BAYESIAN EXAMPLE BASED SEGMENTATION USING A HYBRID ENERGY MODEL.

Claire Gallagher and Anil Kokaram

Department of Electronic and Electrical Engineering,
Trinity College Dublin,
Ireland.

ABSTRACT

This paper describes a supervised segmentation algorithm which draws inspiration from recent advances in non-parametric texture synthesis. A set of example images which have been segmented *a priori* are used as a guide in the segmentation process. This new algorithm is built on the Bayesian framework and combines the strengths of both parametric and non-parametric modelling techniques. The suitability of the wavelet transform for texture modelling is highlighted and an outlier class condition is introduced as a means to increase the flexibility of the algorithm. Segmentation results demonstrate the potential of this new algorithm.

Index Terms— Image Segmentation, Non-Parametric and Parametric Modelling, Dual Tree-Complex Wavelet Transform, Markov Random Field.

1. INTRODUCTION

The segmentation of an observed image into an unknown number of distinct regions remains a fundamental issue in low-level image analysis. Depending on the amount of information given a priori, the segmentation problem can be classified as unsupervised (observed image only), semi-supervised (observed image and the number of regions present) or supervised (observed image, number and description of the regions present). By its nature unsupervised segmentation is an ill-posed problem. This paper is related to semi-supervised and supervised techniques addressing a problem that is less ambiguous.

Direct applications of a successful segmentation algorithm are broad and varied, including medical and satellite image analysis, content based retrieval and image compression. Spurred on by the benefits of a fully or semi automated segmentation, there have been many different techniques developed and good reviews of these techniques can be found in [1, 2]. Of the previous approaches developed, those based on the Bayesian framework have been found to be the most stable and robust. Bayesian segmentation algorithms can be classified as parametric or non-parametric. Parametric models de-

scribe the image using some definable process, characterising image behaviour using a finite set of parameters each of which need to be estimated. In contrast, non-parametric approaches offer no such definite model and rather attempt to describe image behaviour implicitly by taking measurements from either the image itself or some set of example training images. Because of their efficiency and suitability to image analysis, parametric methods have been the most popular. However, non-parametric approaches have shown impressive results in texture synthesis applications [3, 4, 5] and have been recently applied to the segmentation problem [6].

This paper presents a new supervised segmentation algorithm exploiting an example based likelihood function. The observed image is modelled implicitly using empirical measurements taken from an example set of images of similar content. This non-parametric modelling technique is derived from the texture synthesis algorithm given in [5]. Unlike the fully non-parametric approach proposed by Mignotte [6], the algorithm described here captures the smooth nature of the label field by modelling it as a Markov Random Field (MRF). Both the parametric and non-parametric energies are combined in a Bayesian framework. To improve the flexibility of the algorithm an outlier class condition will be introduced and the benefits of multi-resolution analysis are exploited by performing segmentation in the wavelet domain. Segmentation results demonstrate the potential of this new algorithm.

2. THE LIKELIHOOD ENERGY REQUIRED

Let $\mathbf{I}$ denote the observed image and $\mathbf{L}$ its segmentation. Both $\mathbf{I}$ and $\mathbf{L}$ are defined on the lattice $\mathbf{X}$ and indexed using the spatial vector $\mathbf{x}$. Since it contains both spatial and statistical information, texture is chosen as the feature descriptor for the underlying image model. The observed image can now be to be considered as a mixture of textures where each pixel is given as,

$$I(\mathbf{x}) = T_{L(\mathbf{x})}(\mathbf{x}) \tag{1}$$

where $T_{L(\mathbf{x})}(\mathbf{x})$ represents the intensity value associated with the texture labelled $L(\mathbf{x})$ at position $\mathbf{x}$. The problem of segmentation now becomes that of: (i) estimating the textures

$\mathbf{T_i}$ for $i = 1, \cdots, K$ which make up the observed image and (ii) mapping the label field $\mathbf{L}$ using information from the estimated textures and the observed image. A typical parametric approach would characterise each texture region using a finite number of parameters. For example, the $2D$ Auto Regressive (AR) model given by,

$$T_{L(\mathbf{x})}(\mathbf{x}) = \sum_{k=1}^{P} a_k T_{L(\mathbf{x})}(\mathbf{x} + \mathbf{q}_k) + \epsilon(\mathbf{x}) \qquad (2)$$

where $\epsilon(\mathbf{x}) \sim \mathcal{N}(0, \sigma_\epsilon^2)$, $\mathbf{q}_k$ are offset vectors and the P coefficients of the model are denoted a_k for $k = 1, \cdots, P$. Recent work in texture synthesis [7] has shown that non-parametric modelling process outperform parametric approaches in terms of the accuracy with which they can capture the image behaviour. Based on this research the approach taken here will model the observed image implicitly using only heuristic measurements from some set of example texture images which have been segmented *a priori*. In order for this non-parametric model to be accurate each of these texture images must satisfy the following conditions: (i) they must be statistically similar to the textures found in the image to be segmented and (ii) they must be large enough to capture the underlying statistics of the infinite texture pattern. Figure 1 shows an example of an observed image $\mathbf{I}$ to be segmented and the example training texture set $\{\mathbf{T_1}, \mathbf{T_2}, \mathbf{T_3}, \mathbf{T_4}\}$ which will be used to implicitly model $\mathbf{I}$.

(a) Observed image   (b) Example textures
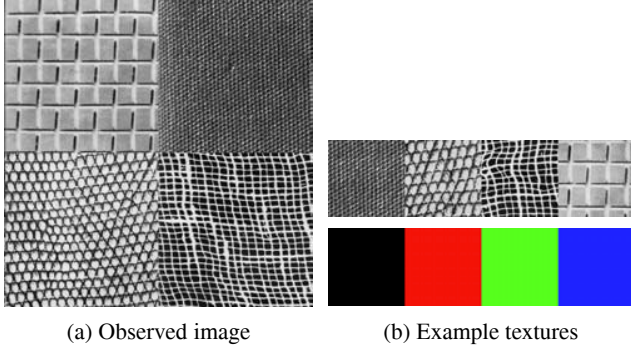
Fig. 1. Example based segmentation input parameters (a) observed image $\mathbf{I}$ and (b) Training textures $\{\mathbf{T_1}, \mathbf{T_2}, \mathbf{T_3}, \mathbf{T_4}\}$.

To apply this non-parametric modelling technique to the segmentation problem, the most likely label at each site is that which minimises the energy between the observed image and example images. That is, to estimate the most likely label $L(\mathbf{x})$ for the pixel $I(\mathbf{x})$, its $w \times w$ non-causal neighbourhood $N(\mathbf{x})$ is constructed and compared to all possible neighbourhoods in each of the example textures. The distance between two neighbourhoods, denoted by $D(N(\mathbf{x}), N(\mathbf{p}))$ is given by the L_2 distance metric. The best matching neighbourhood N_{best} is found and the label associated with the texture from which N_{best} is located is then taken as the most likely label for $I(\mathbf{x})$. This local energy minimisation takes place at each

site $\mathbf{x} \in \mathbf{X}$ resulting in the minimising of the following global energy function,

$$U(\mathbf{I}, \mathbf{T_e}) = \sum_{\mathbf{x} \in \mathbf{X}} \sum_{i=1}^{K} \min_{\mathbf{p} \in \mathbf{X_{e_i}}} D(N(\mathbf{x}), N(\mathbf{p})) \qquad (3)$$

Finding the configuration that minimises expression (3) is equivalent to estimating the maximum likelihood (ML) segmentation of $\mathbf{I}$. To illustrate the strength of this non-parametric energy based model, Figure 2 shows the ML segmentation obtained using (a) the 2D AR model and (b) the non-parametric model. A neighbourhood width of 9×9 was used in both modelling processes. The observed image and the training texture set are given in Figure 1.

(a) Parametric   (b) non-parametric
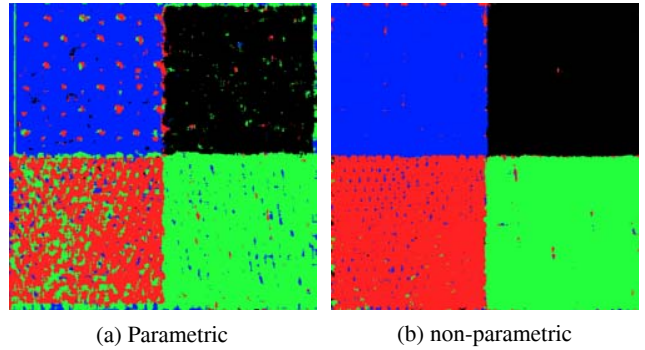
Fig. 2. Comparing parametric and non-parametric ML segmentations.

In terms of efficiency, the parametric approach demands less computational power and is much faster. In terms of accuracy of the modelling process, the non-parametric modelling process performs better and the resultant segmentation contains fewer mis-classified pixels. It should be noted that the strength of the non-parametric approach is very much dependent on the visual similarity of the training and observed textures. In cases where the training textures are dis-similar to the observed textures, the parametric model will offer a better generalisation and so the parametric segmentation will be better.

3. INTRODUCING THE PRIOR

ML segmentation will tend to give a solution similar to the data. Intuitively the label field should be smooth and such smoothness is introduced by imposing regularisation energy over the label field. The segmentation algorithm developed by Mignotte [6] measures this energy using a non-parametric modelling process similar to that used in the intensity image. The result is a fully non-parametric segmentation algorithm. While this is novel, it does suffer from some serious drawbacks which the approach taken here aims to overcome. Firstly, if the label field and intensity field are modelled using only an example image, then the modelling process will only support configurations which are found in the

example image. To illustrate, if region c_0 is located beside region c_1 in the example image, then the modelling process will only support configurations where c_0 is located beside c_1. To overcome this, the approach taken here treats each training image as separate texture feature and the regularisation energy is calculated by enforcing a parametric Markov Random Field (MRF) model over the label field. Because of its ability to capture smooth behaviour, the Potts model was chosen. Combining both the likelihood energy from the non-parametric model and the prior energy from the parametric model yields the following function which an MAP segmentation seeks to minimise,

$$U(\mathbf{I},\mathbf{L},\mathbf{T_e},\mathbf{L_e}) = \sum_{\mathbf{x}\in\mathbf{X}}\left(\sum_{i=0}^{K-1}\min_{\mathbf{p}\in\mathbf{X_e}} D(N(\mathbf{x}),N(\mathbf{p}))\right.$$
$$\left. + \alpha\sum_{k=1}^{P}\beta_k(1-\delta(L_e(\mathbf{p}),L(\mathbf{x}+\mathbf{q_k})))\right) \qquad (4)$$

where α provides a weighting between the data driven and regularisation terms. To obtain an approximation to the global minimum energy configuration, the ICM process is used. At each site $\mathbf{x}\in\mathbf{X}$, the label which minimises the local energy is found and this process is repeated until convergence is achieved. Pixel labels are updated in a checkerboard scanning order and the initial segmentation is obtained using the data driven energy only, i.e. $\alpha = 0$ and all other iterations use $\alpha = 0.5$. A neighbourhood size of 9×9 was used. Before discussing the multi-resolution implementation an outlier class condition which has been introduced will be discussed.

4. THE OUTLIER CLASS

The outlier class is assigned to any texture that is found in the observed image which is not considered to be similar enough to any of the input example textures. This outlier class is useful in cases when only certain regions or objects are of interest in the segmentation such as content based retrieval. To estimate an approximate threshold under below which an observed texture will be assigned an example texture label, the distribution of distances of neighbourhoods which belong to the same class was examined. To calculate this distribution each possible neighbourhood in the example texture was compared to every other possible neighbourhood in the same example texture. For an $M_e \times N_e$ sized texture, there will be $(M_e \times N_e)^2$ distances and this set is composed of the following entries,

$$\Upsilon = \{D(N(\mathbf{p}),N(\mathbf{x})),\forall \mathbf{p},\mathbf{p}\in\mathbf{X_e}\} \qquad (5)$$

where Υ is the distribution of neighbourhood distances and $\mathbf{x}$ and $\mathbf{p}$ are two spatial vectors used to index the lattice $\mathbf{X_e}$ on which each example texture is defined.

The distribution Υ is assumed to be Gaussian and so the mean and variance can be used to calculate confidence intervals within which class membership will lie. Any neighbour-
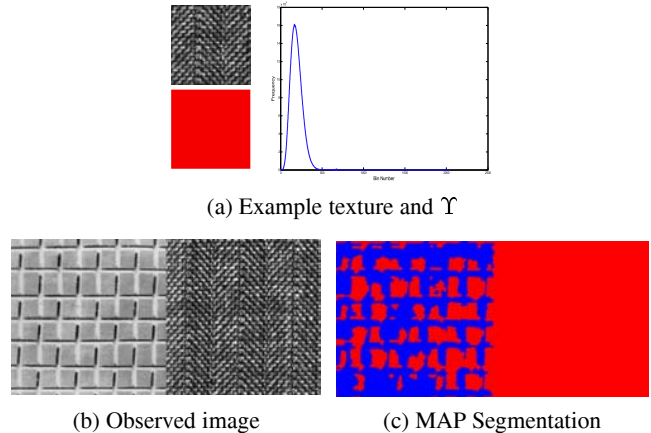

(a) Example texture and Υ

(b) Observed image   (c) MAP Segmentation

Fig. 3. Segmentation showing outlier class.

hoods whose distances fall outside this threshold are assigned the outlier class. Figure 3 (a) shows a distribution of neighbourhood distances for the example texture given and images (b) and (c) show the observed image and its segmentation. The outlier texture is indicated by the blue label.

5. DT-CWT SEGMENTATION

The advantages of multi-resolution image analysis has been well established. With its shift invariance and good directional selectivity, the Dual Tree-Complex Wavelet Transform (DT-CWT) [8] provides a natural domain for the modelling and analysis of textures. Evidence of this is shown in [5] and the means by which the energy based function given in (4) is extended to work within the DT-CWT domain follows on from the method discussed in [5]. At each level l of the DT-CWT, six directionally selective sub-band images denoted $\mathbf{B^l}$ are created. These sub-band images are used in a multi-directional neighbourhood comparison similar to that described in [5]. Neighbourhoods used in the energy minimisation process are now 6D and consist of wavelet coefficients from each of the sub-band images. At each level of the transform, the algorithm seeks to minimise an energy function similar to that given in (4)

Moving from a coarse to fine resolution, energy minimisation takes place at each level l using the segmentation obtained at level $l+1$ as an initialisation. Beginning at the coarse level the initial segmentation is obtained using the data driven term only, i.e. $\alpha = 0$. Figure 4 shows a ML (c) and MAP (d) segmentation of the frame image shown in (a). Each of the example textures (b) were taken from an earlier frame in the sequence. Considering that this is a difficult problem to solve as the lighting conditions change with frame number the algorithm performs well at capturing the main regions in the image.

Figure 5 shows a MAP segmentation using (a) the method proposed by Mignotte [6] and (b) DT-CWT Segmentation.
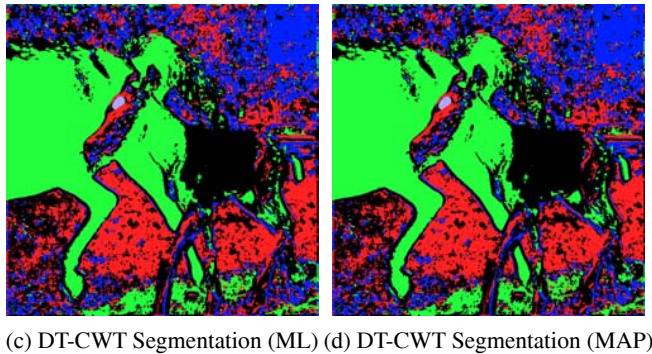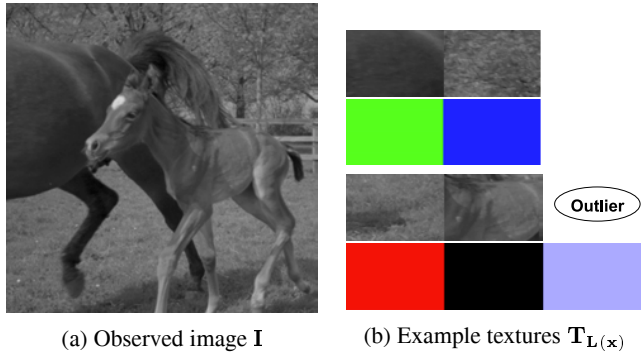
(a) Observed image **I**

(b) Example textures $\mathbf{T_{L(x)}}$

(c) DT-CWT Segmentation (ML) (d) DT-CWT Segmentation (MAP)

Fig. 4. Finding objects in sequences.

(a) Observed image **I**

(b) Example textures $\mathbf{T_{L(x)}}$

(c) DT-CWT Segmentation   (d) Mignotte [6] Segmentation

Fig. 5. Comparing segmentation results.

Note the poor segmentation at the boundary regions in segmentation (c) due to different configurations between the example and observed images. The result in (d) avoids this. Future work will consider how the smoothness of the DT-CWT Segmentation can be improved by using an alternative model over the label field.

6. FINAL COMMENTS

We cannot claim that this method as presented is computationally effective given the large amounts of neighbourhood searches to be conducted, however given the success of efficient implementations of the Efros class of synthesis algorithms, we are confident that this is not a major issue. Neither can we claim that we have demonstrated substantially better segmentation than other supervised techniques, but our results are certainly an improvement over the work of Mignotte. However we have demonstrated that there is interesting potential here for exploring the usefulness of example based segmentation. The outlier class condition certinaly increases the flexibility of the algorithm. One of the weaknesses of the nonparametric approach is its need for the example and observed textures to be perceptually similar. This problem affects all supervised algorithms in that the training set must be rich enough for segmentation to be useful. Certainly there is scope for combining parametric and example-based techniques for the purpose of better generalisation. Example based segmentation remains an interesting and intriguing idea as compute
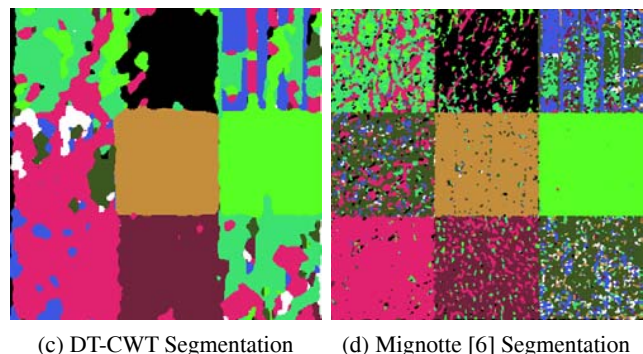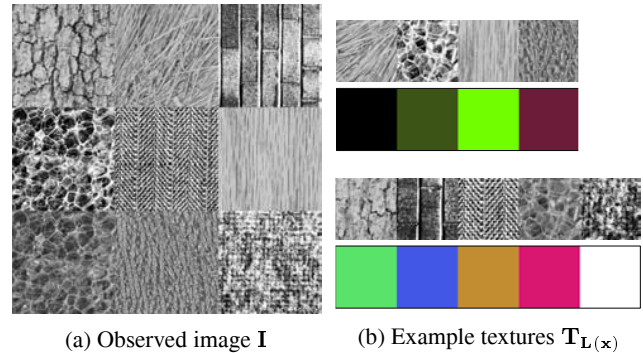
resources and database sizes increase.

7. REFERENCES

[1] J. Serra, "Image segmentation," in *IEEE International Conference on Image Processing (ICIP)*, September 2003, vol. 1, pp. 345–348.

[2] N. Pal and S. Pal, "A review on image segmentation techniques," *IEEE Pattern Recognition*, vol. 26, pp. 1277–1294, 1993.

[3] A.A. Efros and T.K. Leung, "Texture synthesis by nonparametric sampling," in *IEEE International Conference on Computer Vision*, Corfu, Greece, September 1999, pp. 1033–1038.

[4] L.Y. Wei and M. Levoy, "Texture synthesis over arbitrary manifold surfaces," in *ACM SIGGRAPH*, 2001.

[5] C. Gallagher and A.C. Kokaram, "Nonparametric wavelet based texture synthesis," in *IEEE International Conference on Image Processing (ICIP)*, Genova, Italy, September 2005.

[6] M. Mignotte, "Nonparametric multiscale energy-based model and its application in some imagery problems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 2, pp. 184–197, February 2004.

[7] C. Gallagher, *Example Based Image Processing*, Ph.D. thesis, University of Dublin, Trinity College, October 2006.

[8] N. Kingsbury, "Shift invariant properties of the dual-tree complex wavelet transform," in *IEEE Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 1999.