Model Library for Deformable Model-Based Segmentation of 3-D Brain MR-Images

Juha Koikkalainen[1] and Jyrki Lötjönen[2]

[1] Laboratory of Biomedical Engineering, Helsinki University of Technology, P.O.B. 2200, FIN-02015 HUT, Finland
`jkoikkal@cc.hut.fi`
[2] VTT Information Technology, P.O.B. 1206, FIN-33101 Tampere, Finland
`Jyrki.Lotjonen@vtt.fi`

Abstract. A novel method to use model libraries in segmentation is introduced. Using similarity measures one model from a model library is selected. This model is then used in model-based segmentation. The proposed method is simple, straightforward and fast. Various similarity measures, both voxel and edge measures, were examined. Two different segmentation methods were used for validating the functionality of the proposed procedure. Results show that a statistically significant improvement in segmentation accuracy was achieved in each study case.

1 Introduction

In some applications the slow manual segmentation of complicated pathological organs may be acceptable. However, the manual segmentation is usually too slow for the segmentation of normal healthy organs, especially in studies where a large number of cases is necessary. Many kinds of automatic segmentation methods have been developed. One group of methods is segmentation using deformable models [1]. A commonly known problem with deformable models is the need for a good initialization. If the initialization is not good, a model may attract to wrong features or it may have problems in converging to complicated boundaries.

One way to make segmentation results more accurate and robust is to use a bigger set of models, i.e. a model library or a training set. Cootes and Taylor proposed an approach based on statistical models [2]. The idea is to allow all legal transformations of shape and spatial relations of structures, but at the same time prohibit all deformations that are not typical to the organ. Information on shapes, their spatial relations and gray-level appearance in the training set are used in establishing the statistical model. The normal procedure is to use principal component analysis (PCA) to calculate from the training set the modes of shape variation, which reflects variations in the training set. Usually a small number of modes, those with the highest eigenvalues, can explain most of the variation. Maximum variations in the eigenmode space can be also limited [3]. Wang and Staib used this kind of statistical shape information from the training set in non-rigid registration [4]. The fundamental problem with statistical models is the need for a point correspondence. The manual definition of

landmarks is slow and subjective, especially in three dimensions (3-D). One possibility to automate this is to use the parameterization of shapes, like spherical harmonics [3]. Davies *et al.* found the point correspondence via an optimization problem [5]. Rueckert *et al.* applied PCA to the deformation fields that were got by registering the volumes in a training set to a reference volume [6].

We propose a new way to utilize training set to improve segmentation accuracy. This method partly overcomes initialization problems. Our method has similarities with object recognition, like face recognition (e.g. [7]) and shape recognition (e.g. [8]), i.e. the model, which was the most similar to a target volume, i.e. the volume to be segmented, in terms of some similarity measure was searched. Instead of transforming the training set into a statistical form, a normal unconstrained segmentation was performed. The model used in segmentation was selected from the model library. The target volume was first rigidly transformed to the coordinate system of the model library. Thereafter, the model that was the most similar to the target volume was selected from the library based on a similarity measure, and used in segmentation. The proposed method is simple and straightforward. There is no need for defining the point correspondence. Selecting one model gives more freedom to the shape of the segmentation target. The topology of the model is not anymore bound, assuming that the model library is a representative sample of the target shape. The method can be used in segmentation as well as in registration.

2 Methods

2.1 Model Library

In this study, eight T1-weighted brain MR-volumes were used to build a model library. The size of the volumes was $128 \times 128 \times 90$ and the voxel size was about 2 mm $\times$ 2 mm $\times$ 2 mm. In each volume five organs were manually segmented using triangulated surfaces: skin, brain envelope, cerebellum, corpus callosum (CC) and midbrain. The intensity of the volumes was normalized by scaling histograms between values 0 and 255.

For establishing the model library, one volume was selected to be a reference volume. All other volumes were aligned with this volume using a registration method that optimizes seven parameters: 3 for translation, 3 for rotation and 1 for isotropic scaling. The optimization was based on mutual information. One slice from each model in the library is shown in Fig. 1.

An alternative possibility for initialization would be to use a coordinate system based on anatomical landmarks. The most popular approach is the Talairach-coordinate system which was used by Keleman *et al.* [3] for global alignment of head volumes. In this system 12 affine transformations are done based on anterior and posterior commissures (AC and PC) and the extreme limits of the brain.
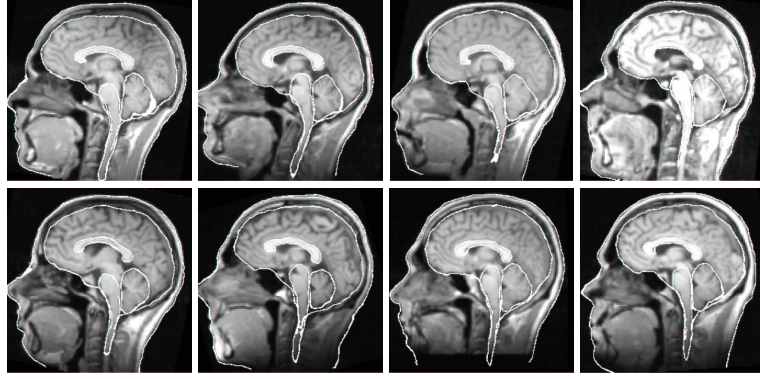
Fig. 1. An example of the model library. Eight volumes are manually segmented and registered in to a common coordinate system

2.2 Model Selection

A new target volume was first rigidly aligned with the reference volume. The model that was used in segmentation was selected based on similarity measures. Similarity measures are commonly used in image registration in cost and energy functions. These measures can be divided into two classes: voxel similarity measures which use gray-level information and edge measures which use edge information.

Voxel similarity measures are very popular in different registration applications. There exists a large selection of different measures [9, 10]. Measures based on correlation, like mean-square-distance (MSD), Pearson's cross correlation (NCC) and correlation ratio (CR) [11], are mainly for monomodal registration, except CR. Another class is voxel similarity measures based on entropy: e.g. entropy of the difference image (EDI), mutual information (MI) and normalized versions of MI, like normalized mutual information (NMI) [12]

$$\mathrm{NMI}(F, G) = \frac{H(F) + H(G)}{H(F, G)} \tag{1}$$

and entropy correlation coefficient (ECC) [13]

$$\mathrm{ECC}(F, G) = \frac{2MI(F, G)}{H(F) + H(G)} \ . \tag{2}$$

In these equations $H(F)$ and $H(G)$ are the entropies of the target volume F and the model volume G, respectively. $H(F, G)$ is the joint entropy of these volumes. These measures are applicable also in multimodal registration, because they make no assumptions about the form of the relationship between gray-levels in two images.

The idea of edge measures is to measure how near the model surface is to a strong edge. After all, the optimal segmentation is usually close to strong edges.
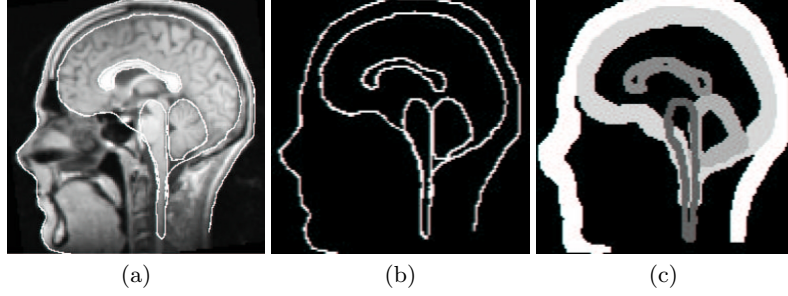
Fig. 2. a) A manually segmented volume. b) A binary volume of surfaces. c) Dilated volumes that were used as masks in computing similarity measures

Information on edge directions can be also incorporated in these measures, so that surfaces must be both near strong edges and edges in the model and the target volumes must be parallel. In this work we tested 24 edge measures such as

$$\text{edge1}(F, G) = \frac{1}{N} \sum_{i=1}^{N} \frac{\nabla f_i \cdot \nabla g_i}{|\nabla f_i||\nabla g_i|} |\nabla f_i| \tag{3}$$

and [14]

$$\text{edge2}(F, G) = \frac{1}{N} \sum_{i=1}^{N} \frac{\nabla f_i \cdot \nabla g_i}{|\nabla f_i||\nabla g_i|} \min(|\nabla f_i|, |\nabla g_i|) \ . \tag{4}$$

N is the number of nodes in the model surface and f_i and g_i are voxels in the target volume and the model volume, respectively. Edge volumes were made using Canny-Deriche operator, which gives also edge directions.

The values of these similarity measures are usually computed from the whole overlapping image volume in registration. Since the segmentations of separate organs are searched for in this study, similarity measures should take into account only the organ of interest. In this work a binary volume was made from the model's boundary (Fig. 2 (b)). Then a dilation operation was executed 1-3 times, depending on the size and the surroundings of the organ (Fig. 2 (c)). This volume was used as a mask when the similarity between the model volume and the target volume was computed. In this way only the alignment and the shape of that organ affected on the similarity measure. Edge-based similarity measures were computed mainly on the model surface.

The complementary information produced by voxel- and edge-based measures was merged to improve the accuracy and robustness in [14, 15]. The approach was adopted also in this work. The optimal combination of different measures was determined using regression analysis. The dependent variable was the final segmentation error and the independent variables were the values of similarity measures. A group of regression equations was achieved using the stepwise regression. The regression equation predicts the final segmentation error based on similarity measures. Therefore, the model that produced the lowest predicted value was chosen in the model selection phase.

2.3 Final Segmentation

The model selection procedure was validated by two segmentation algorithms, i.e. the chosen model was used as a prior model in the algorithms [16, 15].

In [16], a surface template consisting of triangulated surfaces was non-rigidly matched to edges in MR volumes using a free-form deformation (FFD) grid in the multiresolution framework. The energy measure to be minimized consisted of two components: 1) the distance of the model surfaces from the edges in MR volumes using oriented distance maps, and 2) the change of the model shape during the deformation.

A volumetric template consisting of a gray-scale volume and triangulated surfaces of objects of interest was used in [15]. Instead of applying FFD, the deformation was accomplished by deformation spheres. The energy term had three components. The first one was a voxel similarity measure; MSD was used in this monomodal study. The second term was a gradient term, which took into account edge intensities and directions in target and model volumes. The third term regulated the change of the model shape.

3 Results

Our image database was composed of nine volumes. Since the database was reasonably small, the jack-knife procedure was used to validate the model selection procedure, i.e. each volume was once regarded as a target volume and the rest eight volumes composed the model library (Fig. 1). The target volume was initialized to the common coordinate system and all similarity measures (in total 31) were calculated for all model volumes. Then the final segmentation was done and the segmentation error was determined as an average distance from the nodes of the model surface to the manually segmented target surface.

The following procedure was used to validate the model selection. For each target volume i, where $i \in 1, 2, ..9$, each library model ($N = 8$) was used separately as a prior model in the segmentation algorithm. The minimum m_i and the average a_i errors were defined. The minimum error is achieved if the best model was chosen in the model selection. The average gives an idea of the error if the model is chosen arbitrarily from the library. As the model selection was used, the error was s_i. The quality of the model selection is defined by the improvement percent, $p_i = 100\frac{a_i - s_i}{a_i - m_i}$. If the best model is chosen, $p_i = 100\%$. To get an idea of the real error values, $a = 0.986$, $m = 0.686$ and $s = 0.724$ voxels (for the similarity measure NMI) for the skin as the values were averaged over all target volumes.

Tables 1 and 2 present results averaged over nine target volumes for both segmentation methods and for all organs. The results are reported only for similarity measures that produced the best improvement percents. The second row is the percentage of cases, where the best model was selected, and the last row the percentage of cases, where a model better than the average was selected. In the two last columns, the same similarity measure was used for all organs.

Table 1. Results using surface-based segmentation

	skin	brain env	cerebellum	CC	midbrain	many models	one model
improvement (%)	65	90	61	47	78	53	26
best model (%)	22	33	33	22	44	36	18
better than average (%)	89	100	89	78	100	84	71

Table 2. Results using intensity-based segmentation

	skin	brain env	cerebellum	CC	midbrain	many models	one model
improvement (%)	87	94	57	60	92	63	35
best model (%)	44	67	11	44	67	44	29
better than average (%)	100	100	89	78	100	84	71

In the column "many models", each organ was chosen independently from the model library. In the column "one model", one model was selected from the model library and it was used to segment all organs. The improvements in the accuracy were statistically significant ($p < 0.05$), as tested with Wilcoxon signed rank test.

In overall, the best similarity measures were voxel measures. The best results using one similarity measure for all organs were achieved by using NMI. ECC was also good, as were NCC and EDI in some cases. The edge similarity measures presented in Eqs. 3 and 4 were almost as good as voxel measures. One and the same similarity measure was not the best one for each organ. However, using the same measure for all organs gave a reasonable good improvement, as can be seen from the two last columns.

The best model was not found every time. In some cases the selected model was even worse than the average. However, in these cases the model was almost as good as the average. These results indicate, that the selected model was generally good but not necessarily the best. Models, which gave a very poor segmentation, were never selected. So when a surface is attracted severely to a wrong edge, this model is discarded, as in the example in Fig. 3.

Tables 3 and 4 correspond to Tables 1 and 2, except that regression equations were utilized in the model selection. In these equations, two to five similarity
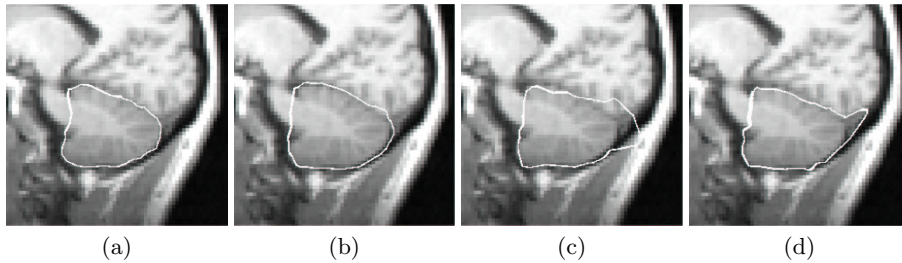

(a) (b) (c) (d)

Fig. 3. An example of the segmentation of cerebellum: a) and b) the best two models, and c) and d) the worst two models after the final segmentation

Table 3. Results using regression analysis and surface-based segmentation

	skin	brain env	cerebellum	CC	midbrain	many models	one model
improvement (%)	74	91	72	54	84	55	33
best model (%)	33	44	56	33	56	33	29
better than average (%)	100	100	89	89	100	84	73

Table 4. Results using regression analysis and intensity-based segmentation

	skin	brain env	cerebellum	CC	midbrain	many models	one model
improvement (%)	86	94	58	60	95	64	43
best model (%)	44	67	33	44	56	40	27
better than average (%)	100	100	100	89	100	87	71

measures were used. Despite of the complementary information in the selection, the results are only slightly better than using only one similarity measure.

The time needed for computing the values of similarity measures is short compared to the computation time of the segmentation methods. The initial alignment is necessary in these segmentation methods even if the model selection is not done. So the method proposed for model selection does not increase the computational burden of the segmentation practically at all. Obviously, if the library contained hundreds of models, the computation time of similarity measures might become significant.

4   Discussion

Deformable model-based segmentation is a powerful tool but suffers from problems related to the model initialization and complicated shapes. The results show convincingly that the presented method improves the segmentation accuracy by selecting one model from the model library. These improvements were statistically significant.

The proposed method is very simple and easy to realize. Any parameterization or definition of point correspondence is not necessary. The topology of segmented subject can vary, if this variation is presented in the model library as well. Also, this method can be combined with methods that use training set to make a statistical model of the shape. The model library could consist of statistical models from which the best one would be selected.

Initialization was done using rigid registration. Another possibility is to use the Talairach-coordinate system. According to our studies with a smaller database, this method may be better than the rigid registration. However, it can be used only for brain images whereas the method used in this paper can be generalized to other applications. Also elastic registration methods could be used in initial registration instead of linear transformations. This would improve results, but at the cost of the computation time.

In the future the library must be enlarged, manual segmentation improved and new, more complex organs studied. New segmentation methods, parame-

ters and especially different initialization methods should be also studied. The method will be tested also in other applications.

Acknowledgements

References

1. T. McInerney, D. Terzopoulos. Deformable Models in Medical Image Analysis: a Survey. *Medical Image analysis*, 1(2): 91–108, 1996.
2. T.F. Cootes, C.J. Taylor. Statistical models of appearance for medical image analysis and computer vision. *Proc. SPIE*, 2001.
3. A. Kelemen, G. Székely, G. Gerig. Elastic Model-Based Segmentation of 3-D Neuroradiological Data Sets. *IEEE Trans. Medical Imaging*, vol 18(10):828–839, 1999.
4. Y. Wang, L.H. Staib. Elastic Model Based Non-Rigid Registration Incorporating Statistical Shape Information. *Proc. MICCAI 1998*, 1162–1173, 1998.
5. R.D. Davies, T.F. Cootes, J.C.Waterton, C.J. Taylor. An Efficient Method for Constructing Optimal Statistical Shape Models. *Proc. MICCAI 2001*, 57–65, 2001.
6. D. Rueckert, A.F. Frangi, J.A. Schnabel. Automatic Construction of 3D Statistical Deformation Models Using Non-rigid Registration. *Proc. MICCAI 2001*, 77–84, 2001.
7. B. Moghaddam, A. Pentland. Face Recognition using View-Based and Modular Eigenspaces. *Automatic Systems for the Identification and Inspection of Humans, SPIE*, vol 2277, July 1994.
8. S. Sclaroff, A.P. Pentland. Modal Matching for Correspondence and Recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol 17(6): 545–561, 1995.
9. D.L.G. Hill, D.J. Hawkes. Across-Modality Registration Using Intensity-Based Cost Functions. Handbook of Medical Imaging Processing and Analysis, I.N. Bankman ed., Academic Press, 2000.
10. M. Holden, D.L.G. Hill, E.R.E. Denton, J.M. Jarosz, T.C.S. Cox, T. Rohlfing, J. Goodey, D.J. Hawkes. Voxel Similarity Measures for 3-D Serial MR Brain Image Registration. *IEEE Trans. Medical Imaging*, vol 19(2): 94–102, 2000.
11. A. Roche, G. Malandain, X. Pennec, N. Ayache. Multimodal Image Registration by Maximation of the Correlation Ratio. Rapport de recherche, INRIA, 1998.
12. C. Studholme, D.L.G. Hill, D.J. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32(1): 71–86, 1999.
13. J. Astola, I. Virtanen. Entropy correlation coefficient. A measure of statistical dependence for categorized data. *Proc. Univ. Vaasa, Discussion Papers*, 44, 1982.
14. J.P. Pluim, J.B.A. Maintz, M.A. Viergever. Image Registration by Maximation of Combined Mutual Information and Gradient Information. *IEEE Trans. Medical Imaging*, vol 19(8): 809–814, 2000.
15. J. Lötjönen, T. Mäkelä. Elastic matching using a deformation sphere. *Proc. MICCAI 2001*, 541–548, 2001.
16. J. Lötjönen, P-J. Reissman, I.E. Magnin, T. Katila. Model extraction from magnetic resonance voluma data using the deformable pyramid. *Medical Image Analysis*, vol. 3(4): 387–406, 1999.