

Simulation of Ground-Truth Validation Data Via Physically- and Statistically-Based Warps

Ghassan Hamarneh, Preet Jassi, and Lisa Tang

Medical Image Analysis Lab., Simon Fraser University, Burnaby, BC, V5A 1S6, Canada
{hamarneh, preetj, lisat}@cs.sfu.ca

Abstract. The problem of scarcity of ground-truth expert delineations of medical image data is a serious one that impedes the training and validation of medical image analysis techniques. We develop an algorithm for the automatic generation of large databases of annotated images from a single reference dataset. We provide a web-based interface through which the users can upload a reference data set (an image and its corresponding segmentation and landmark points), provide custom setting of parameters, and, following server-side computations, generate and download an arbitrary number of novel ground-truth data, including segmentations, displacement vector fields, intensity non-uniformity maps, and point correspondences. To produce realistic simulated data, we use variational (statistically-based) and vibrational (physically-based) spatial deformations, nonlinear radiometric warps mimicking imaging non-homogeneity, and additive random noise with different underlying distributions. We outline the algorithmic details, present sample results, and provide the web address to readers for immediate evaluation and usage.

Keywords: validation, segmentation, deformation, simulation, vibration, variation, non-uniformity.

1 Introduction

Medical images provide a wealth of data about internal anatomy and physiology essential for computer-aided modeling, diagnosis, treatment, and tracking of diseases. This, in turn, imposes high demands for automated, accurate, fast, and robust medical image analysis methods. This need has resulted in a plethora of alternative medical image segmentation, registration, and shape correspondence algorithms. Ironically, the wealth of data, from the ever growing high-dimensional images of millions of pixels and meshes with thousands of vertices, is also the cause of scarcity of ground-truth data sets of labeling, spatial transformation, and point correspondences, on which the medical image analysis algorithms must be evaluated and validated.

Aside from a few exceptions [26][34][35], assessing the performance of image analysis algorithms requires ground-truth data, such as expert-labeled images, physical or computational phantoms with known segments, deformations, or corresponding intrinsic landmarks or external markers [36][33][20]. Some of the notable efforts for providing simulated and ground-truth data include BrainWeb simulated brain MR images [4], the internet brain segmentation repository [15], PET-SORTEO for

simulated PET data [24], and STAPLE for image segmentation validation [31]. Frameworks for evaluating and validating medical image registration techniques include the retrospective image registration evaluation projects [32][12][21][14], the non-rigid image registration evaluation project (NIREP) [3], and others [23][2]. Gerig et al. were the first to assess and visualize differences between multiple segmentations through their publicly available VALMET software [13]. Standard approaches for evaluating segmentation results given ground-truth segmentation include the Hausdorff distance, Dice coefficient [9], and the Jaccard index [16]. More recently, alternative approaches were proposed [25][10][11][30]. Evaluating point correspondence between pairs or within a group of shapes is also of interest in the computational anatomy community, which is primarily based on either geodesic distances between corresponding points found by the algorithm and known ground-truth, or via assessing the statistical shape model’s generality, specificity, or description length [17][28][29][8].

In addition to validating medical image analysis algorithms, large amounts of ground-truth data is important for machine learning and statistical modeling techniques, such as learning relationships (e.g. regression) between shape and appearance [5]. In medical imaging, learning techniques suffer from the problem of high-dimensional, small sample size datasets (“the curse of dimensionality”), in which even the smallest of typical 2D scalar medical images (few hundred pixels, squared) can be seen as samples in tens-of-thousands dimensional space. Clearly, the situation is much more severe for 3D, or 4D (3D+time) data sets with vector (e.g. color or displacements) or matrix (e.g. diffusion tensor) entries.

To the best of our knowledge, none of the existing ground-truth databases or validation methods allows the user to simulate novel data from a single reference dataset (an image, its segmentation, and its landmark points) capturing the exact anatomy on which the developed algorithms need to be trained or validated. In this work, we contribute to addressing the problem of scarcity of ground-truth data through the simulation of an arbitrarily large number of novel ground-truth datasets from a single reference data set and the creation of a web interface to this simulation tool. Although it is difficult to evaluate the validity and realism of the simulated data, we employ a physically- and statistically-based generative model to ensure data realism to a large degree by adopting the formulation proposed in [7]. We also extend [7] to operate on images rather than landmarks along contours of shapes, and extend it from 2D to 3D. While we focus on data simulation for validating or training generic, modality-independent segmentation algorithms, the simulated images will possess appearance characteristics according to the modality of the reference data set.

2 Methods

Combinations of statistically- and physically-based spatial deformations, smoothly varying intensity non-uniformity warps, and random noise are applied to a reference data set which is uploaded by the user through our web-interface, in order to generate new ground-truth data. The reference set consists of a reference image I_i , its corresponding binary or multi-label segmentation S_i , and a set of landmark coordinates L_i . A few user-selected parameters, including the number of desired ground-truth data,

are also supplied as input to the algorithm. The user is then notified via email of the URL from which to download the generated novel ground-truth data.

2.1 Spatial Transformations

Spatial deformations of the image are generated as follows¹. An $M \times N$ uniform grid of control points $\mathbf{x} = \{(cx_{ij}, cy_{ij}); i=1 \text{ to } M; j=1 \text{ to } N\}$ is initialized in the image plane. The control points are displaced as described below and the displacements are interpolated over the image plane. Displacements can be either random, where each component of the displacement vector field is sampled from a uniform random distribution $([-a, a] \times [-a, a])$; from a statistically-driven point distribution model (PDM) [6]; through a physically-based, vibrational model; or a combined model [7] as described below.

2.1.1 Generative Statistically-Based Model

Given a training set of deformed grids of control points, each represented as a $2MN$ -vector, their linear variational modes can be obtained via principal component analysis of their covariance matrix \mathbf{S} [6]. A PDM approximates control point grids as the sum of a mean grid and $t < 2MN$ main modes of variation: $\mathbf{x} = \bar{\mathbf{x}} + \mathbf{P}\mathbf{b}$, where $\bar{\mathbf{x}}$ is a $2MN$ -vector of average locations of the control points and \mathbf{P} is a $2MN \times t$ matrix of principal components. By choosing different weights (within $\pm 3\text{std}$) for the t -vector \mathbf{b} , new control point locations are obtained and used to synthesize new images.

2.1.2 Generative Physically-Based Model

We follow Cootes and Taylor's approach to calculating the vibrational modes of shapes [7]. However, in this work, the coordinates of the grid of control points are treated as the shape's landmarks. Physically-based vibrational modes of the grid are generated through modal analysis of a finite element model (FEM). This gives rise to control points' displacements, which are then used to synthesize new deformed images. The grid of control points are considered nodes with masses interconnected by springs with constant stiffness and with rest lengths equal to the distance between the locations of the control points in the original, undeformed grid. New control point locations are then generated using $\mathbf{x} = \hat{\mathbf{x}} + \mathbf{\Phi}\mathbf{u}$, where $\hat{\mathbf{x}}$ is a $2NM$ -vector of the original locations of the coordinates in the grid, $\mathbf{\Phi}$ is a $2MN$ square matrix of eigenvectors representing the vibrational modes, and \mathbf{u} is a $2MN$ -vector of weights. The matrix $\mathbf{\Phi}$ is the solution of the generalized eigen-system $\mathbf{K}\mathbf{\Phi} = \mathbf{M}\mathbf{\Phi}\mathbf{\Omega}^2$, where \mathbf{K} is a $2MN$ square stiffness matrix calculated as in [7], $\mathbf{M} = \mathbf{I}$ is a $2MN$ square mass matrix, $\mathbf{\Omega}^2 = \text{diag}(\omega_1^2, \omega_2^2, \dots, \omega_{2MN}^2)$ is a matrix of eigenvalues associated with the eigenvectors in $\mathbf{\Phi}$, and ω_i is the frequency of i^{th} vibrational mode.

2.1.3 Combined Vibrational-Variational Model

FEM generates new control point locations through vibrations of a single grid. PDM generates new grids by sampling the allowable space constructed from a training set.

¹ Our simulation can generate 2D or 3D data. We only write the 2D equations for clarity.

We generate new grids with displaced control points similar to the PDM approach but relying on a combined covariance matrix $\mathbf{S}_c = \mathbf{S} + \alpha \left(\frac{1}{m} \sum_{i=1}^m \Phi_i \Lambda \Phi_i^T \right)$, where $\Lambda = \text{diag}(\omega_i^{-2})$ is used to generate more low frequency vibrations and less high frequency, m is the current number of available grid samples, and α balances variational vs. vibrational deformations. Starting with a single uniform grid ($m=1$), $\mathbf{S} = \mathbf{0}$ and hence only vibrational modes generate new shapes. As more grids are available, we gradually reduce the effect of vibrational modes by setting $\alpha = \alpha_1 / m$, where α_1 is a constant.

2.2 Intensity Non-uniformity Via Radiometric Warps

Intensity non-homogeneity is modeled via a smoothly varying, additive intensity field, parameterized by the number U of modes (minima or maxima) in the field. U spatial locations $\{(u x_i, u y_i); i=1, 2, \dots, U\}$ are sampled from a bivariate uniform distribution extending throughout the image domain, or selected on a uniform grid. At each location, a random, uniformly-distributed bias in $[-b:b] \times I_{\max}$ is generated, where I_{\max} is the maximum intensity bias.

2.3 Spatial and Radiometric Interpolation

Different types of interpolation are involved in the simulation of novel ground-truth data. We interpolate the spatial displacements of the gray-level and segmentation image domains, and of landmark coordinates using bi-cubic interpolation. There are several possible interpolation methods for interpolating the pixel intensities from the original spatially un-warped intensity image [19]. To generate the results in this paper, we used bi-linear intensity interpolation in 2D and thin plate splines in 3D. We used nearest-neighbor interpolation for warping the ground-truth labels to avoid erroneous interpolated labels. We interpolate the intensity non-uniformity fields from the intensity biases at the non-uniformity mode centers using bi-cubic interpolation. To resemble ground-truth segmentation, the warped segmentation images are not affected by intensity non-uniformity or by noise.

2.4 Additive Noise

The warped intensity images are deteriorated by additive Gaussian noise [18] or other distributions (e.g. uniform or impulse noise) by specifying the appropriate parameters (e.g. the mean and variance of Gaussian noise).

2.5 Web Simulation Tool and Implementation Details

The simulation code is written in MATLAB (Mathworks, Natwick, MA). MATLAB Compiler for Linux is used to compile M files into stand-alone executables. The web-server interface and MATLAB Component Runtime (MCR) alleviate the need for

MATLAB during run-time. The web interface is created using XHTML, CSS and DHTML. PHP handles the form submissions, uploading the reference images, invoking the simulation code, creating and compressing the appropriate folders and ground-truth files, sending email notifications, and updating an XML file queue to handle the different user submissions. To access the web-based tool, visit <http://mialweb.cs.sfu.ca/>. There, the user selects 2D or 3D simulation, uploads an image (I) and corresponding segmentation (S) and landmark coordinates (L), and specifies simulation parameters. The parameters are: density of the control grid ($M \times N$); extent of random spatial displacement (a); number of variational modes (t), or alternatively, fraction of explained variance; parameter balancing variational vs. vibrational modes (α_1); number of modes for the intensity non-uniformity (U), and whether the mode locations are at uniform grid points; extent of non-uniformity (b); noise parameters (e.g. mean and variance of Gaussian); and number of simulated samples to generate (G). The user is then notified by email of the URL from which the generated files can be downloaded. This data includes G samples of each of the following: spatially warped intensity images; spatially warped segmentation labels; warped intensity + non-uniformity; warped intensity + noise; warped intensity + non-uniformity + noise; the noise field; the non-uniform intensity bias field; and the displacement vector field.

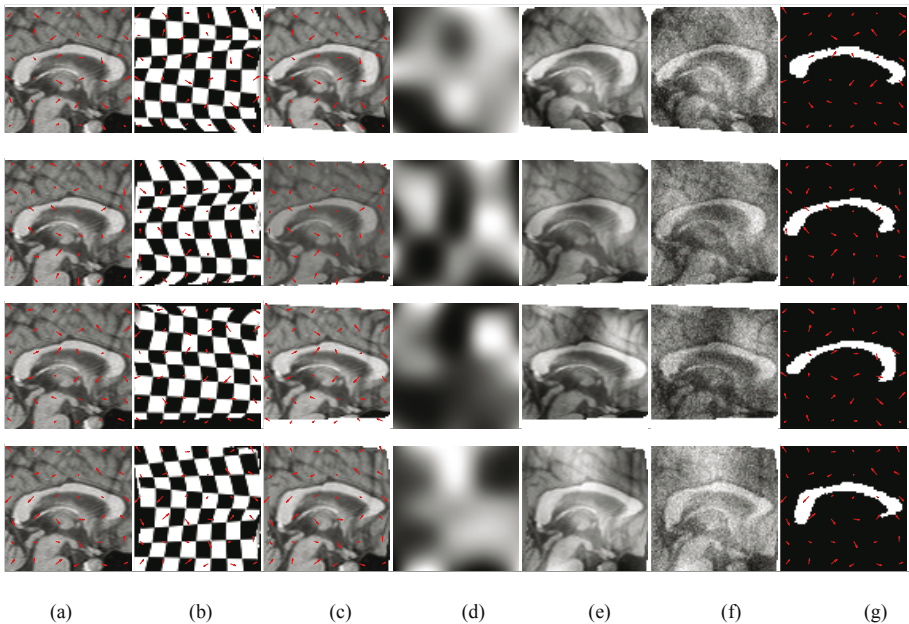


Fig. 1. 2D Simulation: (a) Reference image with different grid displacements (small arrows), (b) deformed checkerboard, (c) deformed image, (d) non-uniformity field, (e) 'c+d', (f) 'e+noise', (g) simulated segmentation (reference segmentation not shown)

3 Results

Figure 1 shows sample simulations of novel 2D MR brain images, deformations and displacement vector fields, segmentations, non-uniformity maps, and noisy images. Figure 2 shows sample simulations of 3D pelvic CT data. Figure 3 shows sample simulations of 2D and 3D ground-truth point correspondences. Figure 4 shows snapshots of the progress of using the simulation web-interface.

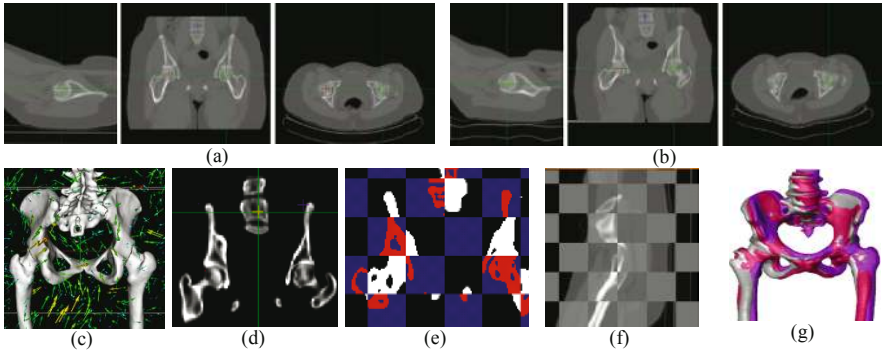


Fig. 2. 3D Simulation: Three orthogonal views of (a) reference CT pelvic volume, (b) one of the simulated volumes, (c) displacement vector field and surface rendering of simulated pelvic and neighboring bone, (d) close-up on simulation result, (e) simulated segmentation (checkerboard overlay with reference segmentation), (f) checkerboard overlay of reference and simulated data, (g) several other simulated datasets shown with different colors

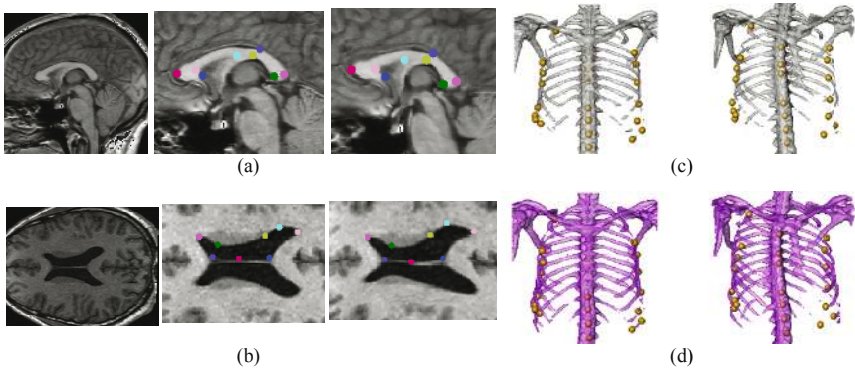


Fig. 3. Simulating 2D and 3D ground-truth point correspondence. (a,b) 2D reference image (left), close up on reference landmarks (middle), and simulated corresponding landmarks (right). (c,d) Two examples of reference thoracic segmentation and landmarks (left) and simulated segmentation and corresponding landmarks (right).

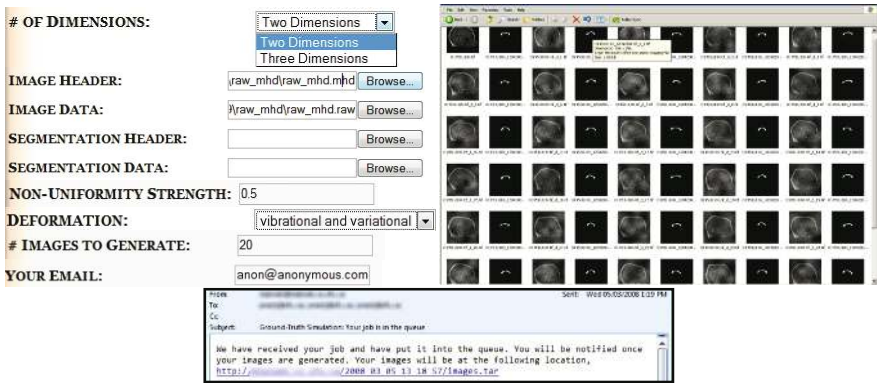


Fig. 4. Simulation progress. The user fills a web-form (sample fields are shown top-left) and receives an email notification (bottom) to download the simulated images (top-right).

4 Conclusions

Lack of sufficient ground-truth data in medical imaging is evident, despite the pressing need to validate a large number of medical image analysis algorithms and to address the high dimension, small sample size problem plaguing machine learning. We present a preliminary proof-of-concept system for simulating ground-truth data, which requires a single reference dataset, generates physically and statistically plausible deformations, applies radiometric warps and noise, and is easily accessed through a web-interface. Although the resulting simulated data may not represent “real” changes, they still appear realistic and are useful for validation, benchmarking, and machine learning (e.g. capturing relationships between spatial warps and intensity variations). Further, the so-called “unrealistic” simulations may stand in lieu of unpredictable, pathological cases, which unfortunately have not been properly addressed by almost all existing algorithms, let alone validated. The work presented here is only a first step towards a more elaborate ground-truth simulator under development. Future work includes adopting advanced models of intensity non-uniformity (e.g. [27][22]), modal analysis reflecting real tissue properties (from the literature or via MR elastography measurements), modality-specific customization of the simulated data (e.g. to simulate diffusion tensor images, one must ensure that the diffusion tensor in each voxel is transformed correctly [1]).

Acknowledgements

We thank Parisa Shooshtari and Omer Ishaq for assistance in code development, James Peltier for assistance in setting up the web server, and NSERC for funding.

References

- [1] Alexander, D., Pierpaoli, C., Basser, P., Gee, J.: Spatial transformations of diffusion tensor magnetic resonance images. *IEEE TMI* 20(11), 1131–1139 (2001)
- [2] Chou, Y., Skrinjar, O.: Ground truth data for validation of nonrigid image registration Algorithms. In: *ISBI*, pp. 716–719 (2004)

- [3] Christensen, et al.: Introduction to the Non-Rigid Image Registration Evaluation Project (NIREP). In: Biomedical Image Registration Workshop, pp. 128–135 (2006)
- [4] Cocosco, C., Kollokian, V., Kwan, R., Evans, A.: BrainWeb: Online Interface to a 3D MRI Simulated Brain Database. *NeuroImage* 5(4), part 2/4, S425 (1997)
- [5] Cootes, T., Edwards, Taylor, C.: Active Appearance Models. *PAMI* 23(6), 681–685 (2001)
- [6] Cootes, T., Taylor, C., Cooper, D., Grahamet, J.: Active Shape Models - Their Training and Application. *Computer Vision and Image Understanding* 61(1), 38–59 (1995)
- [7] Cootes, T., Taylor, C.: Combining point distribution models with shape models based on finite element analysis. *Image and Vision Computing* 13(5), 403–409 (1995)
- [8] Davies, R., Twining, C., Cootes, T., Waterton, J., Taylor, C.: A minimum description length approach to statistical shape Modeling. *IEEE TMI* 21(5), 525–537 (2002)
- [9] Dice, L.: Measures of the Amount of Ecologic Association Between Species. *Ecology* 26(3), 297–302 (1945)
- [10] Everingham, M., Muller, H., Thomas, B.: Evaluating image segmentation algorithms using monotonic hulls in fitness/cost space. *BMVC*, 363–372 (2001)
- [11] Everingham, M., Muller, H., Thomas, B.: Evaluating image segmentation algorithms using the Pareto front. *ECCV (IV)*, 34–48 (2002)
- [12] Fitzpatrick, et al.: Visual assessment of the accuracy of retrospective registration of MR and CT images of the brain. *IEEE TMI* 17, 571–585 (1998)
- [13] Gerig, G., Jomier, M., Chakos, M.: VALMET: A new validation tool for assessing and improving 3D object segmentation. In: Niessen, W.J., Viergever, M.A. (eds.) *MICCAI 2001*. LNCS, vol. 2208, pp. 516–523. Springer, Heidelberg (2001)
- [14] Hellier, et al.: Retrospective Evaluation of Inter-subject Brain Registration. *IEEE TMI* 22(9), 1120–1130 (2003)
- [15] Internet Brain Segmentation Repository,
<http://www.cma.mgh.harvard.edu/ibsr/>
- [16] Jaccard, P.: Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bulletin de la Société Vaudoise des Sciences Naturelles* 37, 547–579 (1901)
- [17] Karlsson, J., Ericsson, A.: A ground truth correspondence measure for benchmarking. In: *ICPR*, pp. 568–573 (2006)
- [18] Kwan, R., Evans, A., Pike, B.: MRI Simulation-Based Evaluation of Image-Processing and Classification Methods. *IEEE TMI* 18(11), 1085–1097 (1999)
- [19] Lehman, T., Gonner, C., Spitzer, K.: Survey: Interpolation Methods in Medical Image Processing. *IEEE TMI* 18(11), 1049–1075 (1999)
- [20] Maintz, J., Viergever, M.: A survey of medical image registration. *MIA* 2(1), 1–36 (1998)
- [21] Maurer, C., Fitzpatrick, J., Wang, M., Galloway, R., Maciunas, R., Allen, G.: Registration of head volume images using implantable fiducial markers. *IEEE TMI* 16(4), 447–462 (1997)
- [22] Pawluczyk, O., Yaffe, M.: Field nonuniformity correction for quantitative analysis of digitized mammograms. *Medical Physics* 28(4), 438–444 (2001)
- [23] Pennec, X., Thirion, J.: A Framework for Uncertainty and Validation of 3-D Registration Methods based on Points and Frames. *IJCV* 25(3), 203–229 (1997)
- [24] Reilhac, et al.: PET-SORTEO: validation and development of database of Simulated PET volumes. *IEEE Transactions on Nuclear Science* 52(5), part 1, 1321–1328 (2005)
- [25] Rosenberger, C.: Adaptive evaluation of image segmentation results. In: *ICPR*, pp. 399–402 (2006)
- [26] Schestowitz, R., Twining, C., Petrovic, V., Cootes, T., Crum, B., Taylor, C.: Non-Rigid Registration Assessment Without Ground Truth. In: *MIUA*, vol. 2, pp. 151–155 (2006)

- [27] Sled, J., Pike, B.: Understanding Intensity Non-uniformity in MRI. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) MICCAI 1998. LNCS, vol. 1496, pp. 614–622. Springer, Heidelberg (1998)
- [28] Styner, M., Rajamani, K., Nolte, L., Zsemlye, G., Székely, G., Taylor, C., Davies, R.: Evaluation of 3D Correspondence Methods for Model Building. In: Taylor, C.J., Noble, J.A. (eds.) IPMI 2003. LNCS, vol. 2732, pp. 63–75. Springer, Heidelberg (2003)
- [29] Twining, C., Cootes, T., Marsland, S., Petrovic, V., Schestowitz, R., Taylor, C.: Information-Theoretic Unification of Groupwise Non-Rigid Registration and Model Building. In: MIUA, pp. 226–230 (2006)
- [30] Unnikrishnan, R., Pantofaru, C., Hebert, M.: A Measure for Objective Evaluation of Image Segmentation Algorithms. In: CVPR Workshop on Empirical Methods in Computer Vision, p. 34 (2005)
- [31] Warfield, S., Zho, K., Wells, W.: Simultaneous Truth and Performance Level Estimation (STAPLE): An Algorithm for the Validation of Image Segmentation. *IEEE TMI* 23(7), 903–921 (2004)
- [32] West, et al.: Comparison and evaluation of retrospective inter-modality brain image registration techniques. *J. Computer Assisted Tomography* 21(4), 554–566 (1997)
- [33] Zhang, Y.: A review of recent evaluation methods for image segmentation. In: ISSPA, pp. 148–151 (2001)
- [34] Zhang, H., Fritts, J., Goldman, S.: An entropy-based objective evaluation methods for image segmentation. In: SPIE, vol. 5307, pp. 38–49 (2003)
- [35] Zhang, H., Cholleti, S., Goldman, S.: Meta-Evaluation of Image Segmentation Using Machine Learning. In: CVPR, pp. 1138–1145 (2006)
- [36] Zhang, Y.: A survey on Evaluation Methods for Image Segmentation. *Pattern Recognition* 29(8), 1335–1346 (1996)