# A Generative Mold for Image Segmentation Based on Tag Union

***T.V. SAIKRISHNA[#1], Dr.A.YESUBABU[#2], Dr.A.ANANDARAO[#3]***
***Assoc. Professor, HOD & Professor, Principal &Professor,***
***Computer Science Department, QIS College of Engineering & Technology, Ongole, A.P., India,***
***Computer Science Department, SIR C R Reddy College of Engineering, Eluru, A.P., India,***
***Computer Science Department, JNTU College of Engineering, Anantapur, A.P., India,***

***Abstract***--We propose a nonparametric, probabilistic model for the automatic segmentation of medical images, given a training set of images and corresponding label maps. The resulting inference algorithms rely on pair wise registrations between the test image and individual training images. The training labels are then transferred to the test image and fused to compute the final segmentation of the test subject. Such label fusion methods have been shown to yield accurate segmentation, since the use of multiple registrations captures greater inter-subject anatomical variability and improves robustness against occasional registration failures. To the best of our knowledge, this manuscript presents the first comprehensive probabilistic framework that rigorously motivates label fusion as a segmentation approach. The proposed framework allows us to compare different label fusion algorithms theoretically and practically. In particular, recent label fusion or multitask segmentation algorithms are interpreted as special cases of our framework. We conduct two sets of experiments to validate the proposed methods. In the first set of experiments, we use 39 brain MRI scans with manually segmented white matter, cerebral cortex, ventricles and sub cortical structures—to compare different label fusion algorithms and the widely-used Free Surfer whole-brain segmentation tool. Our results indicate that the proposed framework yields more accurate segmentation than Free Surfer and previous label fusion algorithms. In a second experiment, we use brain MRI scans of 282 subjects to demonstrate that the proposed segmentation tool is sufficiently sensitive to robustly detect hippocampus volume changes in a study of aging and Alzheimer's Disease.

***Index Terms***—Image percolation, image registration, image segmentation.

## I. INTRODUCTION

**T**HIS paper investigates a probabilistic modeling framework to develop automatic segmentation tools that delineate anatomical regions of interest in a novel medical image scan. The objective is to learn a segmentation protocol from a collection of training images that have been manually labeled by an expert. This protocol is then employed by the algorithm to automatically segment a new (test) image. Such supervised segmentation tools are commonly used in many medical imaging applications, including surgical planning and the study of disease progression, aging or healthy development [23], [50], [74]. As an application domain, this paper focuses on magnetic resonance (MR) imaging of the brain. However, most of the ideas we discuss here can be easily extended to other modalities and applications, particularly with the recent development of fast algorithms for pairwise registration in other imaging domains We will thus consider the problem of segmenting the MRI volume scan of a novel subject, based on other subjects' MRI scans that have been delineated by an expert. Early MR segmentation algorithms mainly dealt with the problem of tissue classification, where local image intensity profiles contain a significant amount of the relevant information [10], [15]. A detailed parcellation of the brain anatomy into structurally or functionally defined regions of interest (ROI) typically requires supervision, commonly in the form of labeled training data, since the local appearance of most such structures is not easily distinguishable]. The training data is commonly obtained via a time-consuming and/or expensive procedure such as manual delineation, histology or functional localization experiments Automating the painstaking procedure of labeling improves the reliability and repeatability of the study, while allowing for the analysis of large pools of subjects. One of the simplest ways to automatically segment an image using a single training dataset is to perform a nonrigid, dense registration between the labeled image and test image. The resulting warp can then be used to map the training labels onto the coordinates of the test image [16]. The qualityof such a registration-based approach is limited by the accuracy of the pairwise registration procedure and the anatomical similarity between the labeled and test subjects. To reduce the bias due to the labeled subject and to model anatomical variability, multiple subjects can be employed in the training phase. A common method is to use a parametric model to summarize the training data in a common coordinate system [8]. In this approach the training

data are co-registered to compute probability maps that encode the prior probability of observing a particular label at each point in the common (atlas) coordinates. The test subject is then normalized to the atlas coordinates through a pairwise registration with a template image that represents the average subject. This registration can be completed as a preprocessing step, or can be unified with the segmentation procedure, as in [8]. Once the test subject is spatially ormalized, one can use a variety of models of shape and appearance to devise a segmentation algorithm. Traditionally, generative models have been popular, where simple conditionally independent Gaussian models are used for appearance [24], [48]. More sophisticated shape models that encourage certain opological properties have also been proposed to improve segmentation quality.

The central contribution of this paper is to propose and investigate a generative model that leads to label fusion style image segmentation algorithms. Within the proposed framework, we derive several methods that combine transferred training labelsinto a single segmentation estimate. Using a dataset of 39 brain MRI scans and corresponding label maps obtained from an expert, we experimentally compare these segmentation algorithms. Additionally, we compare against other benchmarks includingv FreeSurfer's whole brain segmentation tool, which has been widely used in a large number of studies , and STAPLE [67], a method that combines multiple segmentation estimates based on a probabilistic performance model. Our results suggest that the proposed framework yields accurate and robust segmentation tools that can be employed on large multisubject datasets. In a second experiment, we used one of the proposed egmentation
algorithms to compute hippocampal volumes in MRI scans of 282 subjects. A comparison of these measurements across clinical groups indicate that the proposed algorithm is sufficiently sensitive to robustly detect hippocampal volume differences associated with aging and early Alzheimer's Disease.

The generative model described in this paper is an extension of the preliminary ideas we presented in recent conference papers. The present paper offers detailed derivations, discussions and experiments that were not contained in those papers. The remainder of the paper is organized as follows. Sections II and III present the generative model and its instantiation, respectively. In Section IV, we develop several label fusion style segmentation algorithms based on the proposed generative model. Section V presents empirical results. In Section VI, we discuss the contributions of the paper along with the drawbacks of the proposed algorithms, while pointing to future research directions. Section VII concludes with a summary.

## II. GENERATIVE MODEL

In this section, we present the probabilistic model that forms the core of this paper.We use $\{\tilde{I}_n\}$ to denote training images with corresponding label maps , . We assume the label maps take discrete values from 1 to (including a "background" or "unknown" label) at each spatial location. While the training images are defined on a discrete grid, we treat them as spatially continuous functions on by assuming a suitable interpolator. Let be a finite grid where the test subject is defined.We denote to be the spatial mapping (warp) from the test subject coordinates to the coordinates of the th training subject. For simplicity, we assume that have been precomputed using a pairwise registration procedure, such as the one described in Appendix A. This assumption allows us to shorthand and with and , respectively, where we drop to indicate that we know the transformation that maps the training data into the coordinates of the test subject. The goal of segmentation is to estimate the label map associated with the test image . This can be achieved via maximum-a-posteriori (MAP) estimation

$$L=\underset{L}{\operatorname{argmax}}\, p(L|I; \{Ln, \tilde{I}n, \Phi n\})$$
$$=\underset{L}{\operatorname{argmax}}\, p(L,I; \{Ln, \tilde{I}n\}) \qquad (1)$$

where denotes the joint probability of the label map and image given the training data. Instead of using a parametric model for , we employ a nonparametric estimator, which is an explicit function of the entire training data, not a statistical summary of it, as shown in Fig. 1. The model assumes that the test subject is
generated from one or more training subjects, the index or indices of which are unknown. This modeling strategy is parallel to Parzen windowdensity estimators, where the density estimate can be viewed as a mixture distribution over the entire training data, and each new sample is associated with a single training
sample, the index of which is unknown and thus is marginalized over. In dealing with images, we may want to allow for this membership index to vary spatially. Therefore we introduce to denote the latent random field that specifies for each voxel in the test image , the (membership) index of the training image it was generated from.

Squares indicate nonrandom parameters, circles indicate random variables. Replications are illustrated with. The in the corner of the plate indicates the variables inside are replicated that many times (i.e., once for each voxel), and thus are

conditionally independent. Shaded variables are observed.

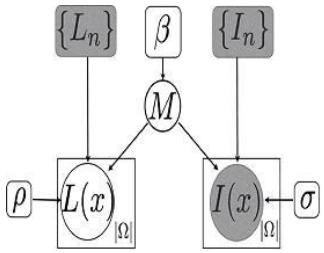



Fig. 1. Graphical model that depicts the relationship between the variables.

In the following, we make the assumption that the image intensity values and labels at each voxel are conditionally independent (as illustrated with a plate around these variables in Fig. 1), given the random field , and the training data . Furthermore, we assume that each voxel is generated from a single training subject indexed, which we will shorthand with and respectively. We can thus construct the conditional probability of generating the test image and label map.

P(L,I|M;{Ln, In})
=Π p(L(x), I(x)|M(x); {Ln,In}) (2)
=Π pM(x) (L(x); Lm(x))pM(x)(I(x);IM(x)) (3)

Given a prior on , we can view the image and label map as generated from a mixture model

P(L,I|M;{Ln, In})= ∑p(L(x), I(x)|M(x); {Ln,In}) (4)

where denotes the marginalization over the unknown random field . Substituting (3) into (4) yields the final cost function

The conditional independence assumption between the label map and image may seem simplistic at first. Yet conditional independence does not imply independence and the relationship between and is given by marginalizing over the unknown as in (4). Therefore, our model implicitly includes complex dependencies between labels and intensity values. For instance , a term commonly modeled explicitly in the segmentation literature can be expressed as

Thus, given a model instantiation, the conditional intensity distribution of a particular label at a location of interest can be estimated by examining the training subjects that exhibit that label in the proximity of the location of interest. This is exactly what atlas-based segmentation algorithms do, which underscores the similarity between the proposed probabilistic model and parametric models used in the literature. But nlike atlas-based methods that use a parametric model for ,the proposed framework explicitly employs the entire training data set.

## III. MODEL INSTANTIATION

This section presents the specific instantiations of the individual terms in that we use in this work to derive segmentation algorithms.

***A. Image Likelihood***

We adopt a Gaussian distribution with a stationary variance as the image likelihood term

For , reduces to an improper distribution , where is a constant. As we discuss in Section IV-B, this simple model leads to the Majority Voting strategy in label fusion, whereas for a finite , yields a weighted averaging strategy.

***B. Label Prior***

In this work, we investigate two representations to define the label prior term . One representation uses the logarithm of odds (LogOdds) model based on the signed distance transform [49]. Let denote the signed distance transform of label in training subject (in the native coordinates), which is assumed to be positive inside the structure of interest.We define the label prior as

where is the slope constant, is the partition function, and is the total number of labels including the background label. The prior encodes the conditional probability of observing label at voxel of the test
image, given that it was generated from the th training image. The second representation, commonly used in the literature, employs the probability vector image of : each voxel is a length- binary vector with the th entry equal to 1 if and 0 otherwise. To define the label prior , the transformation is applied to the robability vector image of In this method, non-grid locations need to be interpolated using a suitable method (e.g., trilinear or nearest neighbor) that ensures positive and normalized probability values. In general, it is well known that trilinear interpolation yields better segmentation results than nearest neighbor interpolation [51], [55]. The LogOdds model of (7) has the advantage of yielding nonzero probabilities everywhere, which makes the use of the logarithmof the probability numerically more stable. As discussed in our
experiments presented in Section V-A, we find that the LogOdds model produces more accurate results.

***C. Membership Prior***

The latent random field encodes the
local association between the test image and training data. We employ a Markov random field (MRF) prior on M

where is a scalar parameter, is a spatial neighborhood of voxel , is the partition function that only depends
on , and is the Kronecker delta. This particular type of MRF is often referred to as the Potts model. In our implementation, includes the immediate six neighbors of each voxel. Similar models have been used in the segmentation literature, mainly as priors on label maps to encourage the spatial relationships of labels observed in the training data. In contrast, we use the MRF prior to (indirectly) pool local intensity information from within a neighborhood in determining the association between the test subject and the training data. Here we adopt a simple form of the MRF that does not include singleton and/or spatially varying terms. This is unlike the common usage of MRFs in the segmentation literature where the label prior typically varies spatially.

The parameter influences the average size of the local patches of the test subject that are generated from a particular training subject. In this work, we consider three settings of the parameter . With , the model assumes that each test image voxel is generated from the training subjects with equal probability and that the membership is voxel-wise independent. forces the membership of all voxels to be the same and corresponds to assuming that the whole test subject is generated from a single unknown training subject, drawn from a uniform prior. A positive, finite encourages local patches of voxels to have the same membership.

## IV. LABEL FUSION BASED IMAGE SEGMENTATION

In this section, we derive several label fusion style image segmentation algorithms based on the model and MAP formulation described above. These algorithms correspond to variations in the image likelihood, label prior and membership prior models described in Section III. ***A. Local Weighted Voting***
Let us assume , which, thanks to the adopted simple MRF form, implies that is independent and identically
distributed according to a uniform distribution over all labels for all

where is the cardinality of the image domain (the number of voxels). Using the image likelihood term of (6), the segmentation problem in (5) reduces to

This optimization problem can be solved by simply comparing numbers at each voxel: the fused label of each voxel is computed via a local weighted (fuzzy) voting strategy. The local image likelihood terms serve as weights and the label prior values serve as votes. Therefore, at each voxel, training images that are more similar to the test image at the voxel *after* registration are weighted more. Interestingly, a similar approach was recently proposed in the context of CT cardiac segmentation by Isgum *et al*. where the transferred training labels are fused in a weighted fashion. The heuristic weights proposed in that paper have a different form however and are spatially smoothed with a Gaussian filter to pool local neighborhood information. In Section IV-D, we discuss a more principled approach to aggregate statistical information from neighboring voxels into the weighted label fusion procedure.

***B. Majority Voting***
Majority voting, which has been widely used as a label fusion method can be derived as a special case of *Local Weighted Voting*. The key modeling assumption is to set in the image likelihood term, effectively using an improper distribution and assigning equal weight to all training subjects, which reduces (10) to

If we use the probability vector image of to define the label prior,we arrive at the traditional ajorityVoting algorithm where each training image casts a single, unit vote, with no regards to the similarity between the training image and the test image. If one uses nearest neighbor interpolation, each vote corresponds to one particular label , whereas tri-linear interpolation yields a fuzzy voting strategy with each vote otentially spread over multiple labels .

***C. Global Weighted Fusion***
Here, we consider . As we now show, this results in an algorithm where, at each voxel, training images that are *globally* more similar to the test image *after* registration are weighted more. With , the membership prior defined in (8) only takes nonzero values if membership values at all voxels are equal, i.e.,

Thus, (4) is equivalent to a mixture model where the test subject is assumed to be generated from a single, unknown training subject

The segmentation problem in reduces to Equation (14) cannot be solved in closed form. However, an efficient solution to this MAP formulation can be obtained via expectation maximization (EM) [20]. Appendix B contains the derivations of the algorithm. Here, we present the summary.

1) *E-Step*
2) *M-Step*

The variational EM algorithm consists of two levels of iterations: the inner loop that repeatedly

computes (19) in the E-step and the outer loop that alternates between the E- and M-steps, until convergence. In the inner loop, at each iteration all ’s are updated using (19) and the neighboring values from the previous iteration. Once this inner loop converges, the algorithm updates the segmentation using (21). To determine the convergence of the outer loop, one can monitor the change in the segmentation estimate. In practice, we terminate the algorithm when less than a predetermined fraction, e.g., 0.01% of voxels change their segmentation estimate from one iteration to the next. Typically convergence is achieved in fewer than 10 iterations.

## V. EXPERIMENTS

In this section, we present two sets of experiments. In the first experiment we compare automatic segmentation results against manual delineations to objectively quantify the accuracy of segmentation algorithms. The second experiment employs a separate collection of brain MRI scans from 282 subjects to demonstrate that hippocampal volume measurements obtained using the proposed label fusion framework can detect subtle volume changes associated with aging and Alzheimer’s disease.

### *A. Experiment I: Comparison Against Manual Segmentation*

The first set of experiments employs 39 brain MRI scans and corresponding manual delineations of nine anatomical regions of interest (ROI) in two hemispheres. The images were selected from a large data set, including an Alzheimer’s cohort, the recruitment of which is described elsewhere. The 39 subjects were selected to span a wide age range and reflect a substantial anatomical variation due to dementia pathology. We note that these are the same subjects used to construct FreeSurfer’s released probabilistic segmentation atlas. Out of the 39 subjects, 28 were healthy and 11 were patients with questionable ( , Clinical Dementia Rating 0.5) or probable Alzheimer’s ( , CDR 1). Ten of the healthy subjects were young (less than 30 years), nine middle-aged (between 30 and 60 years), and nine old (older than 60 years). The MRI images are of dimensions 256 256 256, 1 mm isotropic voxels and were computed by averaging three or four scans. Each scan was a T1-weighted MP-RAGE, acquired on a 1.5 T Siemens Vision scanner. All scans were obtained in a single session. Acquisition details are as follows: TR 9.7 ms, TE 4.0 ms, TI 20 ms, Flip angle 10 . These high quality images were then gain-field corrected and skull-stripped. All the preprocessing steps were carried out using FreeSurfer tools [69]. The anatomical ROIs we used1 are white matter (WM), cerebral cortex (CT), lateral ventricle (LV), hippocampus (HP), thalamus (TH), caudate (CA), putamen (PU), pallidum (PA), and amygdala (AM). The labeling protocol we employed was developed by the Center for Morphometric Analysis and has been published and validated elsewhere [13]. An example segmentation obtained via the local weighted voting method of Section IV-A is visualized in Fig. 2.

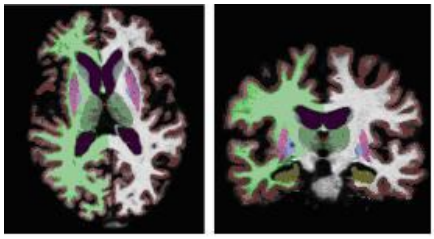


Fig. 2. A typical segmentation obtained with the local mixture model. 2D slices are shown for visualization only. All computations are done in 3D.

where denotes set cardinality. The Dice score varies between 0 and 1, with 1 indicating a perfect agreement between the two segmentations.

***1) Setting the Free Parameters Through Training:*** The proposed label fusion algorithms have two stages, registration and label fusion, each with several input parameters. To set these parameters properly, we initially performed *training* on nine out of the 39 subjects. These nine subjects were then only used as training subjects during the testing phase. Therefore, all results reported in this paper are on the remaining 30 subjects and reflect generalization performance. The registration stage has two independent parameters (as described in Appendix A): controls the step size in the Gauss–Newton optimization and determines the smoothness of the final warp. We registered 20 random pairs of the nine training subjects for a range of values of and . For each pair of subjects, we measured pairwise overlap by computing the Dice score between the warped manual labels of the “moving” subject and the manual labels of the “fixed” subject. We then selected that resulted in the best registration quality as measured by the average pairwise label overlap. The label fusion stage also has several independent parameters, depending on the method used. These include the standard deviation of the image likelihood in (6), the slope of the distance transform used to compute the label prior in (7), and the Markov weight which is nonzero for the semi-local method in Section IV-D and controls the average size of the image patches associated with the same training subject.

To determine , we performed nine leave-one-out segmentations on the training subjects using the Majority Voting method of Section IV-B and label

prior model of (7) for a range of values. The value that achieved the best segmentation accuracy was . We employed Local Weighted Voting (Section IV-A) and lobalWeighted Fusion (Section IV-C) to determine a local and global optimal value for (10 and 30), respectively. The optimal standard deviation for the local model was then used to determine the optimal value for (0.75) for the semi-local model.

We performed leave-one-out cross-validation on the 30 test subjects using these optimal parameters. For each test subject, all remaining 38 subjects were treated as training subjects.

***2) Comparison of Label Prior Models****:* Using the Majority Voting method (Section IV-B), we compare three different label

prior models (Section III-B): the LogOdds (based on the signed distance transform) model of (7) and two instantiations of the common approach that interpolates the vector image of indicator probability vectors, based on nearest neighbor interpolation (e.g., [30]) or tri-linear interpolation Fig. 3 shows a box-plot of Dice scores for these three different models and all the ROIs. These results indicate that the LogOdds representation provides a significantly better label prior for the label fusion framework.

***3) Comparison of Label Fusion Methods and Benchmarks:***

In this section we provide a comparison between the three weighted label fusion algorithms we derived in our framework and four benchmarks.

The second benchmark is the Majority Voting scheme based on the LogOdds prior, which is similar to the shape averaging method proposed in [52] and other voting based algorithms.

The whiskers extend to 2.7 standard deviations around the mean, and outliers are marked individually as a "*."

Finally, the training subjects that had the smallest SSD were used for majority voting. In the results we present here, we fix the number of training subjects that were used to 10 and call the algorithm "Majority10." Later in this section, we investigate the effects of varying the number of training subjects.

Majority Voting which has gained recent popularity [1], performs significantly worse than the weighted label fusion methods.

This result highlights the importance of incorporating image similarity information into the label fusion framework.

We note, however, that the results we report for our Majority Voting implementation are lower than the ones reported in. This might be due to differences in the data and/or registration

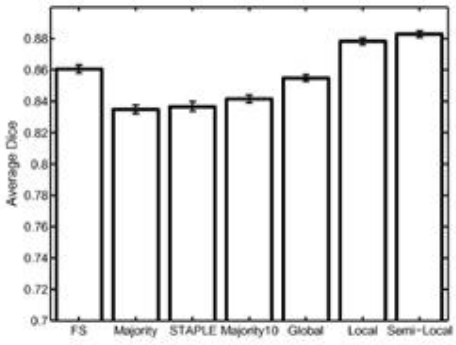
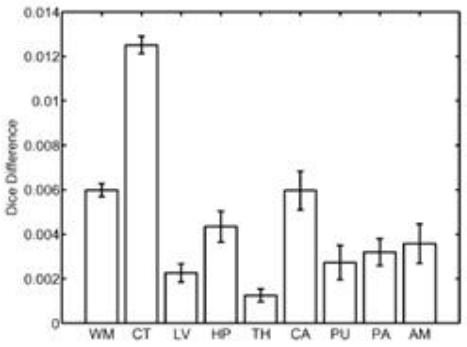



Fig. 5. Average Dice scores for each algorithm (FS: FreeSurfer,Majority: Majority Voting, STAPLE, Majority10, Global: Global Weighted Fusion, Local: Local Weighted Voting, and Semi-Local: Semi-local Weighted Fusion). Error bars show standard error. Each subject and ROI was treated as an independent sample with equal weight.

algorithm. Specifically, normalized mutual information (NMI) was used as the registration cost function in. Entropy-based measures such as NMI are known to yield more robust alignment results. We leave a careful analysis of this issue to future work.

Fig. 6. Average Dice differences: Semi-Local Weighted Fusion minus Local Weighted Voting. Overall, Semi-Local Weighted Fusion achieves better segmentation. Error bars show standard error.




Majority10 performs slightly better than Majority Voting. The improvement is particularly significant in subcortical ROIs such as the caudate. STAPLE, an alternative weighted fusion strategy, also yields slightly better average segmentation accuracy than Majority Voting. STAPLE's performance, however, is significantly worse than the three weighted label fusion algorithms derived based on the proposed probabilistic framework. Once again, this difference underscores the importance of

employing the MRI intensity information in determining the weights for label fusion. FreeSurfer, which we consider to represent the state-of-the art atlas based segmentation, on average, yields better segmentation accuracy than our remaining benchmarks. Yet we stress that FreeSurfer integrates registration and segmentation, while the performance of the remaining benchmarks were limited by our choice of the pairwise registration preprocessing step.

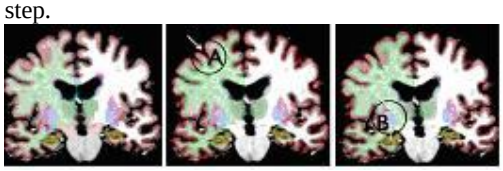



Fig. 7. The segmentations of the subject that Semi-localWeighted Fusion performed the worst on. Left to right: FreeSurfer, Global and Semi-localWeighted Fusion. Common mistakes (indicated by arrows): (A) Global Weighted Fusion tends to over-segment complex shapes like the cortex. (B) Semi-localWeighted Fusion does not encode topological information, as FreeSurfer does. Hence it may assign an "unknown" or "background" label (white) in between the pallidum (blue), putamen (pink), and white matter (green).

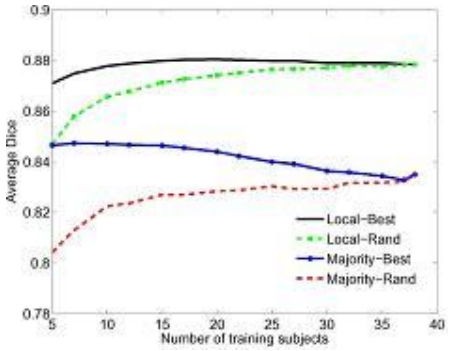



Fig. 8. The average Dice score for Majority Voting (Majority) and Local Weighted Voting (Local) as a function of the number of training subjects. We consider two strategies to select the training subjects: (1) randomly selecting a set of training subjects (Rand), (2) selecting the best training subjects that are globally most similar to the test subject (Best). The average Dice score reaches 83.9% for Majority Voting and 87.8% for Local Weighted Voting, when all 38 subjects are used.

***4) The Effect of the Number of Training Subjects:*** In theprevious section, for all the algorithms we employed a leaveone out validation strategy, where for each test subject all remaining 38 subjects were treated as the training data. In this section, we explore how the accuracy results vary as one varies the number of training subjects. This point has received considerable attention in prior work.

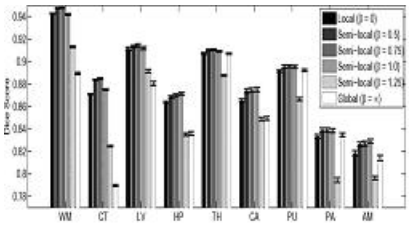



Fig. 9. Average Dice scores for different values in the MRF membership prior of (8). Error bars show standard error.

***5) The Effect of the MRF Prior:*** To investigate the effect of the MRF membership prior we applied the Semi-local Weighted Fusion method with four different values of to the 30 test subjects. Note that during the training phase, we established as optimal. Fig. 9 reports the average Dice scores for Semi-local Weighted Fusion with these values, Global Weighted Fusion, which corresponds to , and Local Weighted Voting, which corresponds to B.

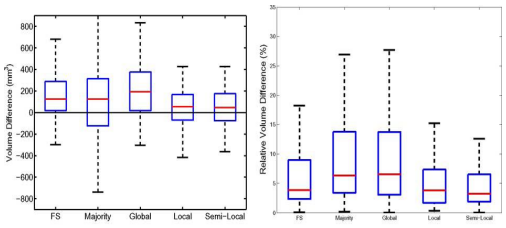
***6) Runtime:*** Table II lists the average run-times for the seven algorithms compared above. Majority10 and FreeSurfer are the fastest algorithms with less than 10 h of CPU time required for each test subject. ajority10 uses only 10 training subjects, which are globally most similar to the test subject as measured by the sum of squared differences after affine-normalization. The initial training subject selection stage takes about an hour.

The remaining three algorithms (STAPLE, Global and Semi-localWeighted Fusion) employ iterative optimization methods (EM,EMand variational EM, respectively) and require longer run-times. It is important to note that these run times can be reduced substantially using the same preselection strategy as
Majority10. In particular, our experiments with LocalWeighted Voting suggest that we can lower the run time of this method by at least a half with almost no reduction in accuracy.

***B. Experiment II: Hippocampal Volumetry***

In a second set of experiments, we aim to demonstrate that the proposed label fusion framework yields accurate volumetric measurements of the hippocampus. Hippocampal volume has been shown to correlate with aging and predict the onset of probable Alzheimer's Disease.

where denotes the computed volume of label in label map . These results indicate that both Local Weighted Voting and Semi-local Weighted Fusion provide more accurate hippocampal volume measurements than Global Weighted Fusion and Majority Voting




(a) (b)

Fig. 10. Hippocampal volume differences on the data from Experiment 1. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles. The whiskers extend to 2.7 standard deviations around the mean. (a) Automatic minus Manual volumes. (b) Relative volume differences [(23)].

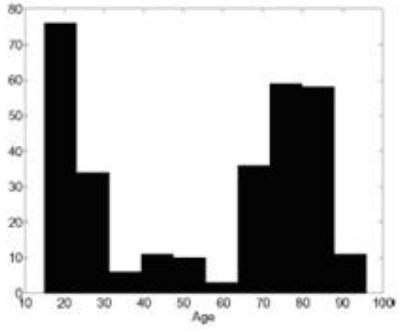
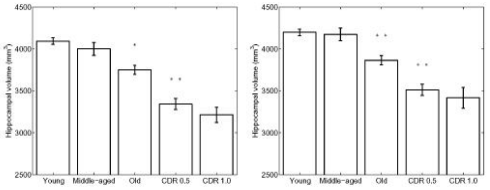



Fig. 11. Age histogram of 282 subjects in Experiment 2.

.




(a) (b)

Fig. 12. Hippocampal volumes for five different groups in Experiment 2.Error bars indicate standard error across subjects. Stars indicate that the volume measurements in the present group are statistically significantly smaller than the measurements in the neighboring group to the left. (Unpaired, single-sided t-test. (a) Left hippocampus. (b) Right hippocampus.

Fig. 12 shows the average hippocampal volume measurements for these five groups. Volumetric reduction due to aging and AD can be seen from this figure. These findings are in agreement with known hippocampal volume changes in AD and aging and demonstrate the use of the proposed label fusion method on a large pool of subjects, for which manual segmentation may not be practical.

## VI. DISCUSSION

Our experiments demonstrate the accuracy and usefulness of the label fusion framework as a segmentation tool. The proposed framework yields better accuracy than current state-of-the-art atlas-based segmentation algorithms, such as FreeSurfer.

The proposed framework should be viewed as an initial attempt to generalize segmentation algorithms based on label fusion, or a multi-atlas approach, which have recently shown promise and gained popularity with hardware advancements and developments of fast registration algorithms. In this paper, we investigated several modeling assumptions and derived four different instantiations of label fusion, one of which is the popular Majority Voting.

Majority Voting simply determines the most frequent label at each voxel, where each training subject gets an equal vote. Yet, recent work suggests that incorporating the similarity between the test image and training subjects can improve segmentation quality. For example, [1] employs a subset of training subjects
that are close in age to the test subject. Alternative strategies include using an image-based measure to quantify anatomical similarity, either at a local or global level. This similarity can then weigh the label votes during fusion, where more similar training subjects are given a larger weight.

Our theoretical development based on the proposed nonparametric probabilistic model yields three such algorithms, which solve the same problem for different settings of a single model parameter . This parameter controls the interactions between neighboring voxels in the Markov prior we construct on the latent membership random field encoding the (unknown) association between the test subject and training data. Smaller values allow for this association to vary more locally. Specifically, treats each voxel independently, whereas corresponds to assuming a single association for the whole image. A finite,
onzero encourages local patches of voxels to have the same membership.

These three cases are solved with different inference algorithms. The most efficient case corresponds to , where the global optimum can be computed via simple voxelwise counting. The other two cases are solved with more expensive iterative optimization methods, such as Expectation Maximization and Variational EM. Exact inference for the finite, nonzero case is intractable, yet our experiments suggest that approximate numerical solutions yield good segmentation accuracy.

The development of the proposed framework makes several simplifying assumptions. In the following, we discuss a number of directions that can be explored to relax these assumptions.We consider these as important avenues for future research, which promise to improve the performance of label fusion style segmentation.

1) In the graphical model of Fig. 1, we made the convenient assumption that the transformations are known and solved for these in a preprocessing pairwise registration step (see Appendix A). Ideally, however, one would like to integrate over all possible transformations, which has a prohibitively high computational cost. Recent work attempted to approximate this integration for a *single* registration [3]. A more practical approach is to compute the registrations jointly with the segmentations, cf. [8]. Here, we avoided this particular route, since the multiple registrations performed between the test subject and training data were already computationally challenging.

2) The simple additive Gaussian noise model presented in Section III-A has two crucial consequences: 1) the registration cost function is a sum of squared intensity differences, and 2) in weighted label fusion, the eights are a function of sum of squared intensity differences, i.e., anatomical similarity is measured based on squared differences of intensity values. This model makes the algorithm sensitive to intensity variations due to imaging artifacts. Thus, the presented algorithms are only suitable for intensity-normalized images. An alternative strategy is to employ a more sophisticated image likelihood model that would motivate
information theoretic similarity measures, such as mutual information.

3) The main drawback of label fusion style algorithms is the computational complexity introduced by the multiple pairwise registrations and the manipulation of the entire training data. Traditional atlas-based segmentation approaches avoid this problem by using parametric models of anatomical variation in a single coordinate system. In recent work, we used a mixture modeling strategy, called iCluster, to model anatomical heterogeneity with multiple atlases. We believe a combination of the label fusion framework presented in this paper and iCluster can be employed to reduce the computational burden by summarizing the training data with a small number of *templates*.

4) An alternative strategy to reduce the computational demand of label fusion is to employ a nonparametric model in a single coordinate system, to which the test subject is normalized with a single registration procedure. This approach, which entails the co-registration of the training subjects akin to atlas-based segmentation, was recently shown to produce accurate segmentation [5], [17]. The application of this strategy within the proposed label fusion framework is a direction to be explored.

5) Another strategy to reduce computational burden is to preselect the most useful training subjects and apply label fusion on these, as recently proposed by Aljabar *et al*. [1]. We explored one particular instantiation of this approach, where the subset of training subjects was selected to include the training subjects globally most similar to the test subject after affine normalization. It is clear that this criterion to preselect the most relevant training subjects is related to our definition of the image likelihood term . Yet, a crucial difference is that the image likelihood term is computed by nonlinearly registering the training and test images, while the preselection is done based on an affine normalization. Alternative preselection strategies should also be investigated.

## VII. CONCLUSION

In this paper, we investigated a generative model that leads to label fusion style image segmentation methods. Within the proposed framework, we derived several algorithms that combine transferred training labels into a single segmentation estimate.With a dataset of 39 brainMRIscans and corresponding label maps obtained from an expert, we empirically compared these segmentation algorithms with FreeSurfer's widely-used atlas-based segmentation tool. Our results demonstrate that the proposed framework yields accurate and robust segmentation tools that can be employed on large multi-subject datasets. In a second experiment, we employed one of the developed segmentation algorithms to compute hippocampal volumes inMRIscans of 282 subjects. A comparison of these measurements across clinical and age groups indicate that the proposed algorithms are sufficiently sensitive to detect hippocampal volume differences
associated with early Alzheimer's Disease and aging.

## REFERENCES

[1] P. Aljabar, R. A. Heckemann, A. Hammers, J. V. Hajnal, and D. Rueckert, "Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy," *Neuroimage*, vol. 46, no. 3, pp. 726–738, 2009.
[2] S. Allassonnière, Y. Amit, and A. Trouve, "Towards a coherent statistical framework for dense deformable template estimation," *J. R. Stat. Soc. B*, vol. 69, pp. 3–29, 2007.
[3] S. Allassonnière, E. Kuhn, and A. Trouvé, "MAP estimation of statistical deformable template via nonlinear mixed effect models: Deterministic and stochastic approaches," *Math.*

*Foundations Computat. Anat. (MFCA) Workshop MICCAI 2008 Conf.*, 2008.
[4] V. Arsigny, O. Commowick, X. Pennec, and N. Ayache, "A log-Euclidean framework for statistics on diffeomorphisms," in *Proc. Of MICCAI*. New York: Springer, 2006, vol. 4190, Lecture Notes Computer Science, pp. 924–931.
[5] X. Artaechevarria, A. Munoz-Barrutia, and C. O. de Solorzano, "Efficient classifier generation and weighted voting for atlas-based segmentation: Two small steps faster and closer to the combination oracle," *SPIE Med. Imag. 2008*, vol. 6914, 2008.
[6] X. Artaechevarria, A. Munoz-Barrutia, and C. O. de Solorzano, "Combination strategies in multi-atlas image segmentation: Application to brainMRdata," *IEEE Trans. Med. Imag.*, vol. 28, no. 8, pp. 1266–1277, Aug. 2009.
[7] J. Ashburner, "A fast diffeomorphic image registration algorithm," *Neuroimage*, vol. 38, no. 1, pp. 95–113, 2007.
[8] J. Ashburner and K. Friston, "Unified segmentation," *Neuroimage*, vol. 26, pp. 839–851, 2005.
[9] B. B. Avants, M. Grossman, and J. C. Gee, "Symmetric diffeomorphic image registration: Evaluating labeling of elderly and neurodegenerative cortex and frontal lobe," in *Biomedical Image Registration*. New York: Springer, 2006, vol. 4057, LNCS, pp. 50–57.
[10] S. Awate, T. Tasdizen, N. Foster, and R. Whitaker, "Adaptive Markov modeling for mutual-information-based, unsupervised MRI brain-tissue classification," *Med. Image Anal.*, vol. 10, no. 5, pp. 726–739, 2006.
[11] D. Blezek and J. Miller, "Atlas stratification," *Med. Image Anal.*, vol. 11, no. 5, pp. 443–457, 2007.
[12] P. Cachier, E. Bardinet, D. Dormont, X. Pennec, and N. Ayache, "Iconic feature based non-rigid registration: The PASHA algorithm," *Comput. Vis. Image Understand.*, vol. 89, no. 2–3, pp. 272–298, 2003.
[13] V. S. Caviness, P. A. Filipek, and D. N. Kennedy, "Magnetic resonance technology in human brain science: Blueprint for a program based upon morphometry," *Brain Develop.*, vol. 11, pp. 1–13, 1989.
[14] G. E. Christensen and H. J. Johnson, "Consistent image registration," *IEEE Trans. Med. Imag.*, vol. 20, no. 7, pp. 568–582, Jul. 2001.
[15] C. Cocosco, A. Zijdenbos, and A. Evans, "A fully automatic and robust brain MRI tissue classification method," *Med. Image Anal.*, vol. 7, no. 4, pp. 513–527, 2003.
[16] D. L. Collins, C. J. Holmes, T. M. Peters, and A. C. Evans, "Automatic 3-d model-based neuroanatomical segmentation," *Human Brain Mapp.*, vol. 3, no. 3, pp. 190–208, 1995.
[17] O. Commowick, S.Warfield, and G. Malandain, "Using Frankensteins creature paradigm to build a patient specific atlas," in *Proc. of MICCAI 2009*. New York: Springer, 2009, vol. 5762, Lecture Notes Computer Science, pp. 993–100.

## Authors Profiles:

**Mr. T.V.Sai Krishna** currently working as Associate Professor in Dept of Computer Science and Engineering. He Received his B.Tech from JNTU, Hyderabad and M.Tech from JNTU, Ananthapur. He had 9 years of Teaching Experience and Life Member of ISTE. He has published several papers at various national and International Journals and Conferences. His Research area includes Image Processing, Data mining, and Network Security.

**Dr. ADIMULAM YESU BABU** Presently working as Professor & Head of the Department of Computer Science & Engineering, Sir CRR College of Engineering,Eluru,Andhra Pradesh, India. He had around 23 years of Academic & Administration experience. Well versed with technical writing and editing. Reviewing research papers for International journals. Reviewing journal papers for Journal of Computational Biology and Bioinformatics Research (**JCBBR).** He had been Reviewing journal papers for International Journal of Biometrics and Bioinformatics (IJBB) **Reviewer for CiiT International Journal of Data Mining Knowledge Engineering.** He had been working as a teacher in various capacities i.e. Lecturer, Reader, Associate professor& Professor since 1988. Well versed with procedures and practices of AICTE, JNTU, AU, NBA and the Government concerning establishment, of new college / courses / renewals / accreditation etc.

**Dr. Anand Rao Akepogu**. recieved B.Sc (M.P.C) degree from Sri VENKATESWARA University, Andhra Pradesh, India. He received B.Tech degree in Computer Science & Engineering from University of Hyderabad, Andhra Pradesh, India and M.Tech degree in A.I & Robotics from University of Hyderabad, Andhra Pradesh, India. He received PhD degree from Indian Institute of Technology, Madras, India. He is currently working as a Principal and also as a Professor of Computer Science & Engineering Department of JNTU College of Engineering, Anantapur, Jawaharlal Nehru technological University, Andhra Pradesh, India. Dr. Rao published more than twenty research papers in international journals and conferences. His main research interest includes software engineering and data mining.