33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

# A FEATURE-BASED APPROACH FOR REFINEMENT OF MODEL-BASED SEGMENTATION OF LOW CONTRAST STRUCTURES

*Arish A. Qazi, John Kim, David A. Jaffray*

Princess Margaret Hospital, Toronto, Canada

*Vladimir Pekar*

Philips Research, Markham, Canada

## ABSTRACT

Accuracy and robustness are fundamental requirements of any automated method used for segmentation of medical images. Model-based segmentation (MBS) is a well established technique, where uncertainties in image content can be to a certain extent compensated by the use of prior shape information. This approach is, however, often problematic in cases where image information does not allow for generating a strong feature response, one example being soft tissue organs in CT data, which typically appear in low contrast. In this paper, we enhance our recently proposed framework for voxel classification-based refinement of MBS using a level-set segmentation technique with shape priors. We also introduce a novel feature weighting methodology that improves the performance of the classifier, demonstrating results superior to the previous feature selection method. Results of fully automated segmentation of low contrast organs in head and neck CT are presented. Compared to our previous approach, we have achieved an increase of up to 22% in segmentation accuracy.

***Index Terms***— Model-based segmentation, classification, radiation therapy planning, level-sets, feature weighting.

## 1. INTRODUCTION

An essential requirement for successful implementation of Intensity Modulated Radiation Therapy (IMRT) [1] is accurate contouring of the target structures as well as organs at risk. Manual contouring is typically very time consuming and suffers from both, intra- and inter- observer variability. Fully automated segmentation methods for radiation therapy planning (RTP) are highly desirable, however, given the high anatomical variability and low contrast, the task is very challenging.

Current automated segmentation methods in RTP are either deformable (model-based) approaches, or atlas-based methods. Deformable models rely on distinct image features, such as edges and may also include prior shape and appearance information [2], however, such methods suffer in regions of low contrast, where object boundary discrimination is not distinct. Atlas-based methods are a popular choice in RTP [3, 4], since prior information is integrated in a simple and effective way. Such methods rely on one or several atlas images, which contain contours labeled by an expert. To segment a new clinical case, the atlas is registered to the image and the structures of interest are then transformed using the mapping determined by the registration. The major drawback of this approach is its reliance on accuracy of image registration which can fail with high variability of the patient anatomy, organ motion, and image artifacts.

Recently, we proposed a hybrid approach which combines registration and model-based segmentation into a common framework [5]. We build an organ-specific probabilistic atlas by affine registration of expert segmentations, and register the atlas with the result of model-based segmentation. The uncertainty area in the transferred atlas is refined using k-nearest neighbor ($k$NN) based voxel classification using a plurality of low-level image features. The features are organ-specific and have been selected using a combinatorial feature selection method. One drawback with the method in [5] is manual selection of a threshold that is used to obtain the final segmentation from the combined classification and atlas probabilities.

In this paper, we propose a variational approach which attempts to address this issue. Another improvement to the previous framework is the introduction of feature relevance weighting, which is aimed at improving the classification performance of the $k$NN method. The introduced methodology is generic and can be utilized in other probabilistic segmentation and classification applications.

The paper is organized as follows. In the next section, we describe our previous segmentation framework, followed by the proposed feature weighting method, and the variational segmentation approach. In Section 3, we present the quantitative and qualitative evaluation of our method when compared to manual expert segmentations. Finally, discussion and conclusions are presented in Section 4.

## 2. METHODS

### 2.1. Previous segmentation framework

As a first step, a collection of manually segmented ground truth data is used to create a organ-specific probabilistic atlas using affine registration [5]. Given a new image, a model-based segmentation (MBS) technique based on energy mini-

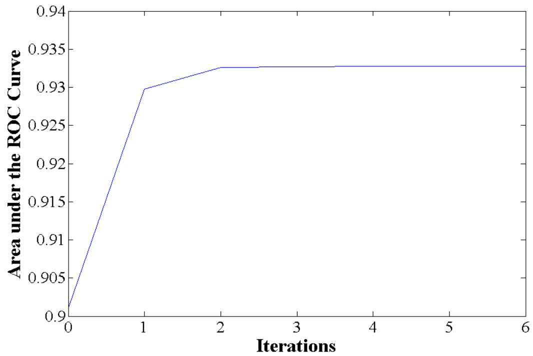


**Fig. 1**. Plot showing the increase in AUC by feature relevance weighting. The AUC value at iteration 0 is achieved from the floating feature selection method.

mization [2] is applied to segment the anatomical structures of interest.

The final step is to refine the resulting segmentations. Each probabilistic atlas is registered with the result of MBS, and the uncertainty area consisting of probabilities between 0 and 1 is refined by voxel classification. An optimal set of organ-specific features selected from a pool of low-level image features describing various structural and textural properties at different scales are used [5]. For classification, an efficient implementation of a $k$NN classifier [6] is applied to compute the probability for a particular voxel to belong to the object of interest. This probability is averaged with the atlas probability, and the result is thresholded at a certain value, e.g. $0.5$, to obtain the final segmentation.

### 2.2. Feature weighting

The prediction quality of the $k$NN method is known to be highly dependent on how well the features are able to discriminate between two different instances. Since the discrimination is based on the Euclidean distance function, it makes the $k$NN classifier to be quite sensitive to presence of redundant, irrelevant, and noisy features. This could possibly be circumvented by doing dimensionality reduction using feature selection, as in our previous method [5]. Feature selection, however, assigns a binary weight to each feature selected and might be useful for filtering out relevant features, and it has been shown that if the features vary in their relevance then classification accuracy can be further improved by using feature weighting [7].

In this paper, we propose a new and novel feature weighting methodology, which combines feature selection with a line search method for optimal feature relevance weighting. As a first step, as in [5], in order to choose the optimal set of features we employ a feature selection step based on the heuristic sequential forward floating selection (SFFS) [8]. The performance of SFFS is comparable to the optimal branch and bound algorithm, while being more computationally efficient. SFFS translates to a forward selection (FS) step, followed by backward selection (BS). FS starts from an empty set and adds features sequentially as long as the performance criterion improves. Subsequently, BS iteratively removes the least significant features according to the performance criterion. The outcome of the performance criterion is evaluated at each iteration and we stop iterating when the dimensionality of the feature space reaches a point after which the improvement is not significant. For the performance criterion, we maximize AUC - the area under the receiver operating characteristic (ROC) curve. The ROC curve is determined by varying the classification threshold and plotting the ratio of false positives vs. the ratio of true positives [9].

After the feature selection step, optimal set of features for a specific structure has been selected; however, each feature has a binary weight. We propose to find the optimal weight (relevance) of each selected feature using line search. Line search methods are typically used to find the minimum value of a function in one dimension [10]. Since, the line search is one dimensional, while searching for the optimal weight for a particular feature, the weights of all other features are kept constant. This process is repeated for all the selected features. We repeat the line search step for several iterations until the optimization criterion cannot be further improved, where the relevance criterion is maximization of the AUC. If during line search the criterion is not improved for a specific feature then its weight remains unchanged. All weights are initialized to 1. For implementation of line search, we use Brent's method [11], which first finds a bracket which contains the desired optima. A bracket is a triple $(x, y, z)$, such that, $AUC(x) < AUC(y) > AUC(z)$. Once the bracket is found using interpolation/Golden section method [11], we iteratively search for the optimal weight.

Fig. 1 illustrates the increase in classification performance achieved by the same features, selected using SFFS, and then weighted using line search technique. It shows that for the

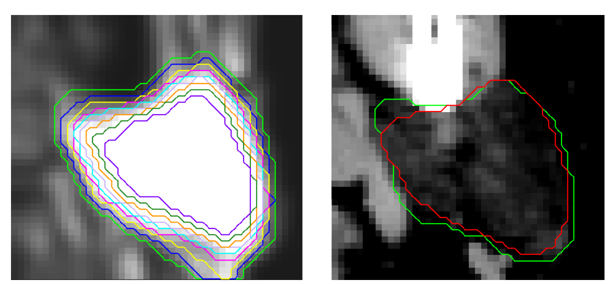

**Fig. 2**. Slice depicting segmentation results of the left parotid. Left: classified probability map of a sub-volume. Overlaid on this image are segmentations obtained from our previous framework by varying the threshold. Right: the ground truth (green) and the segmentation of our new level-set based segmentation (red) overlaid on the raw intensity image.

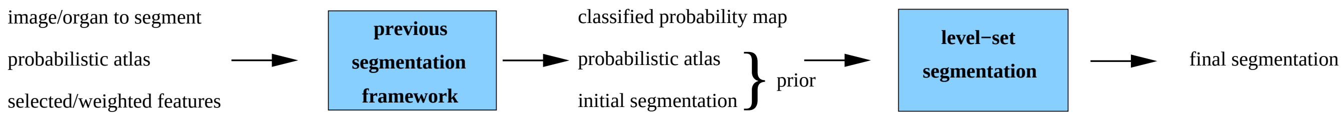


**Fig. 3**. A variational approach for segmentation of low-contrast structures.

first iteration of the line search procedure the AUC increases considerably implying the effect of feature weighting and in successive iterations the AUC reaches to an optimal value.

### 2.3. Variational segmentation of probabilities

A drawback of the previous formulation is the use of a hard threshold to obtain the final segmentation whose optimal choice can vary between different structures. Furthermore, the threshold is not intuitive and can have an adverse impact on the segmentation results, see Fig. 2 (left).

Instead of applying thresholds, we propose to use a level-set formulation, following the approach suggested by Chan and Vese [12] to obtain the final segmentation. For this purpose, all voxels in the region around (and including) the segmentation obtained by the previous framework are classified using $k$NN and the set of selected weighted features, resulting in a probability map. This probability map is then segmented using level sets with a shape prior. Fig. 3 gives an illustrative overview of the proposed method.

Level sets provide an implicit representation of the evolving contour, which is embedded as the zero level set of a signed distance function $\phi$ ($\phi > 0$ inside and $\phi < 0$ outside the contour, and $|\nabla\phi| = 1$). Two-region segmentation is performed by minimizing the following functional [12]

$$E_\nu(c_1, c_2, \phi) = \int_\Omega \{(u - c_1)^2 H(\phi) + (u - c_2)^2 (1 - H(\phi)) + \nu|\nabla H(\phi)|\} dxdydz,$$

where $u : \Omega \to R$ is the classified probability map, $c_1$ and $c_2$ are scalar variables which are updated during the level set evolution, and represent the mean intensity of the two regions (where $\phi$ is positive or negative). $H(\phi)$ denotes the Heaviside function, which is $1$ where $\phi \geq 0$, and $0$ otherwise, and $\nu > 0$ is a parameter that controls the smoothness of the evolving contour.

The above model, however, is purely based on the probabilities present in the map and will fail to segment a meaningful region if the probabilities are less pronounced. We adopt the model of Cremers et al. [13] which elegantly incorporates shape information into the Chan-Vese functional. The shape prior is represented as a signed distance function, $\phi_0$ and is enforced on all the domain $\Omega$ by the $L_2$ distance

$$E_{shape}(\phi) = \int_\Omega (\phi - \phi_0)^2 dxdydz,$$

where $\phi_0$ can be respresented as a mean shape, however, due to high anatomical variability we want the prior to be in close proximity to the anatomy of the underlying subject. Therefore, we use the segmentation output from our previous method, as explained in section 2.1, as the shape prior.

In addition to the shape prior, we also introduce a weighting term, where the weights are probabilities quantified from the probabilistic atlas (Section 2.1). In this way the distance function will be penalizing regions of low probabilities, which is advantageous in situations of distorted shape prior. We modify the term as

$$E_{shape}(\phi) = \int_\Omega (1 - m)(\phi - \phi_0)^2 dxdydz,$$

where m denotes the probabilistic atlas. Thus, combining the terms, the segmentation functional becomes

$$E(c_1, c_2, \phi) = E_\nu(c_1, c_2, \phi) + \lambda E_{shape}(\phi),$$

where $\lambda \geq 0$ determines the influence of the shape prior term. As in Chan-Vese's method [12], minimizing the above functional with respect to $\phi$ is implemented by gradient descent

$$\frac{\partial\phi}{\partial t} = \delta(\phi)\left[\nu \operatorname{div}\left(\frac{\nabla\phi}{|\nabla\phi|}\right) - (u - c_1)^2 + (u - c_2)^2\right] - 2\lambda(1 - m)(\phi - \phi_0),$$

where $\delta(\phi)$ is the derivative of the Heaviside function $H(\phi)$.

## 3. RESULTS

Image data used for the experiments consisted of head and neck CT scans of 25 patients acquired at Princess Margaret Hospital in Toronto, Canada. All datasets were acquired using a standard field of view and do not contain large neck deformations due to disease. Scan resolution for all datasets was approximately $1\times1\times2$ mm$^3$. For voxel classification training, feature selection, and parameter settings for the level-set formulation, 15 datasets were used, and the remaining 10 datasets were used to validate the method. Both feature selection and weighting are carried out by randomly dividing the training data into two subsets. The classifier is trained for a combination of features on the first set and the performance is then evaluated on the second set. Note that unlike using non-intuitive and organ-dependent thresholds to obtain the final segmentation [5], the parameters in the level-set formulation do not require tuning for every new case.

| Dataset | Brainstem (DSC) | Brainstem (HD) | L Par (DSC) | L Par (HD) | R Par (DSC) | R Par (HD) | L Sub (DSC) | L Sub (HD) | R Sub (DSC) | R Sub (HD) |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.89 | 3.1 | 0.81 | 4.3 | 0.74 | 7.4 | 0.82 | 3.0 | 0.69 | 5.0 |
| 2 | 0.91 | 2.9 | 0.78 | 7.9 | 0.84 | 4.0 | 0.73 | 5.2 | 0.66 | 5.1 |
| 3 | 0.86 | 3.3 | 0.84 | 3.8 | 0.84 | 5.6 | 0.80 | 4.2 | 0.80 | 4.8 |
| 4 | 0.91 | 2.9 | 0.89 | 3.5 | 0.90 | 4.0 | 0.85 | 3.5 | 0.85 | 3.7 |
| 5 | 0.87 | 2.1 | 0.82 | 4.1 | 0.83 | 6.2 | 0.78 | 5.9 | 0.73 | 6.5 |
| 6 | 0.91 | 2.2 | 0.86 | 6.4 | 0.84 | 8.0 | 0.87 | 2.5 | 0.84 | 2.8 |
| 7 | 0.90 | 3.9 | 0.87 | 4.4 | 0.83 | 6.9 | 0.82 | 4.9 | 0.81 | 5.7 |
| 8 | 0.91 | 2.2 | 0.83 | 7.0 | 0.81 | 5.9 | 0.84 | 2.9 | 0.75 | 5.2 |
| 9 | 0.91 | 2.8 | 0.88 | 3.5 | 0.87 | 4.0 | 0.86 | 3.3 | 0.85 | 3.5 |
| 10 | 0.90 | 2.9 | 0.89 | 3.9 | 0.88 | 4.9 | 0.84 | 3.5 | 0.85 | 4.2 |
| **Mean** | **0.90** | **2.8** | **0.85** | **4.9** | **0.84** | **5.7** | **0.82** | **3.9** | **0.78** | **4.6** |

**Table 1**. Segmentation results: the table lists the DSC, and the median HD (mm) of our approach vs. expert segmentations. Mean DSC overlap and average median HD distance (last row). Left(L), Right(R), Par(Parotid), Sub(Submandibular gland).

To quantitatively evaluate the resulting segmentations, we compare them to the manual segmentations done by a clinical expert. The evaluation is carried out by estimating two common measures on 10 datasets: volume overlap fraction or the Dice similarity coefficient (DSC), and a geometrical metric, the Hausdorff distance (HD), which is evaluated slice-wise. Table 1 lists these quantitative measures for all structures. Compared with the results of the previous approach [5], an improvement in volume overlap of around 7% has been achieved for the brainstem, 18% for the left and 22% for the right parotid. Quantitative validation for the submandibular glands has been done for the first time in the present paper, so no direct comparison with the previous approach can be made, but taking their fuzzy visual appearance in CT data into account, current results can be considered as promising.

## 4. CONCLUSION

We have proposed a variational scheme by enhancing our previous approach for segmentation of low contrast structures in the head and neck region. We segment voxel classified probability maps by extending the Chan-Vese functional to incorporate weighted shape prior information. The primary advantage of our approach is that it removes the reliance on organ-specific user based thresholds, which are difficult to select due to high anatomical variability. Our results demonstrate that when compared to the old methodology we are able to achieve superior segmentation accuracy. Future work will involve a multi-center validation study incorporating more structures in the head and neck region.

## 5. REFERENCES


[1] W.W. Chou, D.R. Puri, and N.Y. Lee, "Intensity-modulated radiation therapy for head and neck cancer," *Expert Rev Anticancer Ther*, vol. 5, pp. 515–521, 2005.

[2] V. Pekar, T.R. McNutt, and M.R. Kaus, "Automated model-based organ delineation for radiotherapy planning in prostatic region," *Int J Radiat Oncol Biol Phys*, vol. 60, pp. 973–980, 2004.

[3] P.Y. Bondiau, G. Malandain, S. Chanalet, P.Y. Marcy, J.L. Habrand, F. Fauchon, P. Paquis, A. Courdi, O. Commowick, and I. Rutten, "Atlas-based automatic segmentation of mr images: validation study on the brainstem in radiotherapy context," *Int J Radiat Oncol Biol Phys*, vol. 61, pp. 289–298, 2005.

[4] X. Han, M.S. Hoogeman, P.C. Levendag, L.S. Hibbard, D.N. Teguh, P. Voet, A.C. Cowen, and T.K. Wolf, "Atlas-based auto-segmentation of head and neck CT images," in *Med Image Comput Comput Assist Interv*, 2008, pp. 434–441.

[5] A.A. Qazi, J.J. Kim, D.A. Jaffray, and V. Pekar, "Probabilistic refinement of model-based segmentation: Application to radiation therapy planning of the head and neck," in *Med Imag Augmented Reality*, 2010, pp. 403–410.

[6] S. Arya, D.M. Mount, R. Netanyahu, N.S. Silverman, and A.Y. Wu, "An optimal algorithm for approximate nearest neighbor searching fixed dimensions," *J ACM*, vol. 45, pp. 891–923, 1998.

[7] D. Wettschereck, D.W. Aha, and T. Mohri, "A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms," *Art Intel Rev*, vol. 11, pp. 273–314, 1997.

[8] P. Pudil, J. Novovicová, and J. Kittler, "Floating search methods in feature selection," *Pat Recog Let*, vol. 15, pp. 1119–1125, 1994.

[9] M.H. Zweig and G. Campbell, "Receiver-operating characteristic (roc) plots: a fundamental evaluation tool in clinical medicine," *Clin Chem*, vol. 39, 1993.

[10] J. Nocedal and S.J. Wright, *Numerical Optimization*, Springer Verlag, 1999.

[11] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling, *Numerical Recipes*, Cambridge University Press, 2007.

[12] T.F. Chan and L.A. Vese, "Active contours without edges," *IEEE Trans Image Proc*, pp. 266–277, 2001.

[13] D. Cremers, N. Sochen, and C. Schnörr, "Towards recognition-based variational segmentation using shape priors and dynamic labeling," in *Scale Space Met Comp Vis*, 2003, pp. 388–400.