

Automated Segmentation of CBCT Image Using Spiral CT Atlases and Convex Optimization

Li Wang¹, Ken Chung Chen², Feng Shi¹, Shu Liao¹, Gang Li¹, Yaozong Gao¹,
Steve GF Shen³, Jin Yan³, Philip K.M. Lee⁴, Ben Chow⁴,
Nancy X. Liu⁵, James J. Xia², and Dinggang Shen¹

¹ Department of Radiology and BRIC, University of North Carolina at Chapel Hill, NC, USA

² The Methodist Hospital Research Institute, Houston, Texas, United States

³ Shanghai Jiao Tong University Ninth Hospital, Shanghai, China

⁴ Hong Kong Dental Implant & Maxillofacial Centre, Hong Kong, China

⁵ Peking University School and Hospital of Stomatology, Beijing, China

Abstract. Cone-beam computed tomography (CBCT) is an increasingly utilized imaging modality for the diagnosis and treatment planning of the patients with craniomaxillofacial (CMF) deformities. CBCT scans have relatively low cost and low radiation dose in comparison to conventional spiral CT scans. However, a major limitation of CBCT scans is the widespread image artifacts such as noise, beam hardening and inhomogeneity, causing great difficulties for accurate segmentation of bony structures from soft tissues, as well as separating mandible from maxilla. In this paper, we presented a novel fully automated method for CBCT image segmentation. In this method, we first estimated a patient-specific atlas using a sparse label fusion strategy from predefined spiral CT atlases. This patient-specific atlas was then integrated into a convex segmentation framework based on *maximum a posteriori probability* for accurate segmentation. Finally, the performance of our method was validated via comparisons with manual ground-truth segmentations.

1 Introduction

Segmentation of the cone-beam computed topographic (CBCT) image is an essential step of generating 3D models in diagnosis and treatment planning of patients with craniomaxillofacial (CMF) deformities. It requires segmenting bony structures from soft tissues, as well as separating mandible from maxilla. Unlike expensive conventional spiral CT scanners, CBCT scanners, usually cost around \$200K, are getting popularly used clinically, even in private practice settings. However, CBCT image quality is significantly inferior to the spiral CT. Besides extremely low signal-to-noise ratio, CBCT scans have severe image artifacts, including noise, beam hardening, inhomogeneity, and truncation, thus affects image contrast and the accuracy of subsequent segmentation [1]. Furthermore, in order to better quantify the deformity, CBCT scans are usually acquired when the maxillary (upper) and mandibular (lower) teeth are in maximal intercuspation, which brings even more challenges to separate the

mandible from the maxilla [2]. To date, in order to use CBCT clinically for diagnosis and treatment planning, the segmentation must be completed manually by experts.

Manual segmentation is a tedious, time-consuming and user-dependent. Previous automated segmentation methods are mainly based on threshold and morphological operations [3], which are sensitive to the presence of the artifacts. Recently, shape information has been utilized for robust segmentation [4-6], e.g., Zhang et al. [6] proposed a deformable segmentation via sparse shape representation. However, these approaches are only applicable to objects with relatively regular shapes (e.g., mandible), but not the objects with complex shapes (e.g., maxilla). Interactive segmentation approaches [2, 7] were also provided to take advantage of both manual and automatic segmentation. To our best knowledge, there is no existing method that is able to automatically and simultaneously segment both maxilla and mandible from CBCT.

In this paper, we propose a fully automated CBCT segmentation method to 1) segment bony structures from the soft tissues, and 2) further separate the mandible from the maxilla. Specifically, we first employ a sparse label propagation strategy to estimate a patient-specific atlas from the spiral CT atlases. There are two reasons that why the spiral CT subjects are employed as the atlases: 1) although image formation process between spiral CT and CBCT is different, they share the same patterns of anatomical structures, which are captured in a patch fashion in our method to estimate the probability; 2) the images from CT scanners have better signal contrast and less noise than those from CBCT scanners, therefore less time is needed to construct atlases of manual segmentations from spiral CT images (i.e., ~30 minutes per image) than from CBCT images (i.e., ~12 hours per image) by an experienced operator. Finally, the patient-specific atlas is then integrated into a convex segmentation framework based on *maximum a posteriori probability* (MAP) for accurate segmentation.

2 Method

In this study, we aim to segment a CBCT scan into three structures/regions: mandible, maxilla (the skull without the mandible), and background. We consider image segmentation as the task of partitioning the image domain Ω into a set of three disjoint regions, such that $\Omega = \bigcup_{i=1}^3 \Omega_i$.

2.1 Subjects

We used 30 (15 males/15 females) spiral CT images with manual segmentations as atlases, and 13 (4 males/9 females) CBCT images as testing images. The 30 spiral CT images of the subjects with normal facial appearance scanned at maximal intercuspation were randomly selected from our HIPAA de-identified CT database. Their ages were 22 ± 2.6 years (range: 18-27 years). The CT matrix was 512×512 (resolution: $0.488 \times 0.488 \times 1.25$ mm³). 13 CBCT images were collected from the patients with non-syndromic dentofacial deformities and scheduled for double-jaw orthognathic surgery for their treatment. Their age were 24 ± 10 years (range: 10-49 years). The CBCT matrix was 400×400 (resolution $0.4 \times 0.4 \times 0.4$ mm³). These 30 CTs

and 13 CBCTs were labeled by two CMF surgeons who are experienced in segmentation using Mimics 10.01 software (Materialise NV, Leuven, Belgium).

2.2 Estimating Patient-Specific Atlas from Spiral CT Atlases

Atlas-based segmentation has demonstrated its robustness and effectiveness in many medical image segmentation problems [8]. Conventionally, a population-based atlas is constructed from multiple training images (e.g., manually segmented CBCTs). However, a population-based atlas often fails to provide useful guidance especially in the regions with high inter-subject anatomical variability, and leads to unsatisfactory segmentations results. One way to overcome this problem is to integrate the patient-specific information in the atlas construction. To this end, we propose to construct a patient-specific atlas by combining both population and patient information as follows.

We propose to estimate the prior, i.e., the patient-specific atlas, using a patch-based representation technique [9, 10]. The rational is that an image patch generally provides richer information, e.g., anatomical pattern, than a single voxel. Specifically, $N = 30$ spiral CT images I_j and their corresponding segmentation maps S_j ($j = 1, \dots, N$) are nonlinearly aligned onto the space of the testing image I using ELASIX [11]. Then, for each voxel x in the testing image I , its corresponding intensity patch with size $w \times w \times w$ can be represented as a column vector $\mathbf{Q}(x)$. An initial codebook $\mathbf{B}(x)$ can be constructed with respect to all these patches $\{\mathbf{Q}_j(x) | j = 1, \dots, N\}$ from all aligned templates at the same location, i.e., $\mathbf{B}(x) = [\mathbf{Q}_1(x), \mathbf{Q}_2(x), \dots, \mathbf{Q}_N(x)]$. To alleviate the effect of registration error, the initial codebook can be extended to include more patches from the neighboring search window $\mathbf{W}(x)$, i.e., $\mathbf{B}(x) = \{\mathbf{Q}_j(y), \forall j = 1 \dots, N, y \in \mathbf{W}(x)\}$, where each patch is represented by a column vector and normalized to have the unit ℓ_2 norm. To represent the patch $\mathbf{Q}(x)$ by the codebook $\mathbf{B}(x)$, its coding vector $\mathbf{c}$ could be estimated by many coding schemes, such as vector quantization, locality-constrained linear coding [12], and sparse coding [13]. Here, we utilize the sparse coding scheme [13] to estimate the coding vector $\mathbf{c}$ by minimizing a non-negative Elastic-Net problem [14],

$$\min_{\mathbf{c} \geq 0} \frac{1}{2} \|\mathbf{Q}(x) - \mathbf{B}(x)\mathbf{c}\|_2^2 + \lambda_1 \|\mathbf{c}\|_1 + \frac{\lambda_2}{2} \|\mathbf{c}\|_2^2 \quad (1)$$

where the first term is the least square fitting term, the second term is the ℓ_1 regularization term which is used to enforce the sparsity constraint on the reconstruction vector $\mathbf{c}$, and the last term is the ℓ_2 smoothness term to enforce the similarity of coding coefficients for the similar patches. Each element of the coding vector $\mathbf{c}$, i.e., $c_j(y)$, reflects the similarity between the target patch $\mathbf{Q}(x)$ and the patch $\mathbf{Q}_j(y)$ in the codebook. Based on the assumption that the similar patches should share similar labels, we use the sparse coding $\mathbf{c}$ to estimate the prior probability of the voxel x belonging to the i -th structure/region, i.e., $p_i(x) = \sum_j \sum_{y \in \mathbf{W}(x)} c_j(y) \delta_i(S_j(y))$, where $\delta_i(S_j(y)) = 1$ if the label $S_j(y) = i$; otherwise, $\delta_i(S_j(y)) = 0$. Finally, $p_i(x)$ is normalized to ensure $\sum_i p_i(x) = 1$.

2.3 Convex Segmentation Based on MAP

After obtaining the prior probability from the patient-specific atlas, we integrate it with the intensity distribution of the testing image itself for better segmentation. In the testing image, to accurately label each voxel x in the image domain Ω , we jointly consider its neighboring voxels $y \in \mathcal{O}(x)$, where $\mathcal{O}(x)$ is the neighborhood of voxel x . The regions $\{\Omega_i\}$ produce a partition of the neighborhood $\mathcal{O}(x)$, i.e., $\{\Omega_i \cap \mathcal{O}(x)\}_{i=1}^3$. We first consider the segmentation of $\mathcal{O}(x)$ based on *maximum a posteriori probability* (MAP). According to the Bayes rule:

$$p(y \in \Omega_i \cap \mathcal{O}(x) | I(y)) = \frac{p(I(y) | y \in \Omega_i \cap \mathcal{O}(x)) p(y \in \Omega_i) p(y \in \mathcal{O}(x))}{p(I(y))} \quad (2)$$

where $p(I(y) | y \in \Omega_i \cap \mathcal{O}(x))$, denoted by $p_{i,x}(I(y))$, is the structure probability density in region $\Omega_i \cap \mathcal{O}(x)$. Note that $p(y \in \Omega_i)$, i.e., $p_i(y)$, is the a priori probability of y belonging to the region Ω_i , which has been estimated in section 2.2. $p(y \in \mathcal{O}(x)) = \mathbf{1}_{\mathcal{O}(x)}(y)$ is the indicator function and $p(I(y))$ is independent of the choice of the region and can therefore be neglected. Thus, Eq. (2) can be simplified as $p(y \in \Omega_i \cap \mathcal{O}(x) | I(y)) = p_{i,x}(I(y)) p_i(y)$. Assuming that the voxels within each region are independent, the MAP will be achieved only if the product of $p_{i,x}(I(y)) p_i(y)$ across the region $\mathcal{O}(x)$ is maximized: $\prod_{i=1}^3 \prod_{y \in \Omega_i \cap \mathcal{O}(x)} p_{i,x}(I(y)) p_i(y)$. In fact, we can use a Gaussian kernel K_ρ with scale ρ to indicate the neighborhood $\mathcal{O}(x)$ [15]. Taking a logarithm transformation and integrating all the voxels x , the maximization can be converted to the minimization of the following energy,

$$\mathcal{E}^D = - \sum_{i=1}^3 \int_{\Omega} \int_{\Omega_i} K_\rho(x - y) \log p_{i,x}(I(y)) p_i(y) dy dx \quad (3)$$

Based on [16], we can use multiple variables which take values between 0 and 1 to derive a convex formulation for Eq. (3). Since in our project, there are only 3 different regions of interest: mandible, maxilla, and background, we need only 2 segmentation variables $u_1 \in [0, 1]$ and $u_2 \in [0, 1]$ to represent the partitions $\{\Omega_i\}$: $M_1 = u_1$, $M_2 = u_2$, $M_3 = (1 - u_1)(1 - u_2)$. Therefore, Eq. (3) can be converted as follows,

$$\mathcal{E}^D = - \sum_{i=1}^3 \int_{\Omega} \int_{\Omega} K_\rho(x - y) \log p_{i,x}(I(y)) p_i(y) M_i(y) dy dx \quad (4)$$

There are many options to estimate the $p_{i,x}(I(y))$. In this paper, we utilize a Gaussian distribution model with the local mean $\mu_i(x)$ and the variance $\sigma_i^2(x)$ [17] to estimate it: $p_{i,x}(I(y)) = \exp\left(-(\mu_i(x) - I(y))^2 / 2\sigma_i^2(x)\right) / (\sqrt{2\pi}\sigma_i(x))$.

Based on the assumption that there should be no overlap between mandible and maxilla, we propose the following penalty constraint term,

$$\mathcal{E}^P = \int u_1(x) u_2(x) dx \quad (5)$$

In addition, the length regularization term [18] is defined as weighted total variation of functions u_1 and u_2 ,

$$\mathcal{E}^{\mathcal{R}} = \int g(I(x))(|\nabla u_1(x)| + |\nabla u_2(x)|) dx \quad (6)$$

where g is a non-edge indicator function that vanishes at object boundaries [18].

Finally, we define the entire energy functional below, which consists of the data fitting term $\mathcal{E}^{\mathcal{D}}$, the overlap penalty term $\mathcal{E}^{\mathcal{P}}$ and the length regularization term $\mathcal{E}^{\mathcal{R}}$:

$$\min_{u_1, u_2 \in [0, 1]} \{\mathcal{E}(u_1, u_2) = \mathcal{E}^{\mathcal{D}} + \alpha \mathcal{E}^{\mathcal{P}} + \beta \mathcal{E}^{\mathcal{R}}\} \quad (7)$$

where α and β are the positive coefficients. Based on [16], the energy functional (Eq. (7)) can be easily minimized in a fast way with respect to u_1 and u_2 .

3 Experimental Results

The parameters used in this paper were determined experimentally via cross validation. In fact, the results are relatively insensitive to the variation of the parameters. We finally chose the following parameters for all the experiments: the weight for ℓ_1 -norm term $\lambda_1 = 0.1$, the weight for ℓ_2 -norm term $\lambda_2 = 0.05$, the patch size $9 \times 9 \times 9$, the search window size $5 \times 5 \times 5$, $\rho = 3$ for the Gaussian Kernel K_ρ , and the weights $\alpha = 10$ for the overlap penalty term $\mathcal{E}^{\mathcal{P}}$ and $\beta = 10$ for the length regularization term $\mathcal{E}^{\mathcal{R}}$.

Fig. 1 demonstrates the segmentation results of different methods for one typical subject. In the first row, the volume rendering of the original intensity image, and surfaces rendering of the segmentation result obtained by majority voting (MV), patch-based fusion method [10], the proposed method by directly using the maximum class probability from step 1 (prior estimation) as the segmentation result, the proposed method with both step 1 (prior estimation) and step 2 (convex segmentation), and the manual segmentation, are shown from left to right. The following rows show the corresponding results on slices and zoomed views for better visualization, from which the artifacts can be clearly observed. Due to possible errors during image registration, the surface by MV is far from accurate, which incorrectly labels some upper teeth as lower teeth, by referring to the manual segmentation. Due to the closed-bite position and large intensity variations, the patch-based fusion method [10] cannot accurately separate the mandible from maxilla, thus mislabeling the upper teeth and lower teeth. Note that, to be a fair comparison, we have performed the similar searching scheme to derive the optimal parameters for the patch-based fusion method [10], by obtaining the final patch size of $9 \times 9 \times 9$ and the final search window size of $5 \times 5 \times 5$. Overall, the proposed method produces much more reasonable results.

We then quantitatively evaluate the performance of different methods on 13 subjects by employing Dice ratio. The average Dice ratios of different methods are shown in Fig. 2. As also shown in Table 1, with the convex segmentation integrating the estimated prior, the proposed method achieves the highest Dice ratios. To further validate the proposed method, we also evaluate its accuracy by measuring the average

surface distance error, as plotted in Fig. 3. Additionally, the Hausdorff distance is also used to measure the maximal surface-distance errors of each of 13 subjects. The average Hausdorff distance on all 13 subjects are shown in Table 1, which again demonstrates the advantage of our proposed method.

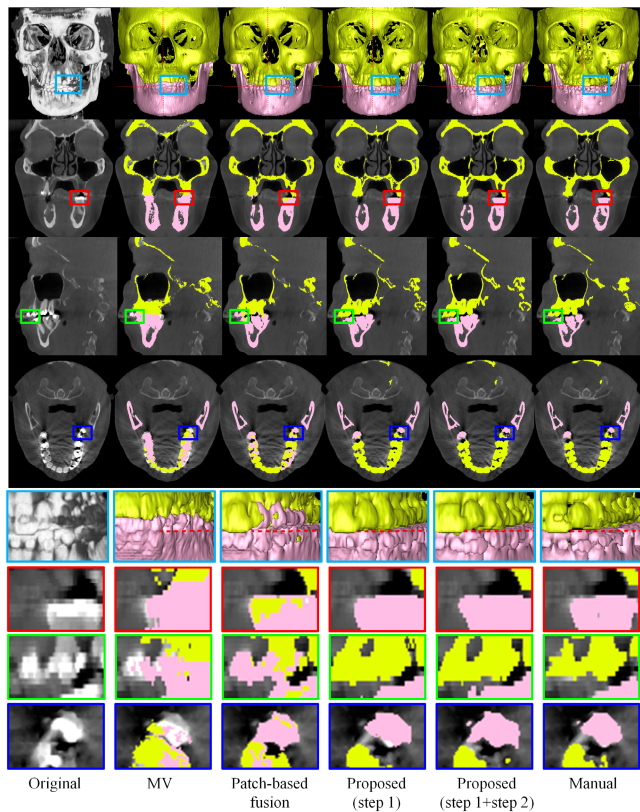


Fig. 1. Comparisons of segmentation results of 4 different methods on a typical CBCT image

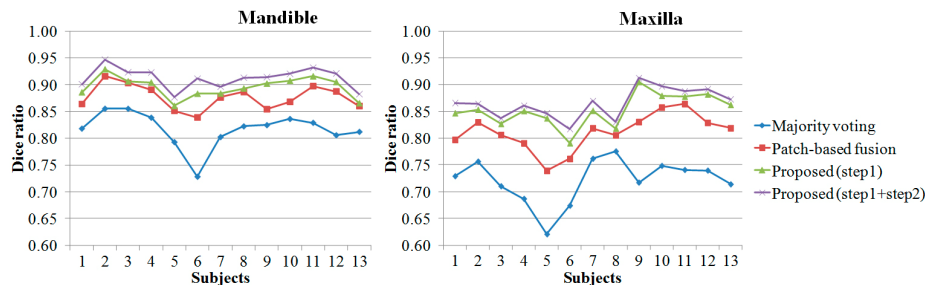


Fig. 2. Dice ratios on mandible and maxilla by 4 different methods

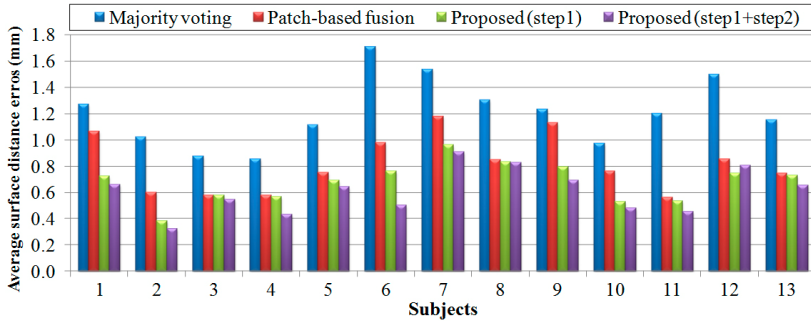


Fig. 3. Average surface distances from the surfaces obtained by 4 different methods to the ground-truth surfaces on 13 subjects

Table 1. Average Dice ratio and surface distance error (in mm) on 13 subjects

		Majority voting	Patch-based fusion [10]	Proposed (step 1)	Proposed (step 1+2)
Dice ratio	Mandible	0.82±0.03	0.88±0.02	0.89±0.02	0.91±0.02
	Maxilla	0.72±0.04	0.81±0.03	0.85±0.03	0.87±0.02
Average distance error	Mandible	1.21±0.25	0.81±0.21	0.67±0.15	0.61±0.17
Hausdorff distance error	Mandible	3.61±1.53	2.22±1.1	1.14±0.54	0.92±0.47

4 Discussion and Conclusion

We have successfully developed and validated a novel fully automated method for CBCT segmentation. We first estimated a patient-specific atlas from spiral CT atlases using a sparse label fusion strategy. Then, we integrate it into a convex segmentation framework based on MAP for segmentation. Comparing to the state-of-the-art label-fusion methods, our method achieved the best results.

The success of applying normal CT atlases to the CBCT subject with CMF deformity can be mainly attributed to the following two factors: 1) The deformation between the subject with the CMF deformity and the normal subject is first alleviated by nonlinear image registration; 2) After registration, the structure probabilities for the CBCT subject with CMF deformity are then robustly estimated by the proposed patch-based sparse technique in Section 2.2.

In our experiment, we found that increasing the number of atlases would generally improve the segmentation accuracy. For example, the combination of CT and CBCT atlases achieves slightly improved Dice ratios: 0.005 higher for mandible and 0.0046 higher for maxilla. However, more atlases also bring in larger computational cost.

In our future work, we will validate the proposed method on more dataset and further compare with other methods such as [19]. We will also improve the robustness of the proposed method by increasing the variability of the atlases such as including more datasets with different CMF deformities.

References

1. Loubele, M., Maes, F., Schutyser, F., Marchal, G., et al.: Assessment of bone segmentation quality of cone-beam CT versus multislice spiral CT: a pilot study. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology* 102, 225–234 (2006)
2. Le, B.H., Deng, Z., Xia, J., Chang, Y.-B., Zhou, X.: An Interactive Geometric Technique for Upper and Lower Teeth Segmentation. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009, Part II. LNCS*, vol. 5762, pp. 968–975. Springer, Heidelberg (2009)
3. Hassan, B.A.: *Applications of Cone Beam Computed Tomography in Orthodontics and Endodontics* (2010)
4. Kainmueller, D., Lamecker, H., Seim, H., Zinser, M., Zachow, S.: Automatic Extraction of Man-dibular Nerve and Bone from Cone-Beam CT Data. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009, Part II. LNCS*, vol. 5762, pp. 76–83. Springer, Heidelberg (2009)
5. Gollmer, S.T., Buzug, T.M.: Fully automatic shape constrained mandible segmentation from cone-beam CT data. In: *ISBI*, pp. 1272–1275 (2012)
6. Zhang, S., Zhan, Y., Dewan, M., Huang, J., Metaxas, D.N., Zhou, X.S.: Deformable segmentation via sparse shape representation. In: Fichtinger, G., Martel, A., Peters, T. (eds.) *MICCAI 2011, Part II. LNCS*, vol. 6892, pp. 451–458. Springer, Heidelberg (2011)
7. Suebnukarn, S., Haddawy, P., Dailey, M., Cao, D.: Interactive Segmentation and Three-Dimension Reconstruction for Cone-Beam Computed-Tomography Images. *NECTEC Technical Journal* 8, 154–161 (2008)
8. Aljabar, P., Heckemann, R.A., Hammers, A., Hajnal, J.V., et al.: Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. *NeuroImage* 46, 726–738 (2009)
9. Rousseau, F., Habas, P.A., Studholme, C.: A Supervised Patch-Based Approach for Human Brain Labeling. *TMI* 30, 1852–1862 (2011)
10. Coupé, P., Manjón, J., Fonov, V., Pruessner, J., et al.: Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *NeuroImage* 54, 940–954 (2011)
11. Klein, S., Staring, M., Murphy, K., Viergever, M.A., et al.: elastix: A Toolbox for Intensity-Based Medical Image Registration. *TMI* 29, 196–205 (2010)
12. Wang, J., Yang, J., Yu, K., Lv, F., et al.: Locality-constrained Linear Coding for image classification. In: *CVPR*, pp. 3360–3367 (2010)
13. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., et al.: Robust Face Recognition via Sparse Representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 31, 210–227 (2009)
14. Zou, H., Hastie, T.: Regularization and variable selection via the Elastic Net. *Journal of the Royal Statistical Society, Series B* 67, 301–320 (2005)
15. Li, C.M., Kao, C.Y., Gore, J.C., Ding, Z.H.: Minimization of Region-Scalable Fitting Energy for Image Segmentation. *TIP* 17, 1940–1949 (2008)
16. Goldstein, T., Bresson, X., Osher, S.: *Geometric Applications of the Split Bregman Method: Segmentation and Surface Reconstruction*. CAM Report, UCLA (2009)
17. Brox, T., Cremers, D.: On Local Region Models and a Statistical Interpretation of the Piecewise Smooth Mumford-Shah Functional. *IJCV* 84, 184–193 (2009)
18. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. *IJCV* 22, 61–79 (1997)
19. Liao, S., Gao, Y., Shen, D.: Sparse patch based prostate segmentation in CT images. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) *MICCAI 2012, Part III. LNCS*, vol. 7512, pp. 385–392. Springer, Heidelberg (2012)