OBJECT-BASED IMAGE SEGMENTATION USING DWT/RDWT MULTIRESOLUTION MARKOV RANDOM FIELD

Lei Zheng, J. C. Liu, A.K. Chan, W. Smith*[+]
Department of Electrical Engineering
*Department of Computer Science
Texas A&M University
College Station, Texas 77843
[+]Naval Research Lab, Stennis Space Center

ABSTRACT

This paper introduces a segmentation algorithm for object-based image coding techniques. This scheme is based on *Discrete Wavelet Transform (DWT)/Redundant Discrete Wavelet Transform (RDWT)* and *Multiresolution Markov Random Field (MMRF)*. DWT based MMRF works well for noise-free images. It merges details in the original image with their respective visual objects and divides the image into different segments according to their textures. The RDWT based MMRF is a generalization of the DWT based MMRF. When the noise level is high, RDWT based MMRF reduces the influence of noise in the segmentation procedure and generates much better results at some computing costs. The proposed algorithm has been successfully integrated with our DWT based Region of Interest (ROI) compression coder, the Generalized Self-Similarity Trees (GST) codec, for networking applications.

1. INTRODUCTION

Object-oriented image coding is a new and promising technique for visual communications [1][2]. It has key features for applications such as Video-on-Demand and other visual communications for the next generation Internet. Object-oriented image coding techniques usually consist of two major steps: segmentation and coding. In the first step, images are divided into different segments according to a given segmentation model. These segments are then coded with parameters that are based on different criteria. Because of this, the segmentation result greatly affects the effectiveness of the whole coding scheme.

Recently, Markov Random Field (MRF) based image segmentation method has attracted a lot of attention in the literature and proven to be very successful [3][4]. It provides a convenient and consistent way to model context-dependent entities in image processing. Most images in visual communications are composed of different textural regions. For example, in videophone applications, speakers' faces have a different texture than that of the background; in telemedicine applications, diseased organelles have different appearances than their normal counterparts. Given an image, MRF optimizes the segmentation output according to the image texture based on the *maximum a posteriori* (MAP) probability criteria. In addition, Discrete Wavelet Transform has been widely used as an efficient way for image coding. Some of the well-known schemes are the Embedded Zerotree Wavelet (EZW) algorithm and Set Partitioning in Hierarchical Trees (SPIHT) algorithm. They can achieve a high compression ratio with good visual quality. Because of its efficiency, the DWT based image coding schemes are also widely used in video coding and visual communications.

In this paper, we describe an image segmentation algorithm suitable for object-oriented image coding based on DWT/RDWT and MMRF. It can be easily integrated with DWT based object-oriented image coders for object-oriented coding. By extending our scheme with RDWT, we show that the proposed algorithm can be used in various conditions for image communications.

2. MARKOV RANDOM FIELD

Most images in visual communications are composed of different textural regions. Let $Y=\{y_{j,k}\}$ denotes an observed image. Without loss of generality, it can also be re-indexed as $Y=\{y_i\}$, where $i=\{1,2, \ldots,W\times L\}$ and $W\times L$ is the size of image Y. After segmentation, Y is labeled by segmentation result $X=\{x_i\}$, $i=\{1, 2, \ldots, W\times L\}$, where $x_i=k$ means the pixel at position i in the original image Y belongs to label-k in the segmented image X, where $k=\{1, 2, \ldots, K\}$ and K is the total number of different labels. According to Hammersley-Clifford theorem [6], the probability density of X is given by a Gibbs density, which has the following form:

$$\begin{aligned} P(X) &= Z^{-1}\times e^{-U(X)} \\ Z &= \sum_{X\in RX} e^{-U(X)}, \\ U(X) &= \sum_{c\in C} V_c(X) \end{aligned} \qquad (1)$$

Where $U(X)$ is the energy function which is the summation of clique potentials $V_i(X)$ $\{i\in Z\}$ over all possible cliques, C. In MRF, a clique consists of a set of pixels that are neighbors to each other, and the potential function $V_i(X)$ depends on the local configuration of cliques. The energy function of MRF can be expressed as the summation of its cliques according to their sizes:

$$U(X)=\sum_{\{i\}} V_1(x_i)+\sum_{\{i,j\}\in C_2} V_2(x_i,x_j)+\sum_{\{i,j,k\}\in C_3} V_3(x_i,x_j,x_k)+\ldots, \qquad (2)$$

Here, the multilevel logistic (MLL) model is employed for MRF. The potential function for two-pixel cliques in MLL is defined as:

$$V_2(x_i,x_j)=\begin{cases} -\beta_j, & \text{if } x_i=x_j \quad \text{and} \quad i,j\in C, \\ \beta_j, & \text{otherwise} \end{cases} \qquad (3)$$

where β_j is a constant associated with x_j. The potential function for one-pixel cliques is defined as:

$$V_1(x_i) = \alpha_k \quad if\ x_i = k \quad and\ i \in C, \tag{4}$$

where α_k is a parameter associated with label *k*. According to Bayes' theorem:

$$p(X \mid Y) \propto p(Y \mid X)p(X), \tag{5}$$

where $p(X)$ is the *a priori* density of the region, and $p(Y \mid X)$ is the conditional density of the observed image given the distribution of the region. For a given image *Y* of size *W*×*L*, the *a posteriori probability* mass function for pixel labels *X* can be expressed as:

$$p(X \mid Y) = Z^{-1}e^{-U(X|Y)}, \tag{6}$$

where Z^{-1} is a normalizing constant. According to equation (1) and (5), the corresponding energy function is:

$$U(X|Y) = U(X) + U(Y|X), \tag{7}$$

where U(X) equals equation (2). Assume that the observed image *Y* is the combination of a true image *X* and an independent Gaussian noise, we have *Y=X+E*, where $E = \{e_i, 0 < i < W \times L\}$, and $e_i \sim N(x_i, \delta^2)$. Then, the utility function can be expressed as

$$U(Y|X) = \sum_{i=1}^{WL}[\frac{1}{2}\ln(\delta_{x_i}^2) + \frac{(y_i - x_i)^2}{2\delta_{x_i}^2}], \tag{8}$$
$$X = \{x_1, x_2, ..., x_i, ..., x_{W\times L}\}\ and \quad Y = \{y_1, y_2, ..., y_i, ..., y_{W\times L}\},$$

The objective of the MRF segmentation scheme is to assign a label value to each pixel in the original image Y, such that the energy function *U(X|Y)* is minimized. For more details on MRF theories, the reader is referred to [7].

3. MMRF SEGMENTATION BASED ON DWT AND RDWT

3.1 Discrete Wavelet Transform and Wavelet Frame

Wavelet decomposition for function *f(t)* is expressed as:

$$W_\psi f(b,a) = \int_R f(t)\frac{1}{\sqrt{a}}\overline{\psi}\left(\frac{t-b}{a}\right)dt = \left\langle f(t), \psi_{b,a}(t)\right\rangle, \tag{9}$$

where $\psi_{b,a}(t)$ is the wavelet that satisfies the admissibility conditions $C_\psi = \int_0^\infty \frac{|\overline{\psi}(\omega)|^2}{\omega}d\omega < \infty$. Here, $\psi(\omega)$ is the Fourier transform of $\psi(t)$ and $\psi_{b,a}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right)$ is the wavelet function at the scale *a*, shift-location *b*. One can retrieve the function *f(t)* from $W_\psi f(b,a)$ by the double integration reconstruction formula:

$$f(t) = \frac{1}{C_\psi}\iint W_\psi f(b,a)\psi_{b,a}(t)db\frac{da}{a^2}, \tag{10}$$

By defining L = *(g(n))* and *H=(h(n)), n*∈ *Z,* the discrete wavelet transform can be implemented by quadrature mirror filters (QMF). Here, L denotes a halfband lowpass filter, H denotes a halfband highpass filter and $\phi(x)$ is the scale function corresponding to wavelet $\psi(x)$. *g(n)* and *h(n)* are defined as:

$$g(n) = 1/2 < \phi(x/2), \phi(x-n) >, \tag{11}$$

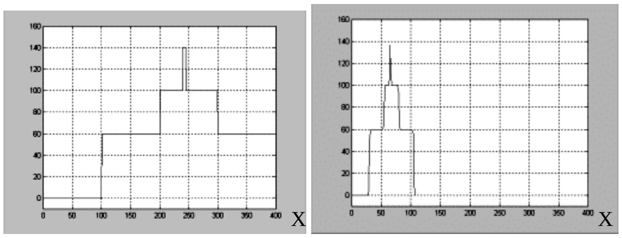
$$h(n) = (-1)^n g(1-n), \tag{12}$$

We use the 2-dimensional version of the DWT (2D-DWT) to analyze digital images. The easiest way to obtain the two-dimensional wavelets and scaling functions is to use the tensor product of 1-D wavelets and 1-D scaling function along the vertical and horizontal directions. In QMF, this means the L and H filters are applied to the image in both horizontal and vertical directions. The outputs are subsampled by a factor of two and consist of three high-pass subbands, HH, LH, HL and one low-pass subband LL. The process is repeated on the LL band to generate the next level of the decomposition using the same method. More detailed discussion on the above statements can be found in [5].

3.2 MMRF segmentation based on DWT

When we use object-based image processing for videophone application, it is natural to separate the face of the speaker in a videophone frame as one segment, while the body of the speaker and the background as other segments. However, this objective is not easy to achieve by conventional MRF schemes. The eyes, the mouth and the shadows on the face are often mistaken as different segments by conventional approaches, adding unnecessary complexities to the coding scheme. We propose a Multiresolution MRF (MMRF) segmentation scheme, based on DWT, to overcome this problem. The MMRF segmentation technique has been used for image segmentation in several cases, and is proven to be a fast and robust segmentation method [4]. Different from existing algorithms that depend on downsampling techniques to acquire different resolution images, we choose the DWT. By taking LL subband images for MRF segmentation, we can effectively merge those unnecessary small regions and produce only the meaningful regions for subsequent coding.

After an image is decomposed into subband images, the LL subband contains only low frequency information. In other words, small regions in the original image that contribute to the high frequency information in the original image have been smoothed out. These small regions are potentially more likely to be removed during the segmentation procedure in LL subband images. We can illustrate this using Figure 1, which is the one dimensional case. Figure 1 (a) is the original function $f_0(x)$, and Figure 1(b) is $f_1(x)$, which is the L subband of the twice DWT decomposed $f_0(x)$. If we select 0, 60, 100 and 140 as the mean value to initialize the segment procedure for $f_0(x)$, in Figure 1(a), there will be six points from position *x=240* to *x=245* belong to the set with mean value 140. But in Figure 1(b), there are no such points. In 2D cases, this means that the wavelet transform removes those small regions in the original image during the initialization step. Even if the small regions can not be removed in the initialization procedure, it becomes easier to be removed in later processing. This is because when a pixel in a small region is surrounded by larger ones, there will be more neighbors with different labels. Thus, the energy calculated by equation in (7) decreases when this pixel label is switched to those that belongs to the neighboring large regions.



(a) (b)

Figure 1. (a) Function $f_0(x)$. (b) $f_1(x)$.

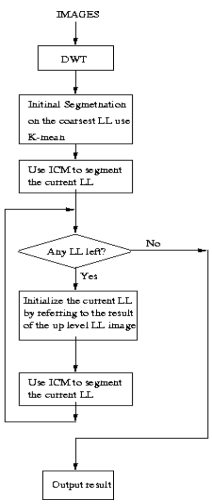


Figure 2. MMRF segmentation procedure using DWT

Figure 2 describes the segmentation procedure. Images are first decomposed into wavelet subbands. The segmentation starts from the LL subband at the lowest resolution level and the *k-means* algorithm is used for the initialization of the MRF at this level. For subsequent finer resolutions, we use results from their previous levels as the initialization of MRF procedure. The initialization is based on the self-similarity map between adjacent resolution levels (Figure 3). That is, assuming that $X_{k-1}[i][j]$ is the segmentation label value of pixel *(i, j)* in the (k-1)th resolution, after initialization, the label value of pixel *(m,n)* in level K can be expressed as $X_k[m][n]= X_{k-1}[i][j]$ (*if* $\lfloor m/2 \rfloor = i$, *and* $\lfloor n/2 \rfloor = j$.). After initialization, *Iterated Conditional Modes* (ICM) is used to get the new segmentation result. This procedure is repeated up until to the finest resolution level.

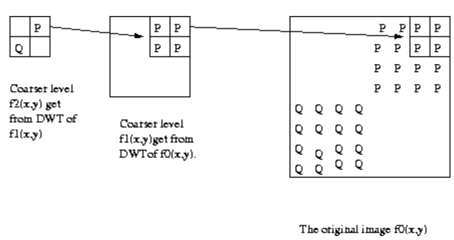


Figure 3. Initialization procedure between different resolutions.

3.3 MMRF segmentation based on Redundant DWT

The segmentation scheme based on DWT and MMRF works well when the foreground of the image frame is clean, such as in the videoconference applications. However, in other cases such as remote sensing and telemedicine, the images to be coded are often polluted by noises from image sources. For example, in ocean geography, seabed structures and sediments are often blurred by sea surface waves. In screening mammography, abnormal tissues may be covered by regular tissues. Strong noise often affects the segmentation results. To alleviate these impacts, we develop the Redundant DWT (RDWT) based MMRF to reduce the impact of noises on segmentation results.

RDWT has been investigated and applied to many fields [8]. In RDWT, the downsampling operation is skipped after the signal has passed through L and H filters. Thus, the size of the image remains unchanged in every processing stage. Figure 4 describes the way we use RDWT to perform MMRF. First, RDWT is applied to the input image. We get the LL subband image which contains only low frequency information but has the same image size as the original image. This image is called *resolution 0*. Then, by downsampling this image by the factor of 2 at each resolution level, coarser images of *resolution 1, resolution 2 …* are generated. The subsequent segmentation uses the MMRF procedure introduced before, starting from the coarsest level and propagating upwards to the finest level *resolution 0*. Since LL subband of RDWT contains only low frequency information of the original image, MMRF segmentation with RDWT removes unwanted noise in the original image. The RDWT procedure serves to remove the noise for an efficient MMRF procedure. Although the RDWT has a computation complexity of *O(nlgn)* while that of DWT is only *O(n)*, the computation speed is not adversely affected for common size images.

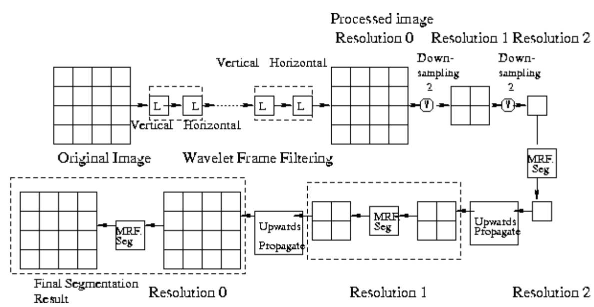


Figure 4. MMRF with RDWT

4. EXPERIMENT RESULTS

The source images for our experiment are *Miss America* and *lab scene* (Figure 5). Both of them are 176*144 gray scale images with 255 gray levels.


(a) Miss America. (b) *lab scene*

Figure 5. Images used for experiments.

Figure 6 compares the segmentation results of normal ICM scheme with proposed MMRF using DWT (using ICM at each resolution level). The normal MRF scheme generates small regions which will increase overheads in object oriented coding procedure. Our proposed method effectively separates the speakers' faces, shoulders and background. By using this result, the ROI based GST codec can correctly set different compression ratios for different objects in original image frames. The speakers' faces are kept at high visual quality when the network bandwidth resources are scarce.

Figure 7 compares results of DWT based MMRF with RDWT based MMRF using three examples. The first column images are original images including a sinusoidal-noise polluted sea-sediment photo, an aerial photo of rocks in seabed with white waves on the sea surface and a mammogram with cancer and fatty tissues. In object-based image coding, we want to distinguish the sediments, the rocks in the seabed and the cancer tissues. When DWT based MMRF is used, the segmentation results are distorted by the noise (Figure 7(a)) or generated many false regions caused by noises (Figure 7(b) and 7(c)). However, RDWT based MMRF can successfully reduce the impacts from unwanted noises and segment the input image successfully.

5. DISCUSSION AND CONCLUSIONS

We introduce an object-based segmentation scheme based on DWT/RDWT and MMRF. From experimental results, we show that using our scheme, image frames can be effectively divided into meaningful image object segments. This technique is useful in object-based image coding and interactive applications for visual communications.

6. ACKNOWLEDGMENT

The research presented is supported by Naval Research Lab, Stennis Space Center, through contract N00173-98-c-6004.

7. REFERENCE

[1] H.G. Musmann, M. Hoetter and J. Ostermann, "Object-Oriented Analysis – Synthesis coding of Moving Images," Signal Processing, Image Communication, 1:117-138, 1989;

[2] M. Kunt, M. Benard and R. Leonardi, "Recent Results on High-Compression Image Coding," IEEE Trans. On Cir. and Sys., 34:1306-1336;

[3] R.C. Dubes, A.K. Jain, S.G. Nadabar, and C.C. Chen, "MRF model based algorithms for image segmentation," Proc. ICCAS, pp. 808-814, 1990;

[4] C. Bouman and B. Liu, "Multiple resolution segmentation of textured images," IEEE Trans. Pattern Anal. Machine Intel. Vol. 13. No. 2, pp. 99-113, 1991.

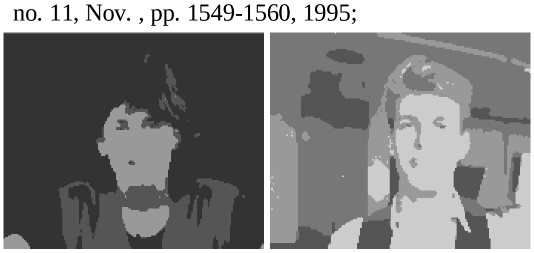
[5] C.K. Chui, "Wavelets: A tutorial in theory and applications," Academic Press, Inc., 1992;

[6] J. Besag, "Spatial interaction and the statistical analysis of lattice systems", Journal of the Royal Statistical Society, Series B 36, pp. 192-236, 1974;

[7] A.K. Jain, "Markov random fields : theory and application," Boston, Mass. : Academic Press, c1993;

[8] M. Unser, "Texture Classification and Segmentation Using Wavelet Frames," IEEE Trans. On Image Processing, vol. 4, no. 11, Nov. , pp. 1549-1560, 1995;

(a)



(b)

(c)

(d)

Figure 6. (a) Segmentation results of common MRF using ICM. (b) Segmentation results of DWT based MMRF segmentation. (c) Region of interest based on segmentation results of *Figure(b)*. (d) Coding result of GST codec with the ROI at compression ratio 1 but the background at 50.

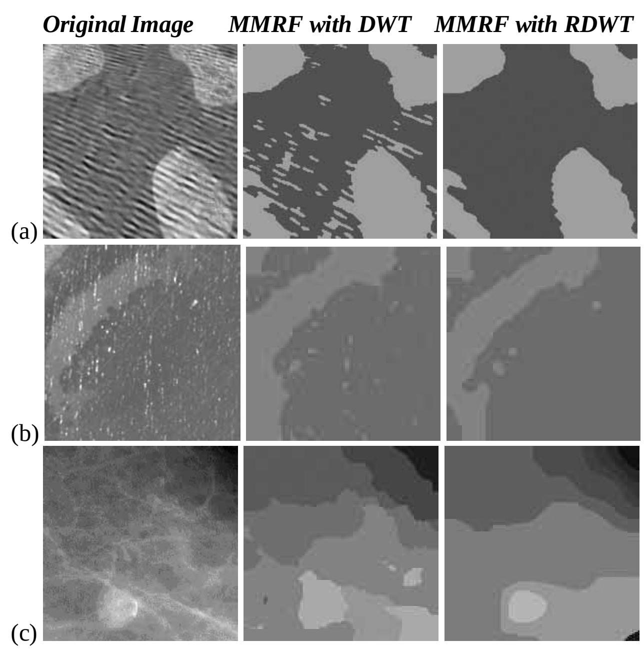


Figure 7. The first (from left to right) column images in (a)–(c): Original image of a sea sediments image polluted by sinusoidal signal; Original image of part of an aerial seabed photo with white waves; Original image of part of a mammogram with cancer and fatty tissues. The second column images in (a)-(c) Results using DWT based MMRF. The third column images in (a)-(c): Results using RDWT based MMRF.