Image Segmentation Using an Efficient Rotationally Invariant 3D Region-Based Hidden Markov Model

Albert Huang
Dept. of ECE, UBC, Canada
alberth@ece.ubc.ca

Rafeef Abugharbieh
Dept. of ECE, UBC, Canada
rafeef@ece.ubc.ca

Roger Tam
Dept. of Radiology, UBC, Canada
roger@msmri.medicine.ubc.ca

Abstract

We present a novel three dimensional (3D) region-based hidden Markov model (rbHMM) for unsupervised image segmentation. Our contributions are twofold. First, our rbHMM employs a more efficient representation of the image than approaches based on a rectangular lattice or grid; thus, resulting in a faster optimization process. Second, our proposed novel tree-structured parameter estimation algorithm for the rbHMM provides a locally optimal data labeling that is invariant to object rotation. We demonstrate the advantages of our segmentation technique by validating on synthetic images of geometric shapes as well as both simulated and clinical magnetic resonance imaging (MRI) data of the brain. For the geometric shape data, we show that our method produces more accurate results in less time than a grid-based HMM framework using a similar optimization strategy. For the brain MRI data, our white and gray matter segmentation results in substantially greater accuracy than both block-based 3D HMM estimation and expectation-maximization hidden Markov random field (HMRF-EM) approaches.

1. Introduction

Hidden Markov models (HMM) have been successfully used in numerous applications such as speech recognition [1][2], document classification [3][4], recognition of characters [5], gestures [6] and facial features [7][8], segmentation of video [9], color images [10] and semantic images [11][12], regional climate forecast [13], target tracking [14], and protein sequence comparisons [15]. The theory of hidden Markov models was originally developed by Baum [16] with an underlying basic one dimensional (1D) Markov chain [17]. In simple terms, a HMM is a Markov source - a conditionally independent process on a Markov chain or a Markov chain on a memoryless noisy channel [18][19]. This Markovian assumption states that the current system state at any given time depends only on the immediate preceding state with fixed transitional probabilities. In an HMM, an observation is generated as a random function of the current state probability distribution; therefore, it is not possible to determine the current state by relying only on the current observation. To estimate the model parameters, the Baum-Welch [20] maximum likelihood algorithm is usually used. An efficient approximation, the Viterbi training [21][22], uniquely assigns the observation to the state with the maximum a posteriori (MAP) probability.

To extend HMM to two dimensional (2D) images, a pseudo 2D HMM [23] has been proposed. The model is pseudo in the sense that instead of a fully connected 2D HMM, a hierarchy of 1D HMM is used. A truly 2D HMM was developed by representing images as Markov meshes [24]. Further extensions to three dimensional (3D) images include the pseudo 3D HMM [8], grid-based 3D HMM [25], and block-based 3D HMM [26].

A popular alternative paradigm for modeling 2D and 3D Markov processes is the Markov random field (MRF) [27][28]. As an undirected graphical model based on a predefined local neighborhood system of sites, an MRF is a probabilistic stochastic relaxation approach that converges approximately to the MAP estimate. MRF-based approaches have been adopted for the labeling of watershed segments [29][30], region adjacency graph processing [31], rotational invariant texture feature classification [32], and the hidden MRF expectation-maximization (HMRF-EM) segmentation of medical images [33].

Ibrahim *et al.* [26] have recently shown that there are significant advantages in the reduced complexity of HMMs over HMRFs, particularly in large computational problems. However, as image resolutions increase with continuous advances in imaging technologies, there is a need for greater efficiency in 3D HMM representations and algorithms as some computational tasks would otherwise still be prohibitively expensive. Recently, Joshi *et al.* [34] derived a computationally efficient 3D HMM parameter estimation method that iteratively decodes each row of voxels using neighboring rows in a 3D rectangular lattice or grid. Though this approach is memory efficient, results show that the method is not invariant to object rotations.

In this paper, we present a new 3D region-based model in a HMM framework (henceforth referred to as rbHMM) and a novel tree-structured parameter estimation algorithm for efficient, accurate and rotationally invariant image

segmentation that can be especially valuable in medical imaging applications where noise and patient orientation in the scanners often complicate segmentation tasks. Our 3D rbHMM organizes volumetric images into homogeneous sub-regions similar to the concept of superpixels [35]. The proposed parameter estimation algorithm then iteratively estimates the model parameters and the MAP labeling of each region using a non-uniform representation of voxel data in a 3D space. The representation is non-uniform in the sense that we do not incur any a priori directionality assumptions on the data. Validation of the proposed segmentation technique was performed on synthetic 3D geometric shapes as well as simulated and clinical T1-weighted brain magnetic resonance image (MRI) volumes. Results demonstrated superior performance in both accuracy and efficiency measures.

The remainder of the paper is organized as follows. In §2, we introduce the proposed 3D rbHMM model. We then present an unsupervised, computationally efficient algorithm for estimating the MAP states using 1D tree-structured Markov chains. Lastly, we present the experimental results in §3 and conclude in §4.

2. Methods

In this section, we present the motivation for grouping voxels into local regions, from which our 3D region-based hidden Markov model is derived. The term regional has previously been used in HMM [13][14] but generally in a geographical sense rather than based on image properties. We describe the proposed model in its general form along with the statistical assumptions made. We then present an efficient extension of the Viterbi algorithm for estimating the 3D rbHMM parameters and the MAP states.

2.1. Computing Image Regions

The proposed algorithm first groups image voxels into contiguous local regions. The motivations behind this grouping is that voxels are typically a discrete representation of a continuous scene and as such are not independent; similar regions exhibit similar properties such as intensity or texture and therefore can be grouped so that the complexity and computational cost are reduced from the number of voxels to the number of regions observed in the image. In this work, we apply the watershed transform [35] for simplicity. Alternative methods such as normalized cuts [36] can be employed to further incorporate texture cues.

Determining the watershed depth based on image gradient magnitude is non-trivial due to the wide variability in image data and image quality. Here, we adopt a procedure similar to the threshold selection process in most edge detectors, namely, by examining the cumulative distribution function of gradient magnitudes. Instead of defining a specific gradient value, a percentage threshold,

T_w, on the cumulative distribution is chosen to ensure over-segmentation and preservation of image structures. This ensures a threshold value that is stable with respect to the image intensity scale and is insensitive to small changes in the threshold value, making it applicable to a wider range of images.

2.2. Region-Based Hidden Markov Model

We propose a 3D HMM governed by the immediate neighborhood of each delineated region in a non-uniform 3D space. The representation does not follow any a priori directionality assumption, and the ordering of the model nodes is data-driven. Unlike that of Li *et al.* [25], which follows a strict lexicographic order of voxels from left-to-right, top-to-bottom, and front-to-back, or that of Li *et al.* [37], which follows a center-to-boundary ordering, the proposed region-based model fundamentally allows for improved computational efficiency and invariance to object rotation. We describe the model as follows.

Given a 3D image U of dimensions $W \times H \times D$, we denote a voxel with 3D coordinates (x, y, z) along the X, Y and Z axes. U can, therefore, be described as a collection of the intensity values of all the voxels in the 3D grid: $U=\{ I(x,y,z): 1 \le x \le W, 1 \le y \le H, 1 \le z \le D \}$. By dividing an image into regions, each voxel in U is assigned a region label $L(x,y,z)$. A region r_i is then defined as $r_i=\{ (x, y, z): L(x, y, z)=i \}$ with $N_i=\|r_i\|$ reflecting the size of the region. Consequently, we can describe U as a set of regions $R=\{ r_i: 1 \le i \le N_R \}$, where $N_R=\|R\|$ is the total number of regions. In addition, we denote the neighboring adjacent regions of r_i in 3D space as $R_i'=\{ r_j: j \neq i, \exists (x,y,z) \in r_i \text{ s.t. } (x \pm 1,y,z) \text{ or } (x,y \pm 1,z) \text{ or } (x,y,z \pm 1) \in r_j \}$ and $N_i'=\|R_i'\|$ is the number of neighbors of r_i.

We can represent each r_i by using a mean regional feature vector, f_i. For grayscale images such as MRI, we can define f_i simply as the mean observed voxel intensities:

$$f_i = \sum_{(x,y,z) \in r_i} I(x,y,z) \Big/ N_i, \qquad 1 \le i \le N_R \qquad (1)$$

If the image is ideal and the image acquisition is perfect, i.e. noiseless and artifact free, all r_i with the same true underlying state s_i will share the same observed feature f_i. However, images are often corrupted by noise such that, given f_i, the true state s_i can be optimally estimated in the MAP sense by using the proposed 3D rbHMM. We inherit the following model assumptions from the classical HMM:

Assumption 1: The observed features for model state $l \in \{ 1, 2, \dots, N_s \}$ given N_s number of true underlying states follows a Gaussian distribution such that the distribution parameters (mean μ_l, and variance σ_l^2) can be readily estimated from the sample observation.

Assumption 2: If s_i is known, then f_i is conditionally independent of the rest of the other regions.

Assumption 3: s_i is governed by an irregular Markov mesh with transitional probabilities: $P(s_i=l \mid n_i)=\alpha_{n_i,l}$, where

$n_i = \{ (s_i'', f_i''): r_i'' \in R_i'' \}$ denotes a set of states and features of the preceding neighbors, R_i'', of r_i in the irregular mesh. R_i'' is a subset of R_i' where states s_i'' are known, and $\| R_i'' \| = N_i''$ is the total number of preceding neighbors.

Figure 1 shows a 2D representation of rbHMM based on the lexicographic order of regions from top-left to bottom-right. This model shows that each node transition is affected only by the immediate preceding neighbors, in this case, $R_i'' = \{ r_j: j < i \}$. This representation is, however, not unique for any given region map as one can traverse the image regions in many different ways, e.g. from bottom-left to top-right, depending on the node number assignment. For real images, the ordering of regions and the transitional dependencies are based on the image features and cannot be pre-determined. The next section explains the computation of this ordering in detail.
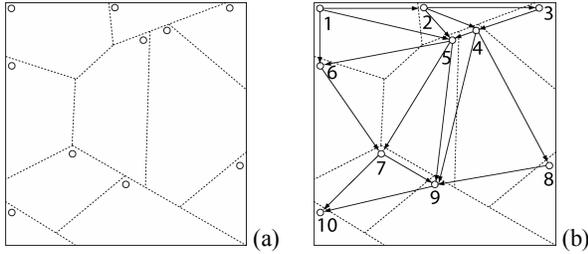

Figure 1: A 2D example of rbHMM. (a) Pre-delineated region map; each region is represented as a node in a hidden Markov model; (b) Representation of connected nodes based on the lexicographic order (indicated by the node number) of regions from top-left to bottom-right.

2.3. Model Parameter Estimation

Unlike a grid-based 3D HMM [25][34], where a total of $N_s^{(W \times H \times D)}$ possible state combinations exist, the proposed 3D rbHMM greatly reduces the state search space to $N_s^{N_R}$. We adopt the Viterbi algorithm used in 1D HMM estimation to iteratively search for the optimal MAP states, $\{ s_i^*: 1 \leq i \leq N_R \}$, given the current model parameters by locally optimizing the likelihood of the observation. The unknown states are estimated, and the model parameters are then updated assuming the MAP states, s_i^*, are the true underlying states.

We initialize s_i^* using K-means clustering [38] on f_i for simplicity. Other alternative unsupervised learning algorithms such as mixture modeling [39] can also be employed. To compute the model parameters $[\mu_l, \sigma_l^2, \alpha_{n_i,l}]$ for each state $l \in \{ 1,2,\ldots,N_s \}$, the mean, μ_l, and variance, σ_l^2, are estimated as the sample mean and the sample variance of all observations, and the transition probabilities, $\alpha_{n_i,l}$, are computed based on the empirical frequencies and subject to the constraint $\sum_{i=1}^{N_R} \alpha_{n_i,l} = 1$ for $1 \leq l \leq N_s$. These model parameters are then iteratively updated for unsupervised estimation without priors or training. The key computational issue in applying the

Viterbi algorithm to the 3D rbHMM is to solve for the MAP states, s_i^*, given a set of parameters with no dependencies or directionality assumption.

Thus, we propose the following iterative algorithm to search for the MAP states, where, to reiterate, R_i' denotes the set of all neighbors of region r_i, R_i'' denotes the set of preceding neighbors of r_i, and s_i and f_i denote the state and observed feature of r_i.

Step 1. Initialize iteration number $t=1$.
Step 2. Initialize $s_i^{(t-1)}$ using K-means for $1 \leq i \leq N_R$.
Step 3. Initialize $\mu_l, \sigma_l^2, n_i, \alpha_{n_i,l}$ for $1 \leq i \leq N_R$, $1 \leq l \leq N_s$.
Step 4. Initialize tree branch depth $d=0$.
Step 5. Initialize current and processed region sets:
$C^d = C^0 = \{ r_i \}$ where i=rand($\{ 1,2,\ldots,N_R \}$). $P^d = P^0 = \{ \varnothing \}$.
Step 6. For all $r_i \in C^d$, determine:
$S_i = \{ s_j^t: r_j \in R_i'' \} \cup \{ s_j^{(t-1)}: r_j \in R_i' \setminus R_i'' \}$
$F_i = \{ f_j: r_j \in R_i' \}$
Step 7. Update the processed region set by including the current region set:
$P^{(d+1)} = \{ P^d \cup C^d \}$.
Expand the current region set with neighbors of the current region set that have not yet been processed:
$C^{(d+1)} = \{ \cup R_i' \} \setminus P^{(d+1)}, r_i \in C^d$
Step 8. If ($\| C^{(d+1)} \| > 0$) d++ and goto Step 6, else t++.
Step 9. Search for MAP states s_i^* given S_i and F_i.
Step 10. Update $s_i^t = s_i^*$ for $1 \leq i \leq N_R$.
Step 11. Update $\mu_l, \sigma_l^2, n_i, \alpha_{n_i,l}$ for $1 \leq i \leq N_R$, $1 \leq l \leq N_s$.
Step 12. If (t<max iterations) go to Step 4.
Step 13. Return s_i^t for $1 \leq i \leq N_R$.

As can be seen above, the algorithm starts at a random region in each iteration t. A nested process (Steps 6 to 8) iteratively constructs a Markov tree by extending one step outward from all nodes that have already been formed towards the immediate neighboring regions. The direction of the tree branching at each node follows a path of the most influential neighbor, which is determined based on the sum of two ratios defined as region adjacency, $a_R(j,i)$, and edge adjacency, $a_E(j,i)$ between neighboring regions r_j and r_i:

$$a_R(j,i) = area(r_j) \Big/ \sum_{r_k \in R_i'} area(r_k) \qquad \begin{cases} 1 \leq i, j \leq N_R \\ r_j \in R_i' \end{cases} \quad (2)$$

$$a_E(j,i) = edge(r_j, r_i) \Big/ \sum_{r_k \in R_i'} edge(r_k, r_i) \qquad (3)$$

where $area(r_j)$ denotes the area of region r_j, and $edge(r_j, r_i)$ denotes the number of edge points between regions r_i and r_j. We note that the direction of the tree branching remains the same under image rotations since the regional neighborhood relationship has not changed; thus, allowing a rotationally invariant description of the image data.

For each tree branch, 1D Viterbi is used to search for the

optimal MAP states, s_i^*, given in equation (4).

$$s_i^* = \arg\max_{s_i} P\left(s_i \middle| f_i, \bar{s}_j, f_j : r_j \in R_i'\right)$$

$$= \arg\max_{s_i} P\left(s_i, f_i, \bar{s}_j, f_j, \bar{s}_k, f_k : r_j \in R_i'', r_k \in R_i' \setminus R_i''\right)$$

$$= \arg\max_{s_i} P\left(f_i \middle| s_i\right) \times P\left(s_i \middle| \bar{s}_j : r_j \in R_i''\right)$$

$$\times P\left(\bar{s}_k \middle| s_i, \bar{s}_j : r_j \in R_i', r_k \in R_i' \setminus R_i''\right) \quad (4)$$

The first term on the right hand side of the last equality sign for equation (4) is the likelihood of the observation belonging to the state, whereas the second and the third terms indicate the transition probabilities. This derivation is based on [34], where $\bar{s}$ is used to distinguish the conditioned states of the neighboring regions from the state, s, of the current region to be optimized. The difference is that instead of locally optimizing 1D rows on a 3D grid, s_i^* is calculated along each 1D model branch. For nodes that lie on multiple branches, a majority voting rule is applied after all the branches have been optimized.

Since we work on an irregular representation rather than a grid mesh, $\alpha_{n_i,l}$ is generated and iteratively updated as a codebook of neighboring states given state l. Each region in the image often yields a unique codebook entry with the probability determined based on the region area ratio. We define an overall neighborhood adjacency overlap ratio, $a_O(i, j)$, in (5) such that any pair of entries, i and j, with significant overlap (>70%), are merged using a weighted sum of both transition probabilities based on the area ratios of their represented regions. Note that this codebook merging is performed for efficient codebook lookup. The lookup is similarly done as a weighted sum of the transition probabilities based on a_O of each codebook entry, and, as a result, preserves accuracy independent of the merging step.

$$a_O(i, j) = \frac{1}{2} \sum_{k=1}^{N_s} \left(\min\left\{ \sum_{m \in R_i'} a_R(m, i), \sum_{n \in R_j'} a_R(n, j) \middle| s_m = s_n = k \right\} + \right.$$

$$\left. \min\left\{ \sum_{m \in R_i'} a_E(m, i), \sum_{n \in R_j'} a_E(n, j) \middle| s_m = s_n = k \right\} \right) \quad (5)$$

Complexity-wise, applying 1D Viterbi to 3D grid-based HMM was shown to be O($WHDN_s^2$) [25], whereas our 3D rbHMM is processed in O($N_N N_B N_s^2$) time, where N_B is the number of tree branches and N_N is the average number of nodes per branch. If no branch overlaps exist, $N_N \times N_B = N_R$. The proposed algorithm is thus, generally, more computationally efficient, since, with proper watershed delineation, $N_N \times N_B \ll W \times H \times D$.

3. Results and Discussion

We applied the proposed 3D rbHMM to 3D image segmentation. We first tested on synthetic geometric shapes similar to [34] with ground truth available. In those images, the object is black (values of 0) on a white (values of 1) background (100×100×100 voxels). Gaussian noise was added with varying standard deviations σ. The shapes were generated using the following equations, where (x, y, z) are the coordinates along the X, Y and Z axes, and a, b and c are parameters governing the shape of each structure.

Torus: $\left(c - \sqrt{(x^2 + y^2)}\right)^2 + z^2 = a^2$ (6)

Ellipsoid: $\left(x^2/a^2\right) + \left(y^2/b^2\right) + \left(z^2/c^2\right) = 1$ (7)

Hyperboloid: $\left(x^2/a^2\right) + \left(y^2/b^2\right) - \left(z^2/c^2\right) = 1$ (8)

We also validated our method by applying the 3D rbHMM to six simulated BrainWeb [40] image volumes (0/1/3/5/7/9% noise, 181×217×181 resolution, 1×1×1mm^3 spacing) and eighteen real clinical Internet Brain Segmentation Repository (IBSR) T1-weighted MRI scans [41] (256×256×128 resolution, 0.837×0.837~1×1mm^2 in-plane spacing, 1.5mm slice thickness).

3.1. Parameter Sensitivity Assessment

We first examined the misclassification rates in the segmentation of 3D geometric shapes when varying the watershed threshold parameter, T_w, by ±13.3% around 0.75 (from 0.65 to 0.85). T_w denotes the ratio of significant edges present in the image, and as its value increases, the size of the regions decrease. By selecting T_w=0.75, we achieved 90% reduction in the number of image regions used, i.e. $N_R \leq 100,000$, as opposed to processing every image voxel ($W \times H \times D$=1,000,000).

Figure 2 shows the results after 3 iterations of the parameter estimation algorithm with σ=0.5 and σ=0.6. Results indicate our method is robust to changes in T_w. The worst case performance (ellipsoid, a=40, b=10, c=40, σ=0.6) was still significantly more stable and accurate than results reported in [34] (5.76%~22.30%).

(a)

3D Ellipsoid Parameters
(a, b and c denote the radius along each axis)

(b)
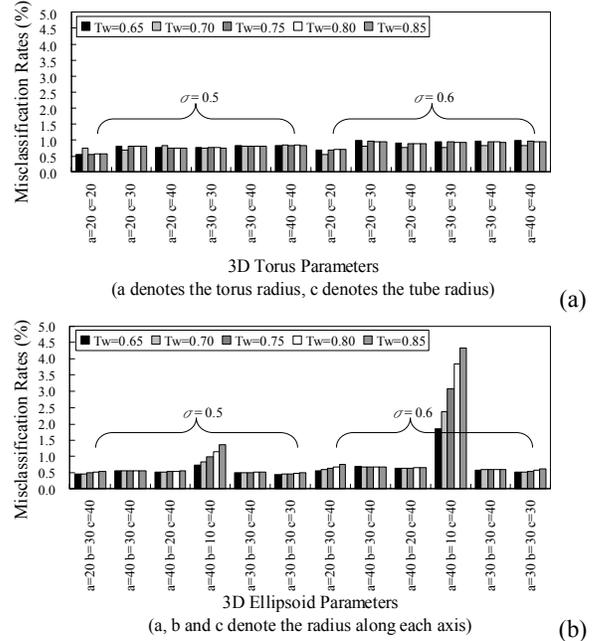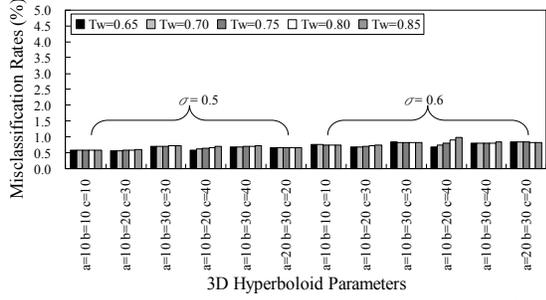
Figure 2: Misclassification error rate of segmentation results obtained by using the proposed 3D rbHMM method on 3D synthetic (a) torii,(b) ellipsoids and (c) hyperboloids. We note the stable performance when varying parameter T_w.

3.2. Segmentation Accuracy

We next compare the highest accuracies of the proposed 3D rbHMM model from §3.1 above to the highest accuracy results reported in [34] for the same shape parameter values tested at σ=0.5 and σ=0.6. Table 1 to Table 3 show that rbHMM outperformed grid-based HMM under various noise conditions in the vast majority of synthetic geometric shapes tested (32 out of 36). One result was equal to and three were only slightly lower than ($\leq$0.38%) those of [34]. Overall, we achieved 66.24% (σ=0.5) and 54.05% (σ=0.6) reductions in the misclassification rates compared to those reported in [34]. Figure 3 to Figure 5 further illustrate the segmentation performance qualitatively.

a	c	3D HMM [34]		3D rbHMM	
		σ=0.5	σ=0.6	σ=0.5	σ=0.6
20	20	0.64%	1.70%	**0.56%**	**0.56%**
20	30	1.04%	1.49%	**0.67%**	**0.81%**
20	40	1.15%	1.64%	**0.75%**	**0.76%**
30	30	3.14%	4.33%	**0.74%**	**0.77%**
30	40	2.33%	2.83%	**0.80%**	**0.82%**
40	40	0.91%	**0.45%**	**0.83%**	0.83%

Table 1. Comparison of segmentation accuracy: grid-based 3D HMM vs. proposed 3D rbHMM on 3D torii with shown parameters. a denotes the radius of the torus and c denotes the radius of the tube. The best performance is highlighted in bold.

a	b	c	3D HMM [34]		3D rbHMM	
			σ=0.5	σ=0.6	σ=0.5	σ=0.6
20	30	40	3.37%	4.61%	**0.45%**	**0.56%**
40	30	40	**0.56%**	0.89%	**0.56%**	**0.67%**
40	20	40	**0.43%**	0.64%	0.52%	**0.63%**
40	10	40	6.26%	5.76%	**0.73%**	**1.86%**
30	30	40	**0.48%**	0.86%	0.49%	**0.58%**
30	30	30	1.69%	4.41%	**0.44%**	**0.51%**

Table 2. Comparison of segmentation accuracy of grid-based 3D HMM vs. proposed 3D rbHMM on 3D ellipsoids with shown parameters. a, b and c denote radii along each axis. The best performance is highlighted in bold.

a	b	c	3D HMM [34]		3D rbHMM	
			σ=0.5	σ=0.6	σ=0.5	σ=0.6
10	10	10	0.80%	1.12%	**0.58%**	**0.74%**
10	20	30	0.90%	2.24%	**0.56%**	**0.67%**
10	30	30	0.74%	3.69%	**0.70%**	**0.81%**
10	20	40	1.43%	1.99%	**0.58%**	**0.68%**
10	30	40	0.97%	1.24%	**0.67%**	**0.79%**
20	30	20	0.74%	1.19%	**0.65%**	**0.82%**

Table 3. Comparison of segmentation accuracy of grid-based 3D HMM vs. proposed 3D rbHMM on 3D hyperboloids with shown parameters. a and b govern the skirt radii and c determines the sharpness. The best performance is highlighted in bold.
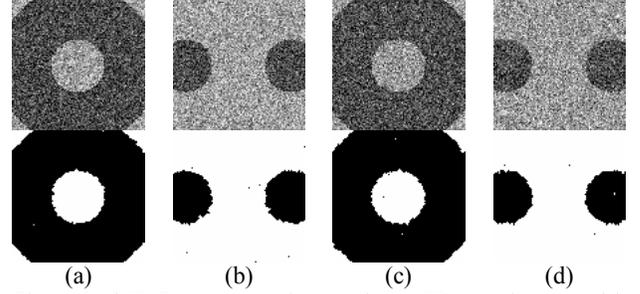

(a)   (b)   (c)   (d)

Figure 3: rbHMM segmentation results on 3D torus images with parameters a=20 and c=40. Top row: original noisy images for (a)-(b) σ=0.5, (c)-(d) σ=0.6; Bottom row: segmentation outputs of different slices.
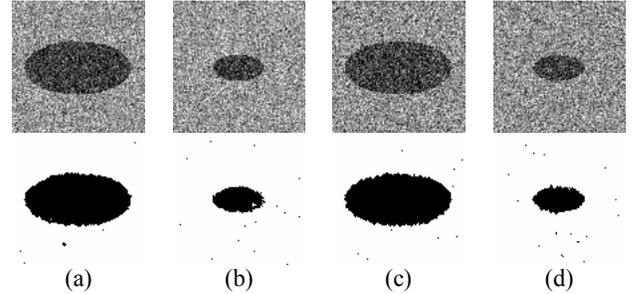

(a)   (b)   (c)   (d)

Figure 4: rbHMM segmentation results on 3D ellipsoid images with parameters a=40, b=20 and c=40. Top row: original noisy images for (a)-(b) σ=0.5, (c)-(d) σ=0.6; Bottom row: segmentation outputs of different slices.
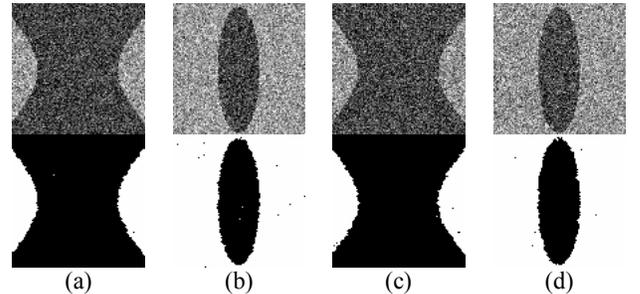

(a)   (b)   (c)   (d)

Figure 5: rbHMM segmentation results on 3D hyperboloid images with parameters a=10, b=30 and c=30. Top row: original noisy images for (a)-(b) σ=0.5, (c)-(d) σ=0.6; Bottom row: segmentation outputs of different slices.

3.3. Robustness to Object Rotation

To illustrate the robustness of rbHMM to object rotation, we demonstrated rotational invariance on a 3D ellipsoid (a=40, b=25, c=30) with noise (σ=0.5). The image was progressively rotated between 0° to 90° along each axis. Our proposed parameter estimation algorithm was then applied for 3 iterations at a fixed T_w=0.75.

Table 4 shows our method is robust to rotation angles. The misclassification error ranged from 0.47% to 0.55%, as compared to a range of 0.40% to 6.70% with the grid-based approach [34]. With only three iterations, we note a minor difference (0.01%) at 0° rotation between each run of the proposed method due to the random initialization nature of the parameter estimation algorithm. Compared to the results shown in [34], Figure 6 shows cleaner segmentation results based on the proposed method without significant observable artifacts such as horizontal streaks along the search direction.

Axis/Angle	0°	10°	20°	30°	40°
X	0.49%	0.50%	0.53%	0.53%	0.54%
Y	0.50%	0.51%	0.51%	0.52%	0.52%
Z	0.50%	0.51%	0.52%	0.53%	0.53%
Axis/Angle	50°	60°	70°	80°	90°
X	0.54%	0.54%	0.52%	0.52%	0.47%
Y	0.52%	0.55%	0.51%	0.51%	0.49%
Z	0.53%	0.52%	0.50%	0.50%	0.47%

Table 4. Effect of rotation on rbHMM segmentation results of 3D ellipsoids. We note the stable performance with varying angles of rotation with $\leq$ 0.07% changes in the misclassification rate.
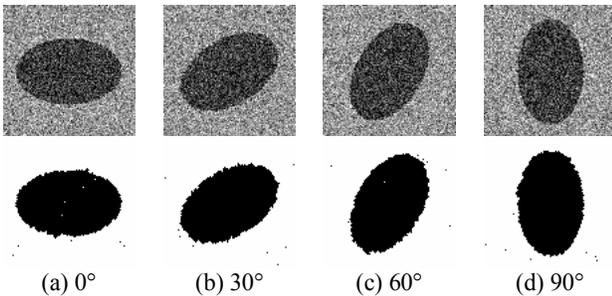

| (a) 0° | (b) 30° | (c) 60° | (d) 90° |

Figure 6: Effect of varying rotational angles (0°~90°) on rbHMM segmentation performance of 3D ellipsoids. Top row: original noisy images for σ=0.5; Bottom row: segmentation results.

3.4. Computational Efficiency

The average running time of our 3D rbHMM for the 18 tested 3D geometric shapes in §3.2 (σ=0.5, 100×100×100 voxels) with 3 iterations at T_w=0.75 was 170s on a 3.0 GHz Xeon processor. This corresponds to roughly a 30% speed increase when compared to the 280s reported based on the grid-based 3D HMM [34], adjusted for CPU speed.

3.5. Application to Noisy Simulated MRI Data

We performed 4-class; background, gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) segmentation on simulated BrainWeb MR images [40] with varying noise levels (0%, 1%, 3%, 5%, 7%, 9%). The proposed rbHMM was applied for 10 iterations at T_w=0.25 which corresponded to approximately 96.7%~98.7% reduction in the number of image regions used (N_R=89,725~234,501) as opposed to processing every image voxel ($W \times H \times D$= 7,109,137).

We compared our segmentation results against HMRF-EM [33] and block-based 3D HMM [26]. The Dice similarity index [42] was used to quantitatively evaluate WM and GM segmentation accuracies (by comparing with provided phantom). Figure 7 illustrates that both our WM and GM segmentation results were consistently and highly accurate and stable, when compared to the HMRF-EM and the block-based HMM models.
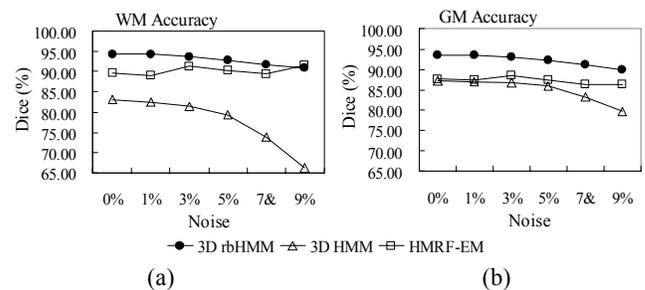

Figure 7: Comparisons of segmentation accuracy of proposed 3D rbHMM, block-based 3D HMM [26], and HMRF-EM [33] on simulated 3D brain MRI images for (a) WM and (b) GM. Our method shows consistently high accuracies and stable results when the noise level is varied from 0% to 9%.

Figure 8 demonstrates rotational invariance of our rbHMM on noisy 3D BrainWeb image (9%) rotated between 15° to 45° around the Z axis. Rotated phantoms are not available for quantitative comparison, and any rotation of the discrete label images would introduce resampling errors so was not pursued. The illustrated segmentation results are provided to qualitatively confirm that no observable artifacts or increases in misclassification resulted from object rotations.

3.6. Application to Real Clinical MRI Data

We also performed 4-class: background, GM, WM, and CSF segmentation on 18 clinical T1-weighted IBSR MR images [41] where 3D rbHMM was applied for 10 iterations at T_w=0.25. Evaluation of the segmentation accuracies was performed by comparing against provided expert-guided manual segmentation labels.
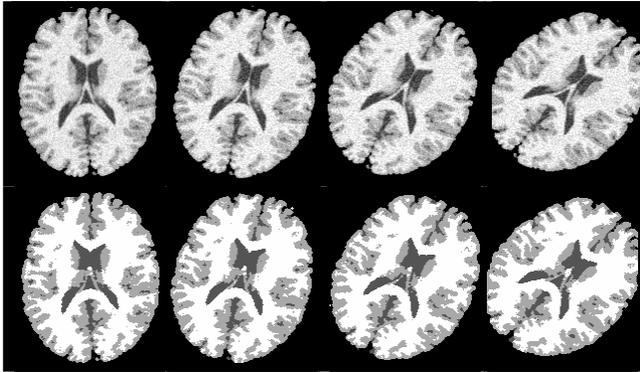
(a) 0°   (b) 15°   (c) 30°   (d) 45°

Figure 8: 3D simulated BrainWeb T1-weighted brain MR images showing effect of varying rotational angles on segmentation performance. Top row: original noisy images (9%); Bottom row: segmentation outputs for rotations of (a) 0°, (b) 15°, (c) 30° and (d) 45°.

Results achieved, on average, true positive (TP) and true negative (TN) ratios of [85.02%, 99.30%] for WM with standard deviations of [3.76%, 0.36%], as compared to [85.30%, 94.34%] with standard deviations of [5.45%, 2.58%] when using HMRF-EM [33]. To statistically evaluate the differences of segmentation results between rbHMM and HMRF-EM, we calculated the p values of TP and TN ratios ($p<0.05$ indicates significant difference). Our method achieved comparable TP ratio ($p=0.8601$) but with a significantly higher TN ratio ($p<0.0001$). Our method also achieved higher stability for both TP and TN ratios, as indicated by the reduced standard deviations.

Quantitative evaluation of GM accuracy is unreliable since the provided expert-guided manual label has been previously reported to misclassify cortical CSF as GM [43]. However, we qualitatively observe very good segmentation of the original images as shown in Figure 9.
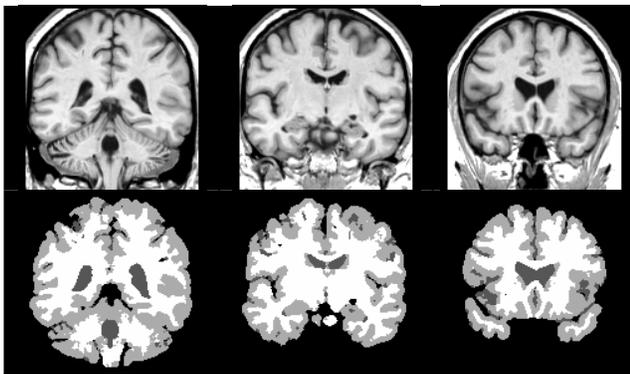
(a) slice 45   (b) slice 65   (c) slice 85

Figure 9: 3D clinical IBSR T1-weighted brain MR images (normal adult scan #06) showing segmentation accuracy of the proposed 3D rbHMM method. Top row: original images; Bottom row: segmentation outputs of slice number (a) 45, (b) 65 and (c) 85.

The average running time of the proposed method was 9.39 minutes a 3.0 GHz Xeon processor, comparable to that of the HMRF-EM approach.

4. Conclusions

In this paper, we proposed a novel 3D region-based hidden Markov model (rbHMM) employing an unsupervised, computationally efficient parameter estimation algorithm. The contributions of this work are: (1) a truly 3D HMM framework based on *irregularly shaped* homogeneous regions, which reduces the overall model complexity; thus, resulting in a more efficient optimization process than approaches based on a rectangular lattice or grid; (2) a novel 3D tree-structured algorithm to provide *rotationally invariant* estimates of the locally optimal MAP states of image regions by employing the Viterbi algorithm. We first demonstrated, using synthetic geometric shapes, improved computational efficiency and increased invariance to object rotation in 3D over a grid-based 3D HMM framework using a similar optimization strategy. We also successfully applied the proposed method to the segmentation of both simulated and clinical brain MRI data. On the simulated MRI dataset, the results showed that rbHMM has improved stability and accuracy over HMRF-EM and block-based HMM approaches. On the clinical MRI dataset, the results showed improved WM stability and accuracy over HMRF-EM. These studies on varying model, shape, noise, and rotation parameters demonstrated the ability of the proposed 3D rbHMM and the parameter estimation algorithm to robustly handle 3D image segmentation.

References

[1] J. K. Baker. The dragon system—An overview, Proc. Int. Conf. Acoust., Speech, Signal Process., 24–29, Feb. 1975.

[2] L. Rabiner, B. H. Juang. Fundamentals of Speech Recognition. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[3] M. Diligenti, P. Frasconi, and M. Gori. Hidden tree markov models for document image classification. IEEE Trans. Pattern Anal. Machine Intelligence, 25(4):519--523, 2003.

[4] J. Li, A. Najmi, R.M. Gray. Image classification by a two dimensional hidden Markov model, IEEE Trans. Signal Processing, 48(2):517-533, Feb. 2000.

[5] E. Levin, R. Pieraccini. Dynamic planar warping for optical character recognition, Proc. Int. Conf. Acoust., Speech Signal Process., San Francisco, CA, 3:149–152, Mar. 1992.

[6] H. K. Lee, J. H. Kim. An HMM-based Threshold Model Approach for Gesture Recognition, IEEE Trans. Pattern Anal. Machine Intelligence, 21(10):961-973, Oct. 1999.

[7] S. Eickeler, S. Muller, G. Rigoll. Improved face recognition using pseudo 2-D hidden Markov models, Proc. Workshop on Advances in Facial Image Analysis and Recognition Technology, Freiburg, Germany, June 1998.

[8] F. Hulsken, F. Wallhoff, G. Rigoll. Facial expression recognition with pseudo-3D hidden Markov models, Proc.

DAGM-Symposium, Lecture Notes in Computer Science, 2191:291-297, 2001.

[9] J. Boreczky, L. Wilcox. A hidden Markov model framework for video segmentation using audio and image features, Proc. IEEE Conf. on Acoustics, Speech, and Signal Processing, 6:3741-3744, 1998.

[10] T. Vlachos, A. G. Constantinides. A graph-theoretic approach to colour image segmentation and contour classification, International Conference on Image Processing and its Applications, Maastricht, Netherlands, 1992.

[11] K. Barnard, D. A. and Forsyth. Learning the Semantics of Words and Pictures. International Conference of Computer Vision, pp. 408-415, 2001.

[12] J. Jiten, B. Mérialdo. Semantic Image Segmentation with a Multidimensional Hidden Markov Model. MMM 1:616-624, 2007.

[13] A. J. Frost. Spatio-temporal hidden Markov models for incorporating interannual variability in rainfall. PhD Doctorate Thesis, the University of Newcastle, 2004.

[14] C.-C. Ke, J. G. Herrero, J. Llinas. Comparative analysis of alternative ground target tracking techniques. Proc. IEEE Information Fusion, 2000.

[15] R. Durbin, S. Eddy, A. Krogh, G. Mitchison. Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids. Cambridge University Press, Cambridge, 1998.

[16] L. E. Baum. An inequality and associated maximization technique in statistical estimation for probabilistic functions of finite state Markov chains, Inequalities III. New York: Academic, pp. 1–8, 1972.

[17] A. A. Markov. An example of statistical investigation in the text of 'Eugene Onyegin' illustrating coupling of 'tests' in chains, Proc. Acad. Sci., 7:153, 1913.

[18] C. E. Shannon. A mathematical theory of communication, Bell Syst. Tech. J., 27:379–423, July 1948.

[19] R. G. Gallager. Information Theory and Reliable Communication. New York, NY: Wiley, 1968.

[20] L. E. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains, Ann. Math. Stat., 41(1): 164–171, 1970.

[21] A. J. Viterbi, J. K. Omura. Trellis encoding of memoryless discrete-time sources with a fidelity criterion, IEEE Trans. Information Theory, IT-20:325–332, May 1974.

[22] G. D. Forney. The Viterbi algorithm, Proceedings of the IEEE 61(3):268–278, March 1973.

[23] S. S. Kuo, O. E. Agazzi. Machine vision for keyword spotting using pseudo 2D hidden Markov models, in Proc. Int. Conf. Acoust., Speech Signal Process., 5:81–84, 1993.

[24] P. A. Devijver. Probabilistic labeling in a hidden second order Markov mesh, Pattern Recognition in Practice II, pp. 113–123, 1985.

[25] J. Li, D. Joshi, J. Z. Wang. Stochastic Modeling of Volume Images with a 3-D Hidden Markov Model, ICIP, 2004.

[26] M. Ibrahim, N. John, M. Kabuka, A. Younis. Hidden Markov models-based 3D MRI brain segmentation, Image and Vision Computing, 24:1065-1079: 2006.

[27] S. Geman, D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, IEEE Trans. Pattern Anal. Machine Intelligence, 6:721–741, Nov. 1984.

[28] S. Z. Li. Markov random field models in computer vision. Proceedings of the European Conference on Computer Vision, Stockholm, Sweden, May 1994.

[29] I. Patras. E. A. Hendriks, R. L. Lagendijk. Video Segmentation by MAP Labeling of Watershed Segments, IEEE Trans. Pattern Analysis and Machine Intelligence, 23(3):326-332, March 2001.

[30] Y. Tsaig, A. Averbuch, A Region-based MRF Model for Unsupervised Segmentation of Moving Objects in Image Sequences, CVPR, 2001.

[31] A. Sarkar, M. K. Biswas, K. M. S. Sharma. A Simple Unsupervised MRF Model Based Image Segmentation Appraoch, IEEE Trans. Image Processing, 9(5):801-812, 2000.

[32] H. Deng, D. A. Clausi. Gaussian MRF Rotation-Invariant Features for Image Classification, IEEE Trans. Pattern Analysis and Machine Intelligence, 26(7):951-955, July 2004.

[33] Y. Zhang, M. Brady, S. Smith. Segmentation of Brain MR Images Through a Hidden Markov Random Field Model and the Expectation-Maximization Algorithm, IEEE Trans. Medical Imaging, 20(1):45-47, 2001.

[34] D. Joshi, J. Li, J. Z. Wang. A computationally efficient approach to the estimation of two- and three-dimensional hidden Markov models,' IEEE Trans. Image Processing, 15(7):1871-1886, 2006.

[35] S. Buecher. Watersheds of functions and picture segmentation, Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, Paris, France, pp. 1928–1931, May 1982.

[36] J. Shi, J. Malik. Normalized cuts and image segmentation, Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 731-737, 1997.

[37] F. Li, Q. Dai, W. Xu. Region-based hidden Markov models for image categorization and retrieval, Proc. SPIE Electronic Imaging, vol.6508, 2007.

[38] S. P. Lloyd. Least Square Quantization in PCM, IEEE Trans. Information Theory, 28(2):129-136, 1982.

[39] G. J. McLachlan, K. E. Basford. Mixture Models: Inference and Applications to Clustering, Marcel Dekker, New York, 1988.

[40] C. A. Cocosco, V. Kollokian, R. K.-S. Kwan, A. C. Evans. BrainWeb: Online Interface to a 3D MRI Simulated Brain Database, NeuroImage, 5(4 Part 2/4):S425, 1997.

[41] A. Worth, The Center for Morphometric Analysis at MGH, http://www.cma.mgh.harvard.edu/ibsr/, November 2007.

[42] L. R. Dice. Measures of the amount of ecologic association between species, Ecology, 26:297-302, 1945.

[43] E. Angelini, T. Song, A. Laine. Homogeneity measures for multiphase level set segmentation of brain MRI, Biomedical Imaging: Macro to Nano, 2006. 3rd IEEE International Symposium on, pp. 746-749, 2006.