

# Weighted Shape-Based Averaging With Neighborhood Prior Model for Multiple Atlas Fusion-Based Medical Image Segmentation

Subrahmanyam Gorthi, Meritxell Bach Cuadra, Pierre-Alain Tercier, Abdelkarim S. Allal, and Jean-Philippe Thiran, *Senior Member, IEEE*

**Abstract**—In medical imaging, merging automated segmentations obtained from multiple atlases has become a standard practice for improving the accuracy. In this letter, we propose two new fusion methods: “Global Weighted Shape-Based Averaging” (GWSBA) and “Local Weighted Shape-Based Averaging” (LWSBA). These methods extend the well known Shape-Based Averaging (SBA) by additionally incorporating the similarity information between the reference (i.e., atlas) images and the target image to be segmented. We also propose a new spatially-varying similarity-weighted neighborhood prior model, and an edge-preserving smoothness term that can be used with many of the existing fusion methods. We first present our new Markov Random Field (MRF) based fusion framework that models the above mentioned information. The proposed methods are evaluated in the context of segmentation of lymph nodes in the head and neck 3D CT images, and they resulted in more accurate segmentations compared to the existing SBA.

**Index Terms**—Atlas-based segmentation, label fusion, medical imaging, MRF, SBA.

## I. INTRODUCTION

**I**N recent years, fusion methods that merge segmentations obtained from multiple atlases (i.e., reference images) have gained significant attention in medical imaging [1], [2]. Multi atlas methods are found to provide more accurate and robust results than single atlas based methods. The widely used fusion methods include majority voting (MV) [2], global weighted voting (GWV) [2], [3], local weighted voting (LWV) [2]–[4], STAPLE [5], and Shape Based Averaging (SBA) [1].

Manuscript received June 25, 2013; revised August 14, 2013; accepted August 19, 2013. Date of publication August 21, 2013; date of current version August 28, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Gustavo Kunde Rohde.

S. Gorthi is with the Signal Processing Laboratory (LTS5), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland (e-mail: subrahmanyam.gorthi@epfl.ch).

M. Bach Cuadra is with the Department of Radiology, CHUV and UNIL, and the Center for Biomedical Imaging (CIBM), CHUV Unit, Lausanne, and also with the Signal Processing Laboratory (LTS5), EPFL, Lausanne, Switzerland.

P.-A. Tercier and A. S. Allal are with the Service de Radio-oncologie, Hôpital Fribourgeois, Fribourg, Fribourg, Switzerland.

J.-P. Thiran is with the Signal Processing Laboratory (LTS5), EPFL, and also with the Department of Radiology, University Hospital Center (CHUV) and University of Lausanne (UNIL), Lausanne, Switzerland.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/LSP.2013.2279269

MV assigns for each pixel, a label that maximum number of atlases agree on. GWV, unlike MV, assigns a weight to the decision of each atlas while counting its vote; the weight for each atlas is determined globally, based on its similarity to the target image. LWV is similar to GWV except that, not a single global weight is assigned to the entire atlas; rather, for each pixel, an individual weight is assigned based on the local similarity to the corresponding pixel in the image to be segmented. Hence, GWV and LWV can be seen as the natural extensions to MV. One of the main problems with the aforementioned voting-based methods is that, although the segmentations obtained from each individual atlas are contiguous, the merged segmentations are often fragmented, containing unwanted holes and islands [6].

SBA, when compared to voting-based methods, looks at the fusion problem from a different perspective. It first computes in all the transformed atlases, how “deep” (inside), or how “far-away” (outside) a given pixel is from the contours of each label; this is tracked using “Signed Euclidean Distance” (SED); it then selects for each pixel a label that has the least resultant SED at that pixel when summed up over all atlases. Please refer to [1] for a pictorial illustration of this idea. SBA generally provides more contiguous segmentations than the voting-based methods [1]; this is because, unlike voting-based methods, SBA is based on the distances from each pixel to the contours of each label, and thus, it *implicitly* includes the neighborhood information of pixels.

Notice the parallels between the developments that took place in voting-based fusion methods (i.e., starting from MV, to GWV, and to LWV) versus SBA: SBA has limitations similar to MV in the sense that, both methods do not benefit from the information regarding how similar the atlas and the target image are; however, unlike GWV and LWV, no such extensions to SBA are proposed in the literature for incorporating the similarity information.

Finally, we notice that, from the transformed atlases, we can actually extract some prior knowledge about the neighborhood labels at each pixel, and for each possible label pair; however, none of the above methods use this additional information. The methods proposed in this letter are based on all the above mentioned observations.

The rest of the letter is organized as follows. In the next section, we present the general MRF-based framework for label fusion and the reformulation of SBA. Section III presents two new fusion methods and a new neighborhood-prior model. In Section IV, an evaluation of the methods in the context of lymph

nodes segmentation is presented. Finally, conclusions are presented in Section V.

## II. SHAPE-BASED AVERAGING (SBA)

In this section, we first present our MRF-based framework for performing label fusion, and then reformulate the existing SBA method for adapting it to the current context.

### A. MRF-Based Fusion Framework

Let  $V$  be the number of voxels in an image. Let  $Y_p$  denote the label assigned to the  $p$ th voxel in the output image. Let  $Y$  be the set containing labels assigned to each voxel in the output image, i.e.,  $Y = \{Y_1, \dots, Y_V\}$ . We will be formulating atlas fusion as an energy minimization problem of the form:

$$Y^* = \arg \min_Y \{E_{\text{data}}(Y) + \lambda E_{\text{smooth}}(Y) + \beta E_{\text{neighbor}}(Y)\}, \quad (1)$$

where the first term is a data term (unary term), and it should be defined in such a way that it reaches to a minimum value when the chosen fusion criteria has been met; the second term is a smoothness term (pairwise term) that performs edge-preserving smoothing of labels; the third term is a prior-information term (pairwise term) that incorporates neighborhood priors for each possible label combination.  $\lambda$  and  $\beta$  are weighting parameters for the smoothness term and neighborhood prior term respectively.

We now briefly describe how we solve the above optimization problem. Each voxel in the target image is represented by a vertex in a graph. Each possible label is represented by an additional node, and these are typically called “terminals.” Each of these terminals are connected to all vertices, and weights of those edges are equal to the resulting data cost for that voxel when it is assigned the label represented by that terminal. The edges that connect a vertex to its neighboring vertices are called “n-links,” and the weights of these n-links represent the costs coming from the corresponding pairwise terms. With such construction, the original problem is transformed into computing a “minimum cost cut” on the graph. Methods like graph cuts can be used for this purpose, provided that all the weights are nonnegative.

Note that although energy formulations similar to the above one were used previously [7] in performing segmentations using individual atlases, it is for the first time such MRF-based energy minimization formulation is used for *merging segmentations* obtained from multiple atlases. The main contributions of this letter are however modeling of the neighborhood prior term, and introducing weighted measures for SBA.

In [6], we have shown how MV, GWV and LWV can be reformulated to fit into the data term of the above formulation. Since the original formulation of SBA [1] can result in negative values as well, it cannot be directly used in the above framework. In order to deal with this problem, we have proposed in [8] an approach based on shifting and thresholding the signed distance values. However, such approach is not very elegant since it requires additionally, careful selection of the threshold value, and it is also not an exact equivalent to the original SBA formulation.

We now present a new logistic function-based approach that results in not only an exact equivalent to original SBA formulation, but also transforms the distances to nonnegative values.

### B. Reformulation of SBA

Let  $N$  be the number of atlases. Let  $X^j$  represent the  $j$ th input labeled image. Let  $u_p^j(l)$  represent the SED at  $p$ th voxel, in  $X^j$ , for label  $l$ . In the original formulation of SBA [1], for each voxel  $p$ , it assigns independently, that label which results in the minimum value of the following summation.

$$Y_p = \arg \min_l \frac{1}{N} \sum_{j=1}^N u_p^j(l). \quad (2)$$

As mentioned earlier,  $u_p^j$  can take any real value, and hence,  $Y_p$  can be negative; the goal here is to map  $Y_p$  to nonnegative values. Our following approach is primarily inspired by the work of [9] where they use Logarithm Odds maps, in a different context, for shape representation. The standard logistic function  $P(\cdot)$  that maps a variable  $Y_p \in \mathbb{R}$  to  $t_p \in (0, 1)$  is given by:

$$P(Y_p) = \frac{1}{1 + \exp(-Y_p)}. \quad (3)$$

Notice that the above logistic function maps negative values to the range  $(0, 0.5)$ , zero to 0.5, and positive values to  $(0.5, 1)$ . Hence, we propose to apply the above mentioned logistic transformation to  $Y_p$  in order to fit the existing SBA method into the data term of our MRF-based framework, and this results in the following data term:

$$E_{\text{data}}(Y) = \frac{1}{N} \sum_{p=1}^V P \left( \sum_{j=1}^N u_p^j(l) \right). \quad (4)$$

## III. NEW FUSION METHODS

This section presents the main contributions of this letter. We first present two new fusion methods, and then present a new neighborhood prior model that we propose to use along with these methods.

### A. Global Weighted Shape-Based Averaging (GWSBA)

As mentioned in Section I, SBA, similar to MV, does not take into account any similarity information of the atlases to the target image. Our goal here is, similar to GWV, to scale the contributions coming from various atlases in accordance to their global similarity to the target image. This results in the following data term:

$$E_{\text{data}}(Y) = \frac{1}{N} \sum_{p=1}^V P \left( \sum_{j=1}^N \hat{w}^j u_p^j(l) \right), \quad (5)$$

where  $\hat{w}^j$  is the normalized global weight assigned to the  $j$ th atlas.

### B. Local Weighted Shape-Based Averaging (LWSBA)

Notice that GWSBA assigns a single *global* weight for each atlas as a measure of the similarity. However, although two images may differ significantly in some particular regions, it is possible that they may be very similar in some other regions,

and LWSBA is indeed based on this observation. Instead of assigning a single similarity measure for the entire image (like in GWSBA), LWSBA computes similarity measure *locally* for each voxel, within a specified neighborhood. So, except in the computation of weight, LWSBA is similar to GWSBA, and thus, the resulting data term of LWSBA is as follows:

$$E_{\text{data}}(Y) = \frac{1}{N} \sum_{p=1}^V P \left( \sum_{j=1}^N \hat{w}_p^j u_p^j(l) \right), \quad (6)$$

where  $\hat{w}_p^j$  is the normalized local weight assigned to the  $j$ th atlas, at the  $p$ th voxel.

We have noticed that in a very recent work [10], a geodesic extension of SBA is proposed in the context of DTI tractography that includes a similarity measure into SBA. However, the goals and formulations of their work are different from our current work. In addition, [10] does not incorporate any neighborhood prior model that we are now going to present.

### C. New Neighborhood Prior Model ( $E_{\text{neighbor}}$ )

Let  $\mathbb{N}_p$  be the set of all voxels in the predefined neighborhood of the  $p$ th voxel. Let  $|\mathbb{N}_p|$  represent the cardinality of  $\mathbb{N}_p$ . Let  $\delta(\cdot)$  is the standard Kronecker delta function. Let  $X_p^j$  represent the label assigned to the  $p$ th voxel in the  $j$ th transformed atlas. The neighborhood prior information that we propose in this letter is as follows:

$$E_{\text{neighbor}}(Y) = \sum_{p=1}^V \sum_{q \in \mathbb{N}_p} \frac{1}{|\mathbb{N}_p|} \zeta(p, q, Y_p, Y_q), \quad (7)$$

where  $\zeta(\cdot)$  is a function that models neighborhood priors; if the label  $Y_p$  never occurs at the  $p$ th voxel in any of the  $N$  atlases, then we set  $\zeta$  to  $\infty$ ; when it occurs at least once, then we define  $\zeta$  as follows:

$$\zeta(p, q, Y_p, Y_q) = 1 - \frac{\sum_{j=1}^N \hat{w}_p^j \delta(X_p^j, Y_p) \delta(X_q^j, Y_q)}{\sum_{j=1}^N \hat{w}_p^j \delta(X_p^j, Y_p)}. \quad (8)$$

To get an intuitive understanding of this formulation, notice that, if we ignore the similarity weights (i.e.,  $\hat{w}_p^j$ ) for a moment, the summation in the denominator of the above equation computes the number of times the label  $Y_p$  occurred at the  $p$ th voxel, among all the  $N$  atlases; in the same way, the summation in the numerator computes the number of times  $Y_q$  occurred at  $X_q^j$  when  $Y_p$  occurred at  $X_p^j$ ;  $\hat{w}_p^j$  is introduced to simply scale the contributions coming from individual atlases based on their respective local similarities to the target image. Further, notice that the value of the ratio of the summations in the above equation is always in the range  $[0, 1]$ , and its value is maximum when the output label pair  $(Y_p, Y_q)$  to be assigned coincides with the label pair that occurred the most among the atlases, at  $(p, q)$  voxels; hence, this ratio is subtracted from 1 in order to transform it into an equivalent minimization problem.

Finally, the standard edge-preserving Potts model is used as the smoothness term since it preserves strong edges while penalizing for discontinuous label distributions; it is given by:

$$E_{\text{smooth}}(Y) = \sum_{p=1}^V \sum_{q \in \mathbb{N}_p} \frac{1}{|\mathbb{N}_p|} (1 - \delta(Y_p, Y_q)). \quad (9)$$

## IV. RESULTS

The evaluations are performed in the context of segmentation of lymph nodes in the 3D Head and Neck (H&N) CT images. 4 important lymph nodes are considered for automated segmentation: (i) IIA-Left, (ii) IIA-Right, (iii) IIB-Left and (iv) IIB-Right. The current database contains 13 images and among them, 8 images are considered for evaluating the fusion methods. The atlases for each target image to be segmented are chosen using the standard leave-one-out strategy, and thus, there are 12 atlases for each target image. The manual delineations of lymph nodes done by an expert oncologist are considered as the ground truth segmentations. Regarding the registration, all the 12 atlases are registered to each patient to be segmented. The registration procedure and the corresponding parameter values are same as those used in [11].

While presenting the results, the methods that additionally use the smoothness terms are denoted with a prime ( $'$ ), and the methods that use both the smoothness and the neighborhood prior term are denoted with a double prime ( $''$ ).

When using both the smoothness and the neighborhood prior terms, we set the weighting parameters such that the smoothness term has 40% of the relative weight given to the neighborhood prior term (i.e.,  $\lambda = 0.4\beta$ );  $\lambda$  value for the voting-based methods and the SBA-based methods is set to 0.5 and 3.0 respectively. Note that the data term is a function of number of votes for the voting-based methods, whereas it is a function of distances for the SBA-based methods; hence, these two classes of methods could require different scalings. All the above mentioned values are chosen empirically, based on visual inspections of the results for one of the images, and are not optimized further. Finally, in order to get some understanding of how sensitive are the results to the weighting parameters, we also present results for an extreme case, where, only the pairwise terms (i.e.,  $E_{\text{neighbor}}$  and  $E_{\text{smooth}}$ ) are used without any data term.

Fig. 1 shows, in one of the axial slices, the ground truth segmentations and the results obtained from the SBA-based fusion methods, both with and without the pairwise terms. The quantitative evaluations are performed over the entire dataset, using ‘‘Dice Similarity Metric’’ (DSM) and sensitivity; DSM is a commonly used metric [11] that measures the percentage of overlap between the ground truth and the automated segmentations; sensitivity is a measure of under-segmentation that computes the ‘‘true positive fraction’’ value. Fig. 2 presents box plots of average DSM values for the results obtained from the voting-based methods, SBA-based methods, and when using the pairwise terms alone. Finally, Table I presents mean and standard deviations of average DSM and sensitivity.

It can be noted from the above results that the proposed GWSBA and LWSBA methods have resulted in more accurate segmentations than the existing SBA. For the SBA-based methods, the inclusion of pairwise prior terms has significantly

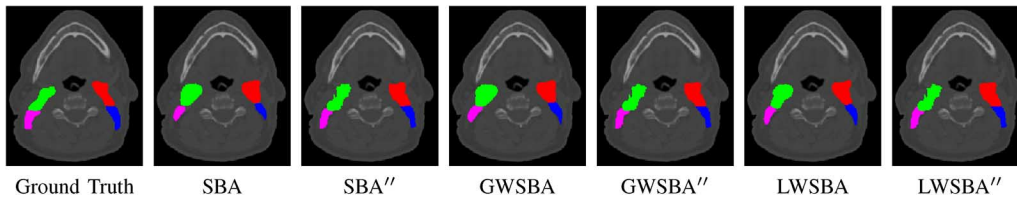


Fig. 1. H&N lymph nodes segmentations obtained from the SBA-based fusion methods in one of the axial slices.

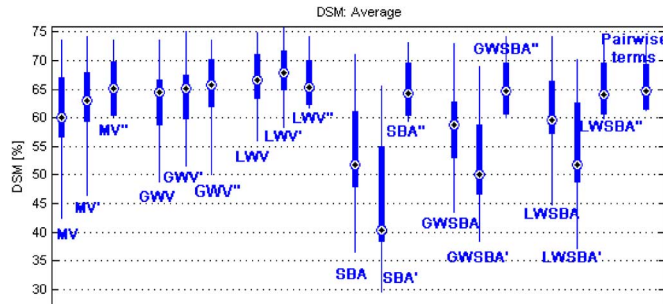


Fig. 2. Box plots of average dice similarity metric computed across all four lymph nodes.

TABLE I  
MEAN AND STANDARD DEVIATIONS OF AVERAGE DSM AND SENSITIVITY VALUES ACROSS ALL FOUR LYMPH NODES

	DSM	Sensitivity
MV	60.5 ± 9	55.7 ± 13
MV'	62.6 ± 8	61.0 ± 12
MV''	63.3 ± 10	56.6 ± 14
GWV	62.7 ± 8	63.7 ± 11
GWV'	64.0 ± 7	65.5 ± 11
GWV''	64.9 ± 8	58.3 ± 11
LWV	66.6 ± 6	65.7 ± 10
LWV'	67.9 ± 6	67.1 ± 10
LWV''	64.2 ± 9	57.3 ± 13
SBA	53.7 ± 11	46.1 ± 13
SBA'	45.3 ± 12	34.0 ± 13
SBA''	62.8 ± 10	55.1 ± 13
GWSBA	58.2 ± 9	53.0 ± 11
GWSBA'	52.4 ± 10	42.2 ± 11
GWSBA''	63.1 ± 10	55.7 ± 13
LWSBA	60.7 ± 9	54.2 ± 12
LWSBA'	54.2 ± 10	43.5 ± 13
LWSBA''	63.0 ± 10	55.5 ± 13
$E_{neighbor}$ and $E_{smooth}$	63.3 ± 10	55.9 ± 13

improved the results when compared to their counterparts that do not use such prior information. Another interesting observation from these results is that, while there is a significant difference among SBA, GWSBA and LWSBA in terms of segmentation accuracy, all these methods resulted in similar accuracy with the inclusion of neighborhood priors.

Among all the fusion methods, LWV' provided the best segmentation results. While the neighborhood prior term is found to be useful for the SBA-based methods, it is not always the case for the voting-based methods. Finally, notice that even when the pairwise prior terms alone are used without any data term, the results are surprisingly good, and are not too far from the best segmentation results obtained from LWV'. It also means that when fusion methods that include similarity information are used, the exact weight to be given to the neighborhood prior term is not very critical in obtaining good segmentation results. We further evaluated the statistical significance of the above mentioned im-

provements using Wilcoxon signed-rank tests. It is found (at 0.05 significance level) that the improvements in DSM from SBA to SBA'', from GWSBA to GWSBA'', from LWSBA to LWSBA'', from SBA to GWSBA, and from SBA to LWSBA are all statistically significant.

## V. CONCLUSIONS

In this letter, we have presented two new label fusion methods that extend the existing shape-based averaging (SBA) by additionally including the similarity information between the atlas and the target images. The proposed methods have significantly improved the accuracy when compared to the original SBA method. We have also proposed an MRF-based neighborhood prior model. The neighborhood prior model is found to be quite useful for the SBA-based methods, and the difference among these methods have become insignificant with the inclusion of this term.

## REFERENCES

- [1] T. Rohlfing and C. Maurer, "Shape-based averaging," *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 153–161, 2007.
- [2] X. Artaechevarria, A. Munoz-Barrutia, and C. Ortiz-de Solorzano, "Combination strategies in multi-atlas image segmentation: Application to brain mr data," *IEEE Trans. Med. Imag.*, vol. 28, no. 8, pp. 1266–1277, 2009.
- [3] M. Sabuncu, B. Yeo, K. Van Leemput, B. Fischl, and P. Golland, "A generative model for image segmentation based on label fusion," *IEEE Trans. Med. Imag.*, vol. 29, no. 10, pp. 1714–1729, 2010.
- [4] I. Işgum, M. Staring, A. Rutten, M. Prokop, M. Viergever, and B. van Ginneken, "Multi-atlas-based segmentation with local decision fusion—Application to cardiac and aortic segmentation in CT scans," *IEEE Trans. Med. Imag.*, vol. 28, no. 7, pp. 1000–1010, 2009.
- [5] S. K. Warfield, K. H. Zou, and W. M. Wells, "Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation," *IEEE Trans. Med. Imag.*, vol. 23, no. 7, pp. 903–921, 2004.
- [6] S. Gorthi, M. Bach Cuadra, U. Schick, P.-A. Tercier, A. S. Allal, and J.-P. Thiran, "Fusion of multi-atlas segmentations with spatial distribution modeling," in *MICCAI Workshop on Multi-Atlas Labeling and Statistical Fusion*, 2011.
- [7] F. van der Lijn, T. den Heijer, M. Breteler, and W. Niessen, "Hippocampus segmentation in MR images using atlas registration, voxel classification, and graph cuts," *NeuroImage*, vol. 43, no. 4, pp. 708–720, 12 2008.
- [8] S. Gorthi, M. Bach-Cuadra, U. Schick, P.-A. Tercier, A. S. Allal, and J.-P. Thiran, "Evaluation of atlas fusion strategies for segmentation of head and neck lymph nodes for radiotherapy planning," in *Proc. IEEE Int. Symp. Biomedical Imaging*, 2012.
- [9] K. Pohl, J. Fisher, M. Shenton, R. McCarley, W. Grimson, R. Kikinis, and W. Wells, "Logarithm odds maps for shape representation," in *Medical Image Computing and Computer-Assisted Intervention*, 2006, pp. 955–963.
- [10] M. Cardoso, G. Winston, M. Modat, S. Keihaninejad, J. Duncan, and S. Ourselin, "Geodesic shape-based averaging," in *Medical Image Computing and Computer-Assisted Intervention*, 2012, pp. 26–33.
- [11] S. Gorthi, V. Duay, X. Bresson, M. Bach Cuadra, S. Castro, F. Javier, C. Pollo, A. S. Allal, and J.-P. Thiran, "Active deformation fields: Dense deformation field estimation for atlas-based segmentation using the active contour framework," *Med. Image Anal.*, vol. 15, no. 6, pp. 787–800, 2011.