An Automated Volumetric Segmentation System
Combining Multiscale and Statistical Reasoning

David W. G. Montgomery and Abbes Amira
The Institute of Electronics, Communications
and Information Technology (ECIT)
Queen’s University Belfast, Belfast, U.K. BT3 9DT
Email: d.montgomery@qub.ac.uk

Fionn Murtagh
Department of Computer Science
Royal Holloway, University of London
Egham, Surrey TW20 0EX, England

Abstract—A fully automated volumetric image segmentation algorithm is proposed. This method is fast and unsupervised, automatically estimating required parameters including optimal segment number selection using Bayesian inference. Primarily Gaussian Mixture Modeling (GMM) is used to achieve a baseline scene estimate. This estimate is then refined to consider spatial correlations using a Markov Random Field Model (MRFM). The application of this system to three-dimensional biomedical image volumes is discussed. This approach delivers promising results in terms of the identification of inherent image features.

I. INTRODUCTION

Image segmentation is the process of classifying each pixel of an image into a set of distinct classes, where the number of classes is much smaller than the number of unique pixel values. The goal of image segmentation is to separate features from each other, where features are items of interest in an image. Applications in a medical context include the separation of different tissue types in a brain image: white matter, grey matter, blood, cerebrospinal fluid etc., the identification of cancerous cells within volumetric data, and also Region Of Interest (ROI) compression, where major image features may be compressed at a lower rate than features which may not be important in the diagnosis of a patient i.e. the image background. Segmentation can be performed manually by a human expert who simply examines an image, determines borders between regions, and classifies each region. This is perhaps the most reliable and accurate method of image segmentation because the human visual system is immensely complex and well suited to the task. This task does however become extremely difficult when the image in question is volumetric [1].

Existing work completed by Collet and Murtagh [2] [3] focused on the segmentation of astronomical multiband imagery using a similar statistical model, however, resolution analysis is supplied via a quadtree structure where decimation is required. Although this leads to a fast segmentation, decimation can lead to the creation of visual artefacts within resulting segmentation maps. Other work in this area has been performed by Stanford and Raftery [1] [4]. Their work, applied only to two dimensional imagery in a spatial context used an analogous model, however multiresolution approaches yield image resolution scale components, which can be useful for noise removal. This type of analysis allows us to evaluate the spatial relationship between objects in an image at multiple scales, and therefore exploit the image characteristics, matched to a particular scale, which might go undetected in other analysis techniques [5].

In our proposed work an integrated automated volumetric segmentation system using a redundant wavelet transform and MRFM is described. This system, using Bayesian inference for model selection can be used to segment three-dimensional medical image volumes into its inherent features.

The novelties presented in this work are as follows;

1) A model is defined which considers neighbouring voxels in three dimensional space. Previous statistical models have processed image volumes slice-by-slice using two dimensional techniques to limit computational burden [6].

2) The system described uses a redundant wavelet transform for multiscale analysis. Other authors have chosen to use popular wavelet filters which require decimation, or a quadtree structure [2] [3]. The usage of each decomposition level is novel as only some can be deemed useful for the identification of principle image features e.g. low frequency decompositions may cause resultant segmentations to be excessively smooth.

3) The statistical model used is initialised using histogram equalisation. In previous publications either random parameters have been computed to establish initial histogram breakpoints or Ward’s methodology has been used. These methods vastly influence overall segmentation quality [1] [4].

The remainder of this paper is structured as follows: Mathematical theory encompassing both multiresolution and statistical methodology is presented in section 2. In section 3 the structure of the proposed system is highlighted. Results to-date are depicted in section 4 and section 5 contains a discussion of results obtained and closing comments.

II. MATHEMATICAL THEORY

A. The À Trous Redundant Wavelet Transform

A wavelet transform performs a series of transformations on a signal, providing a resolution related set of views of the signal.
The wavelet transform of an image produces, at each scale \( j \), a set of zero mean coefficient values \( w_j \). In the à trous method, this set has the same number of pixels as the image and thus this wavelet is a redundant one. Furthermore, using a wavelet defined as the difference between the scaling functions of two successive scales \((\frac{1}{2}\psi(\frac{x}{2}) - \phi(x))\), the original image \( c_0 \), with a pixel at position \( x,y \), can be expressed as the sum of all the wavelet scales and the smoothed array \( c_p \)

\[
c_0(x, y) = c_p(x, y) + \sum_{j=1}^{p} w_j(x, y) \tag{1}
\]

To simplify notation, let us take one index running over all pixels,

\[
c_0(k) = c_p(k) + \sum_{j=1}^{p} w_j(k) \tag{2}
\]

A summary of the à trous wavelet transform algorithm is as follows;

1) Initialize \( j \) to 0 and start with the data \( c_j(k) \).
2) Increment \( j \), and carry out a discrete convolution of the data \( c_{j-1}(k) \) using the filter \( h \). The distance between the central sample and the adjacent ones is \( 2^{j-1} \).
3) After this smoothing, obtain the discrete wavelet transform from the difference \( c_{j-1}(k) - c_{j}(k) \).
4) If \( j \) is less than the number \( p \) of resolutions required, return to step 2.
5) The set \( W = \{ w_1, ..., w_p, c_p \} \) represents the wavelet transform of the data [7].

B. Gaussian Mixture Modeling (GMM)

The mixture density provides a natural way of modeling the observed mixture of features in an image. The mixture density, shown in equation 3 can be used to model the marginal (without spatial information) distribution of grayscale values in an image. In this case we use the gaussian density, as the distribution of grayscale values shown in equation 3 can be used to model the marginal observed mixture of features in an image. The mixture density, is the conditional probability \( \gamma_i | \theta, \lambda \) given the current estimates of \( \theta \) and \( \lambda \) are computed [9].

\[
f(x_i | \theta, \lambda) = \sum_{k=1}^{K} \lambda_k f_k(x_i | \theta_k) \tag{3}
\]

where model parameters \( \theta_k = \{ \mu_k, V_k \} \); \( f_k(\cdot | \theta_k) \) is a gaussian density with mean \( \mu_k \) and variance-covariance matrix \( V_k \); \( \theta = (\theta_1, ..., \theta_K) \); and \( \lambda = (\lambda_1, ..., \lambda_K) \) is a vector of mixture probabilities such that \( \lambda_K \geq 0(k = 1, ..., K) \) and \( \sum_{k=1}^{K} \lambda_k = 1 \) [8].

The Expectation Maximisation (EM) algorithm iterates between the E step and the M step. In the E step, the conditional expectation \( \hat{\gamma}_i \) of \( \gamma \) given the data and the current estimates of \( \theta \) and \( \lambda \) is computed, so that \( \gamma_i | \hat{\theta}, \hat{\lambda} \) is the conditional probability that \( x_i \) belongs to the \( k^{th} \) group. In the M step, conditional maximum likelihood estimates of \( \theta \) and \( \lambda \) given the current \( \hat{\gamma} \) are computed [9].

C. The Markov Random Field Model (MRFM)

In the volumetric image case, consider an unknown, true voxel state, for voxel \( i \), as \( x_i \in \{1, 2, ..., K \} \) for \( K \) states. In this study this is taken as an \( n \) valued vector, which represents each of the selected à trous decomposition levels.

We use Markovian assumptions to introduce spatial dependencies within the data. In this case the model is hidden because the variables \( X \) are only known through the observed \( Y \).

Consider an indicator function, \( I(x_i, x_j) = 1 \) if \( x_i = x_j \) and otherwise = 0. We take \( p(x) \) as being proportional to \( \exp(\phi U(N(x_i), j)) \). This is known as a Potts model. \( \phi \) is a spatial homogeneity parameter, a small value implying randomness, and a large value implying uniformity. A negative value of \( \phi \) implies dissimilarity between neighboring voxels, and is not of interest.

Let \( N(x_i) \) be the neighborhood of \( x_i \), taken in this case as adjacent 3 x 3 x 3 (27) voxels and let \( U(N(x_i), k) \) be the number of neighborhood voxels with state \( k \).

From \( p(x) \) we have the conditional distribution,

\[
p(x_i = j | N(x_i), \phi) = \frac{\exp(\phi U(N(x_i), j))}{\sum_k \exp(\phi U(N(x_i), j))} \tag{4}
\]

Further background on the approach adopted in this study can be found in [1] [4].

D. Bayes Information Criterion (BIC)

A model \( M_k \) is the set of parameters estimated for a given number of segments, \( K \). Consider data \( D \), where the posterior probability of model \( M_k \) is,

\[
p(M_k | D) = \frac{p(D | M_k)p(M_k)}{\sum_{L=1}^{K_{\text{max}}} p(D | M_L)p(M_L)} \tag{5}
\]

\( P(M_K) \) and the influence of \( M_L \) if each model is equi-likely \textit{a priori} can be ignored.

The ratio of posteriors, \( p(D | M_K) / p(D | M_{K'}) \) is referred to as a Bayes factor for model \( M_K \) against model \( M_{K'} \).

The integrated likelihood, \( p(D | M_K) \), is given by,

\[
p(D | M_K) = \int p(D | \theta_K, M_K)p(\theta_K)d\theta_K \tag{6}
\]

where \( \theta_K \) is the set of parameters for the model \( M_K \), \( p(D | \theta_K, M_K) \) is the usual likelihood, and \( p(\theta_K) \) is the prior. Evaluating this integral is combinatorially difficult, so an approximation which is often used is as follows,

\[
2 \log p(D | M_K) \approx BIC \tag{7}
\]

where,

\[
BIC = 2 \log p(D | \hat{\theta}_K, M_K) - N \log (\text{dim}(\theta_K)) \tag{8}
\]

where \( \hat{\theta}_K \) is the maximum likelihood estimator of \( \theta_K \), \( \text{dim}(\theta_K) \) is the number of parameters estimated, and \( N \) is the cardinality of the data [4] [8].
III. PROPOSED SEGMENTATION ALGORITHM

A. System Overview

1) Initially the à trous wavelet transform is applied to the biomedical image volume, using the B-spline of degree three as the scaling function and using a four scale decomposition. This implementation is based on three one-dimensional sets of separable convolutions. From this procedure a four-dimensional result consisting of four three-dimensional volumes with dimensions equating to the original image are achieved. In this work only the second and third image scales are used for the segmentation. We assume that the first scale contains high frequency information, which in many cases constitutes noise, and that the fourth image scale contains averaging coefficients which may lead to an overly smooth segmentation allowing no identification of finer image features (Figure 1(a)).

2) Bayesian inference is used to assess output segmentations to decide on the optimal number of segments to be retained. In this case all parameters resulting from the MRFM are not considered. Instead a simpler, more focused and easily analyzed model is used based on the segmentation maps obtained after the GMM. For this no spatial characteristics are considered, and this is applied only to the third decomposition or scale obtained from the wavelet transform. Mixture modeling is applied iteratively to the third scale using a range of possible segment numbers. The BIC value is calculated for each iteration and a relative maximum value is used as an indication of the optimal number of segments required (Figure 1(b)).

3) On the completion of class number selection, a GMM is reapplied to the third wavelet scale for the optimal number of segments. Although this only considers one of the wavelet scales, this provides a useful baseline scene estimate. A problem associated with mixture modeling is the way in which it is initialised. This can have a major influence on the quality of the overall segmentation result. This initialisation is used as a starting point for the EM algorithm to fit the GMM and find a marginal segmentation through maximum likelihood classification. Histogram equalisation [10] was selected as it provided the basis for a fast and robust method of finding an initial segmentation. This approach divides the grayscale levels within the histogram into \( K \) bins containing roughly equal numbers of voxels. Even when the data is extremely skewed or concentrated on just a couple of grey values, this algorithm is guaranteed to converge as long as \( K \) is less than or equal to the number of distinct grey values in the data.

4) The baseline scene estimate achieved is refined using wavelet scales two and three and a MRFM. This models spatial interaction between neighbouring or nearby voxels. These local correlations provide a mechanism for modeling a variety of image properties. In medical imaging, they are typically used to take into account the fact that most voxels belong to the same class as their neighbouring voxels. In physical terms, this implies that any anatomical structure that consists of only one pixel has a very low probability of occurring under an MRF assumption. The segmentation in this case is performed by maximising the \textit{a posteriori} probability of the segmentation using the Iterated Conditional Modes (ICM) algorithm. This was introduced by Besag [11] and can be used as a method of image reconstruction when local characteristics of the true image can be modeled as a MRF. The algorithm begins with an initial estimate of the true scene \( X \) found primarily through the simple GMM described previously. The algorithm proceeds iteratively to provide an estimate of the required parameters (Figure 1(c)).

![Fig. 1. Volumetric segmentation system.](image)

(a)

(b)

(c)

Original biomedical image volume.

Volume in wavelet domain

(c)

(b)

Fig. 1. Volumetric segmentation system. 1(a) Wavelet transform. 1(b) Iterative application of GMM to find optimal segment selection using BIC. 1(c) Take the GMM as scene estimate along with scales 2 and 3 from the transform and refine using MRFM.

IV. RESULTS

In this study the proposed segmentation system was applied to MRI and PET image volumes.

Test 1 involved the segmentation of an aggregated representative brain, derived from MRI data, with dimensions 91 x 100 x 91. See figure 3(a-c). The Bayes factor approach to model selection provides an objective measure of a number of classes versus another. Section 2D described how the BIC could be computed, and calculated BIC values for the MRI brain volume are displayed in figure 2. To determine the optimal number of classes, we determine BIC for various numbers of classes, \( K = 4, 5, ..., 16 \). We find that \( K = 7 \) provides the optimal segmentation. The segmentation map obtained with \( K = 7 \) classes is shown in figure 3(d-f). The selection of 7 segments corresponds with
Fig. 2. BIC values for varying number of segments, for Brain MRI volume. The value corresponding to 7 classes was used.

Fig. 3. (a-c) Original MRI brain volume, slices 15, 30 and 45. (d-f) Segmented volume into 7 classes using multiresolution MRFM, slices 15, 30 and 45.

Figures 4 and 5 display the results obtained from a PET brain image volume with dimensions 91 x 100 x 60. In this case the BIC criterion determines that \( K = 7 \) provides the most suitable segmentation.

Fig. 4. BIC values for varying number of segments, for Brain PET volume. The value corresponding to 7 classes was used.

Fig. 5. (a-c) Original PET brain volume, slices 15, 30 and 45. (d-f) Segmented volume into 7 classes using multiresolution MRFM, slices 15, 30 and 45.

V. CONCLUSION

The approach addressed in this paper is new, and results are promising. This work describes how the use of a Bayes factor provides a way to objectively decide on the number of appropriate segmentation levels. With the results obtained to date, principle brain structures are clearly apparent, however some aspects must be improved. Some structures are obvious along image boundaries. These could be due to the multidimensionality of the image and the use of only two scale levels for segmentation. Future work will involve the formalisation of a framework for distinguishing the most suitable resolution levels to aid segmentation.

REFERENCES