UNSUPERVISED IMAGE SEGMENTATION VIA MARKOV TREES AND COMPLEX WAVELETS

Cián W. Shaffrey, Nick G. Kingsbury

Signal Processing Laboratory,
Department of Engineering,
University of Cambridge, UK.
{cws23,ngk}@eng.cam.ac.uk

Ian H. Jermyn

Project Ariana (CNRS/INRIA/UNSA),
INRIA Sophia Antipolis, France.
ian.jermyn@sophia.inria.fr

ABSTRACT

The goal in image segmentation is to label pixels in an image based on the properties of each pixel and its surrounding region. Recently Content-Based Image Retrieval (CBIR) has emerged as an application area in which retrieval is attempted by trying to gain unsupervised access to the image semantics directly rather than via manual annotation. To this end, we present an unsupervised segmentation technique in which colour and texture models are learned from the image prior to segmentation, and whose output (including the models) may subsequently be used as a content descriptor in a CBIR system. These models are obtained in a multiresolution setting in which Hidden Markov Trees (HMT) are used to model the key statistical properties exhibited by complex wavelet and scaling function coefficients. The unsupervised Mean Shift Iteration (MSI) procedure is used to determine a number of image regions which are then used to train the models for each segmentation class.

1. INTRODUCTION

In recent years, Content-Based Image Retrieval (CBIR) has emerged as an active research area. This is partly due to the explosion in the amount of image data being generated in many different fields and by the growth of the Web. Successful image retrieval is the key to making this large amount of data useful. However, CBIR also provides a framework within which many important questions of machine vision are brought into focus: successful retrieval is likely to require access to the semantic level. Initial approaches to image retrieval focussed on global properties of the images, but it is apparent that information about the spatial variation of image properties is critical to retrieval. Image segmentation has thus come to play a central role in current approaches to CBIR. The hope is that relatively low-level properties of image regions will be sufficient to characterize semantic value, and hence that the use of segmentations as summaries of image content will capture semantic distinctions between images.

In return, the nature of the CBIR application imposes constraints on the segmentation algorithms to be used. One of the purposes of CBIR is to avoid the time-consuming and expensive business of hand-labelling of images by human operators. This means that the algorithms should be unsupervised: no human interference should be required in adding a new image to the database. Secondly, although the data in different CBIR applications varies, very often the images in the database are highly diverse in their semantics. This means that when a new image is added to the database, there is a significant possibility that it possesses semantics that are new to the database. Consequently, the segmentation algorithm used to generate database indices for retrieval should operate on the data in the current image only, without recourse to the rest of the database.

In this paper, we describe probabilistic models of image texture and colour and use them for unsupervised image segmentation. These models are based on (complex) wavelet decompositions of the image, whose value for image processing and analysis is by now well-established. Modelling of the full joint probability distribution of the wavelet coefficients is impractical, and yet ignoring their mutual dependence completely is unrealistic. Hidden Markov tree (HMT) models provide a compromise between these extremes. They capture the key dependencies that exist between wavelet coefficients at different scales, while remaining computationally tractable. They have proven very successful in image restoration and denoising, and are finding increasing use in segmentation applications also [1, 2]. In [2], HMT models for the coefficients in a wavelet decomposition were used to describe image texture for segmentation. These models use supervised algorithms, and model texture only, rendering them less than ideal in a CBIR context. This paper proposes new models that overcome these limitations and demonstrates their value on real images. The key features of the models we describe are the following:

- The algorithms are unsupervised and operate using the current image only. A Mean Shift Iteration (MSI) procedure partitions the pixels into texturally, chromatically and spatially coherent clusters, which are then used to train the models for colour and texture.
- Texture is described using HMT models for the magnitude of the coefficients of the dual-tree complex wavelet transform (DT-CWT) [3, 4] of the intensity (L) component of the $L^*a^*b^*$ colour space representation of the image. The DT-CWT improves on the standard discrete wavelet transform (DWT) by having much improved translation invariance and better directional sensitivity, both important properties for texture characterization, while maintaining low computational cost. The conditional distributions of wavelet coefficient magnitudes are modelled as Rayleigh distributions, consistent with the real and imaginary parts of each coefficient being modelled by i.i.d. zero-mean Gaussians.
- Colour is described using independent Gaussian models for

This work was supported by EU project MOUMIR (HP108), www.moumir.org

In *Proc. ICIP 2002, Rochester NY, USA, September 22 - 25, 2002.*

© IEEE 2002

the DT-CWT *scaling function* coefficients of the three components of the $L^*a^*b^*$ colour space representation of the image.

In section 2 we give an overview of the entire segmentation process. A description of MSI is provided in §3. The DT-CWT and HMT based texture modelling are discussed in §4.1 and the colour modelling process in §4.2. Typical results of the scheme are presented in §6, with conclusions in §7.

2. OVERVIEW

We first outline the steps of the algorithm, and then describe each step in more detail in subsequent sections.

Mean Shift Iteration (MSI) MSI partitions the pixels into clusters that are approximately coherent, texturally, chromatically and spatially. These clusters will be used to train the colour and texture models. Note that MSI thus determines the number of distinct classes in the image.

Training of HMT texture model The data in each cluster generated by the MSI procedure is used to train the parameters of six 2-state HMT models of the magnitudes of the DT-CWT coefficients in each of the 6 directional subbands.

Training of colour model The data in each cluster generated by the MSI procedure is used to train the parameters of Gaussian models of the scaling function coefficients at each level.

Segmentation The texture and colour models are combined at each level of the decomposition to produce a likelihood that each macro-pixel belongs to a certain class. Initial segmentation occurs by assigning the maximum likelihood class to each macro-pixel. Further refinement of this segmentation is achieved by exploiting inter-scale dependencies using Data Fusion.

3. MEAN SHIFT ITERATION

The MSI procedure provides a general robust clustering method. It groups the image pixels into clusters that tend to be texturally, chromatically and spatially coherent and which can then be used to train the more sophisticated colour and texture models.

MSI is a generalised version of the k-means clustering algorithm and works as follows. In the data space used, a density gradient estimate is obtained using a differentiable kernel, which leads to the calculation of the mean shift vector. Successive evaluations of this vector result in stationary points, which are taken to be the cluster centres. The points located in the vicinity of each cluster centre are assigned to a corresponding class, while the rest of the points are labelled ‘undecided’ and are not used in training. We use a 7-dimensional data space: 3 colour dimensions, 2 texture dimensions and 2 spatial dimensions. The colour components are obtained from the pixel values in the $L^*a^*b^*$ colour space, while the texture components are the two dominant complex wavelet orientations, found using PCA. The spatial components are simply the row and column coordinates of each pixel. Further detail about MSI can be found in [5].

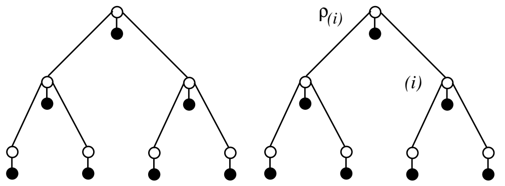


Fig. 1. A 1-D dependence structure where the dependency is associated with the hidden state variable (white node) rather than the actual observed coefficient (black node).

Upon convergence of the MSI procedure, the pixels have been partitioned into a number of clusters. Because of the inclusion of position information in the feature space, these clusters tend to be compact spatially, and define regions in the image with coherent properties. After the use of a smoothing procedure to eliminate small ‘holes’ in the regions, binary masks are constructed for each cluster. These masks are then used in the training phase of the algorithm to ensure that each model is trained on data coming from its respective cluster only. Figure 3 (b) illustrates the masks obtained for figure 3 (a), where each of the four classes is represented by a distinct colour (excluding white).

4. MODELS

4.1. DT-CWT HMT Models for Texture

The complex wavelet coefficients of real world images possess the following important statistical properties:

Non-Gaussian Distribution The frequency distribution of the coefficients tends to be peaky and heavy tailed.

Persistence The size of the coefficients tends to propagate through dependent branches of the wavelet tree, so that, for example, if a coefficient is large in value then it is likely that its children will also be large.

These properties are also reflected in the coefficient magnitudes, ω_i. Both of these properties can be captured by using HMT models of the distribution of the wavelet coefficient magnitudes. The first property is obtained by modelling the marginal distribution of each wavelet coefficient magnitude as a 2-component mixture. The mixture model is realised in its turn by endowing each coefficient with a state taking one of two values, L (large) or S (small), and treating the mixture as the joint distribution for the coefficient magnitude and its state marginalised over the state. The second property is obtained by modelling dependencies between the hidden states at different scales. In graphical terms, the distribution is illustrated in figure 1. The black nodes represent wavelet coefficient magnitudes, while the white nodes represent the hidden states. Edges in the graph represent dependencies. Note that we assume that each coefficient’s magnitude is independent of all other variables given its state, while each state depends only on the states of its parent and its children. We first describe the dependency structure of the hidden states, and then the conditional distributions of the coefficient magnitudes given their states.

4.1.1. Dependency structure of the hidden states

We can describe the distribution simply by giving the conditional probability of the state of a child given the state of its parent. There will be one of these matrices for each edge in figure 1. In addition, we must give the distribution of the root node of the tree. The parameters for the distribution of states are thus the following:

- $p_0(m)$, the probability that the state of the root node $S_0 = m$.
- $\varepsilon^{m,n}_{i,\rho(i)} = P(S_i = m | S_{\rho(i)} = n)$, the conditional probability that the state of node i is m, given that the state of its

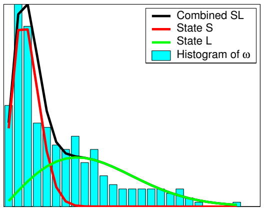


Fig. 2. A histogram of wavelet coefficient magnitudes is displayed in the light blue with the 2 components of the Rayleigh mixture displayed separately in red and green. The full mixture distribution is shown in black.

parent, $\rho(i)$, is n. This is a matrix for each i:

$$\varepsilon_{i,\rho(i)} = \begin{bmatrix} \varepsilon^{SS} & \varepsilon^{LS} \\ \varepsilon^{SL} & \varepsilon^{LL} \end{bmatrix}_{i,\rho(i)} \tag{1}$$

where the persistence property leads us to expect that $\varepsilon^{SS} \gg \varepsilon^{LS}$ and $\varepsilon^{LL} \gg \varepsilon^{SL}$.

In order to enforce translation invariance, the parameters at different positions within one subband are assumed to be the same.

4.1.2. *Conditional distribution of coefficient magnitudes*

Having described the dependency structure of the hidden states, we can now be more specific about the distributions of the coefficient magnitudes given their states. Since these magnitudes are independent given the states, it suffices to describe their individual distributions. These will be modelled using a Rayleigh distribution,

$$P(\omega_i) = \frac{2\omega_i}{\sigma^2_{i,S_i}} \exp\left(\frac{-\omega_i^2}{\sigma^2_{i,S_i}}\right), \tag{2}$$

where the value of σ_{i,S_i} depends on the state S_i of coefficient i and on the particular coefficient i considered, although again $\sigma_i = \sigma_j$ if i and j lie in the same subband.

The use of a Rayleigh distribution is equivalent to assuming that the real and imaginary parts of each coefficient are distributed as i.i.d. zero-mean Gaussians. They are zero mean because each wavelet filter response integrates to zero, and the Gaussian assumption is reasonable in the context of a HMT mixture model. Complex wavelet coefficients can be modelled by pairs of independent Gaussians, since the real and imaginary wavelet bases are approximately in quadrature (i.e. they form an approximate Hilbert pair) and hence are approximately orthogonal to each other. The Rayleigh model gives a simpler model than the pairs of Gaussians with a shared state variable used by Romberg *et al.* [6], and is quite straightforward to include in the HMT probability expressions.

The marginalised distribution of each coefficient magnitude is now given by a 2-component mixture of Rayleigh distributions. Figure 2 shows a typical histogram of wavelet coefficient magnitudes and shows how such a 2-component mixture model captures the distribution.

4.1.3. *Why the DT-CWT?*

The Dual-Tree Complex Wavelet Transform (DT-CWT) [3, 4] employs a dual-tree of real-valued wavelet filters to generate the real and imaginary parts of complex wavelet coefficients. In 2-D, the transform produces 6 subbands which are strongly oriented at angles of $\{\pm 15^\circ, \pm 45^\circ, \pm 75^\circ\}$.

Any method used to characterize texture should be as close to translation-invariant as possible, since the texture may present itself in the image under any translation. This poses a problem for many wavelet-based approaches, since it is well-known that the discrete wavelet transform is not translation-invariant. Thus a small shift in the input texture can cause a quite radical change in the structure of the wavelet coefficients, rendering recognition of the texture quite difficult. In contrast to the DWT, the DT-CWT coefficient magnitudes are approximately invariant to translation, which is one reason we choose to use it.

The second reason is that the DT-CWT possesses much better directional resolution than a standard DWT. It is intuitively clear that an important property of many textures is their directionality. The DT-CWT is able to detect this, as it possesses six directional subbands rather than three. This is because the complex filters are able to separate positive and negative frequencies in 1-D, and hence separate adjacent quadrants in 2-D frequency space.

4.1.4. *Training*

The DT-CWT is computed on the L channel of the $L^*a^*b^*$ representation of the image. Each of the 6 directional subbands of the DT-CWT is modelled over 4 levels using an HMT as described above. These trees are assumed independent. The parameters of the models must be trained: they are taken to be the maximum likelihood estimates. These can be computed using the EM algorithm, where the E step uses a forward-backward algorithm. The algorithm is described in full in [1].

4.2. Colour Modelling

Colour information is very important for producing meaningful results in any general image segmentation technique. Rather than use the wavelet coefficients to model the colour information in the image, it seems more sensible to use the scaling function coefficients: the averaged colour of a region is likely to be more informative than its variation.

The scaling function coefficients are assumed to be mutually independent. They are vector-valued, with three dimensions, one for each of the components in the $L^*a^*b^*$ colour space. Each coefficient is modelled using a Gaussian distribution with a diagonal covariance. The (vector-valued) mean and covariance of the Gaussians are taken to be constant within each level for a given class.

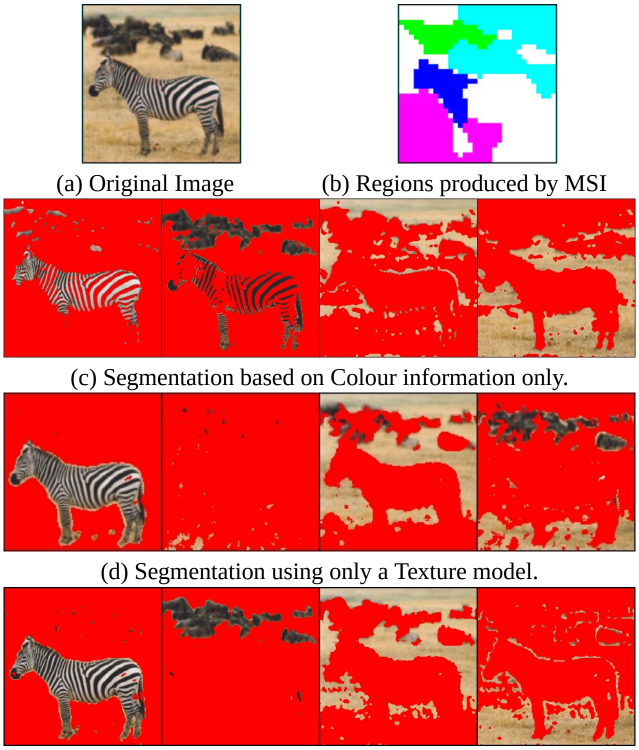
Training in this case is simple, since there are no mixture models and the maximum likelihood estimates are found analytically to be given by the mean and covariance of the data.

5. CLASSIFICATION

Given an image, we compute its 4-level DT-CWT, apply MSI and then for each resulting class, train six HMT models, one for each directional subband, and one Gaussian model for each scaling function coefficient subband. The product of probabilities from these models provides the probability distribution of an image of that class. It also allows us to compute, for any node in the HMT tree of the image, the likelihood of the wavelet and scaling function coefficients at and below that node given that they belong to a particular class. We can thus compute the maximum likelihood class for every node in the tree. The resulting initial estimates are then refined using a cross-scale data fusion technique described in detail in [2] and a 5×5 mode filter eliminates very small regions at the finest scale.

6. RESULTS

We present results using images from two separate data-bases: the Bridgeman Art Library (BAL) collection containing images of museum paintings and the Corel Image Database depicting natural scenes. Due to the lack of simple ground truth data for these databases, it is not easy to evaluate the scheme in an ‘objective’ manner. An attempt to overcome this problem is presented in [7].


(a) Original Image

(b) Regions produced by MSI

(c) Segmentation based on Colour information only.

(d) Segmentation using only a Texture model.

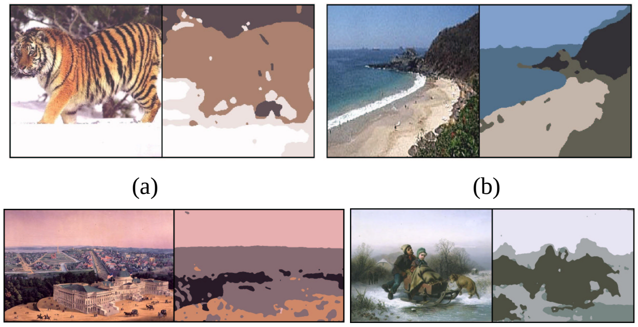
(e) Overall joint colour-texture segmentation.

Fig. 3. The image in (a) is segmented based solely on: colour in (c); texture in (d); and jointly on colour and texture in (e).

The ‘zebra’ image, figure 3 (a), was deemed to contain four classes by the MSI procedure. Each class had an associated training mask, depicted in figure 3(b) by highlighting each class with a separate colour. These masks were then used to isolate relevant regions of the original image when training the texture and colour models for each class. Figure 3 (c) contains the segmentation of the original image using only the colour model. Each of the four classes are displayed in a separate image, from left to right. A pixel which is *not* classified as part of a particular class is coloured in red, so as to contrast sharply with the pixels which are deemed most likely to belong to a particular class.

The colour model incorrectly separates the zebra into two different classes. This contrasts with the results obtained using only the texture model, figure 3 (d), where the textured nature of the zebra is captured correctly but the 2^{nd} and 4^{th} models fail to capture any significant semantic content. However, in figure 3 (e), we see an improved performance when segmenting based upon joint colour-texture information. Note that the foreground grass is in focus and has been ‘correctly’ segmented from the background grass which is blurred.

To provide further examples of the capabilities of our scheme, figures 4 (a), (b), (c) and (d) contain original images and their respective segmentation maps. In general the colour and texture models contribute equally towards the final result. But in a significant proportion of BAL images we noted that the colour information tended to dominate.


(a)

(b)

(c)

(d)

Fig. 4. Typical segmentation results: 2 from the NAT database (a) & (b); and 2 from the BAL collection (c) & (d) (© Bridgeman Art Library).

7. CONCLUSIONS

A segmentation scheme must be unsupervised to render it useful in the context of CBIR. We have presented a technique that uses HMT models of complex wavelet coefficient magnitudes to describe texture, and Gaussian models of scaling function coefficients to describe colour. Due to the MSI procedure, the algorithm operates without human input, and combines both colour and texture information to improve segmentation and retrieval performance. We are currently integrating the algorithm into a CBIR system. The system has the ability to search based on image content using the colour and texture models learned through the segmentation process.

8. REFERENCES

[1] M Crouse, R Nowak, and R Baraniuk, “Wavelet-based statistical signal processing using hidden markov models,” *IEEE Trans. on Signal Processing*, April 1998.

[2] H Choi and R Baraniuk, “Multiscale image segmentation using wavelet-domain hidden markov models,” *IEEE Trans. on Image Processing*, September 2001.

[3] N G Kingsbury, “A dual-tree complex wavelet transform with improved orthogonality and symmetry properties,” in *ICIP*, Vancouver, Canada, September 2000.

[4] N G Kingsbury, “Complex wavelets for shift invariant analysis and filtering of signals,” in *Journal of Applied and Computational Harmonic Analysis*, May 2001.

[5] A Kam and W Fitzgerald, “A general method for unsupervised segmentation of images using a multiscale approach,” in *ECCV*, Dublin, Ireland, June 2000.

[6] J Romberg, H Choi, R Baraniuk, and N Kingsbury, “Multiscale classification using complex wavelets and hidden markov tree models,” in *ICIP*, Vancouver, Canada, September 2000.

[7] C W Shaffrey, I H Jermyn, and N G Kingsbury, “Psychophysical evaluation of image segmentation algorithms,” in *ACIVS*, Ghent, Belgium, September 2002.