G., Lieberman, J., Jones, D., Weinberger, D.: Statistical shape analysis of neuroanatomical structures based on medial models. Med. Im. Ana. 7(3), 207–220 (2003)
17. Styner, M., Lieberman, J.A., Pantazis, D., Gerig, G.: Boundary and medial shape analysis of the hippocampus in schizophrenia. Medical Image Analysis 8(3), 197–203 (2004)
18. Sun, H., Avants, B.B., Frangi, A.F., Sukno, F., Gee, J.C., Yushkevich, P.A.: Cardiac medial modeling and time-course heart wall thickness analysis. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 766–773. Springer, Heidelberg (2008)
19. Sun, H., Frangi, A.F., Wang, H., Sukno, F.M., Tobon-Gomez, C., Yushkevich, P.A.: Automatic cardiac mri segmentation using a biventricular deformable medial model. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part I. LNCS, vol. 6361, pp. 468–475. Springer, Heidelberg (2010)
20. Vera, S., Gil, D., Borras, A., George Linguraru, M., Gonzalez, M.A.: Geometric steerable medial maps. Machine Vision Applications (in press)
21. Vera, S., Gil, D., Borràs, A., Sánchez, X., Pérez, F., Linguraru, M.G., González Ballester, M.A.: Computation and evaluation of medial surfaces for shape representation of abdominal organs. In: Yoshida, H., Sakas, G., Linguraru, M.G. (eds.) Abdominal Imaging. LNCS, vol. 7029, pp. 223–230. Springer, Heidelberg (2012)
22. Vera, S., Gonzalez, M.A., Gil, D.: A medial map capturing the essential geometry of organs. In: Proceedings of ISBI. IEEE (2012)
23. Wade, L., Parent, R.E.: Automated generation of control skeletons for use in animation. The Visual Computer 18, 97–110 (2002)
24. Yao, J., Summers, R.M.: Statistical location model for abdominal organ localization. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part II. LNCS, vol. 5762, pp. 9–17. Springer, Heidelberg (2009)
25. Yushkevich, P.A.: Continuous medial representation of brain structures using the biharmonic PDE. NeuroImage 45(1), 99–110 (2009)
26. Yushkevich, P.A., Zhang, H., Gee, J.C.: Continuous medial representation for anatomical structures. IEEE Trans. Medical Imaging 25(12), 1547–1564 (2006)
27. Yushkevich, P.A., Zhang, H., Simon, T.J., Gee, J.C.: Structure-specific statistical mapping of white matter tracts. NeuroImage 41(2), 448–461 (2008)