MULTI RESOLUTION IMAGE SEGMENTATION FOR SCALABLE OBJECT-BASED WAVELET CODING

Fardin Akhlaghian Tab[1], Alfred Mertins[2] and Habibollah Danyali[1]

[1] **School of Electrical, Computer and Telecommunications Engineering**
University of Wollongong, Wollongong, NSW 2522, Australia, Email: {fat98, hd04}@uow.edu.au

[2] **Signal Processing Group, Institute of Physics, University of Oldenburg, Oldenburg, Germany**
Email: {alfred.mertins}@uni-oldenburg.de

ABSTRACT

This paper introduces a multi resolution image segmentation algorithm for scalable object based wavelet coding. This algorithm is based on discrete wavelet transform and multiresolution Markov random field (MMRF) modelling. The major contribution is to match the spatial scalability features of arbitrary shape wavelet transforms with the segmentation algorithm. To optimize the segmentation/extraction of objects/regions of interest in different resolutions of the wavelet pyramid, with scalability constraint, a multi scale analysis is incorporated into the objective function of MMRF segmentation algorithm. The proposed algorithm improves the segmentation result, especially in lower resolutions of the decomposition, over regular multi resolution segmentation in both objective and subjective tests, in yielding an effective segmentation that supports scalable wavelet coding.

1. INTRODUCTION

With increasing popularity of multimedia in more and more applications on networks and the internet, new multimedia services such as interactivity, manipulation and scalability are necessary. To meet these requirements digital image/video coding has changed from block based to object based [1]. The idea of objects has provided a lot of functionality beside coding efficiency, such as manipulation and scalability. Meanwhile, object based wavelet coding schemes have become increasingly important and gained widespread acceptance.

One major pre-processing objective for any object based coding is image/video segmentation and shape extraction. Therefore, with spatial scalability in mind , it is necessary to extract the objects in multiple resolutions in a way that is useful for scalable object based wavelet coding. In this paper, we propose an image segmentation algorithm which fits multiresolution Markov random field segmentation to object based scalable wavelet coding. The objects/regions of interest are extracted in different resolutions while keeping wavelet scalability as a constraint. A multi scale analysis is incorporated into the objective function of the MMRF segmentation algorithm, in order to align the segmentation with the wavelet scalability constraints. For optimization, the Iterated Condition Mode (ICM) algorithm according to [3], matched to the scalable multi scale analysis, has been used.

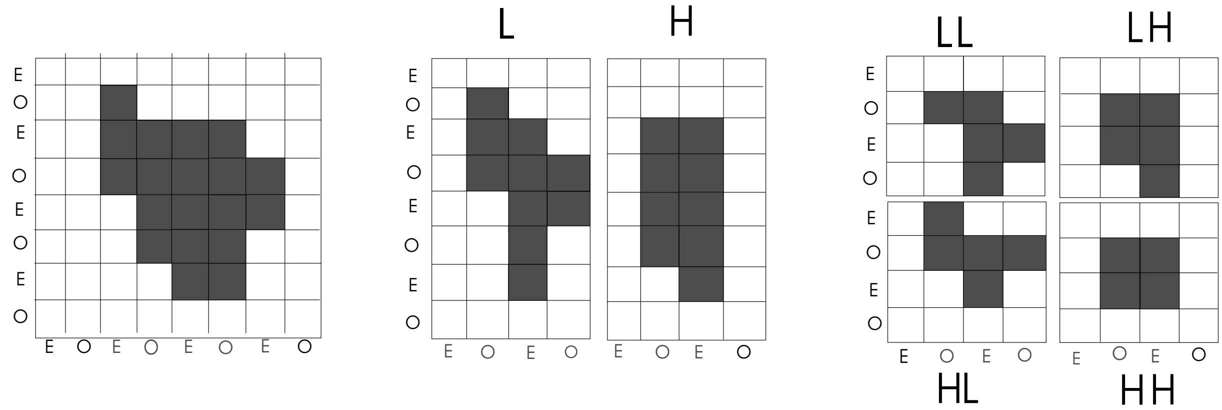
This paper is organized as follows. Section 2 refers to the scalability in wavelet coding. In Section 3 we present a scalable multiresolution segmentation algorithm that includes a statistical image modelling and optimization processes. Some experimental results and discussion are presented in Section 4, and finally, conclusions are drawn in Section 5.

2. OBJECT BASED WAVELET CODING SCALABILITY

Scalability means the capability of decoding a compressed sequence at different data rates. It is useful for image/video communication over heterogenous networks which require high degree of flexibility from the coding system. Scalable image/video coding has also different applications such as web browsing, image/video database systems, video telephony, etc.

In wavelet based spatial scalability applications, due to the self similarity feature of the wavelet transform, the shape in lower scale is the shape in the lowpass (LL) subband. The exact relationship between the full-resolution shape and its low-resolution versions depends on the kind of wavelet transform used for the decomposition. In this paper we use an odd length filter (e.g. 9/7), where all shape points with even indices[1] are downsampled for the lowpass band [4]. Figure 1 further illustrates the wavelet decomposition of arbitrarily shaped objects when using an odd-length filter. The final four-band decomposition is depicted in Figure 1(c). As a result, every shape point with even index has a corresponding point on the lower resolution and every shape point on the lower level has a correspond-

[1] Suppose indices start from zero or an even number



(a) (b) (c)

Figure 1: Decomposition of a non rectangular object with odd-length filters; (a) the object, shown in dark gray; (b) the decomposed object after horizontal filtering; (c) decomposed object after vertical filtering. The letters "E" and "O" indicate the position(even or odd) of a pixel in the horizontal and vertical dimensions.

ing point on the next higher level. By considering the self similarity of the wavelet transform, it is straightforward to suppose that the points of a shape with even indices have the same segmentation classifications as the corresponding points on the lower level.

The wavelet self similarity extends to all low pass subband shapes of different levels. Therefore the discussed relationship between corresponding pixels is extended to shapes on different scales. Corresponding pixels on different levels have the same segmentation class.

3. IMAGE SEGMENTATION ALGORITHM

To solve an image processing problem by MRF technique, a statistical image model has to be fitted to the application which captures the intrinsic character of the image in a few parameters. Then the image/video processing problem, including all uncertainties and constraints, can be converted to a mathematical parameter optimization problem [5].

3.1. Statistical image model

The main challenge in multiresolution segmentation for scalable wavelet based object coding is to keep the same relation between extracted object as it exists between the decomposed object in different resolution in the arbitrary shaped wavelet transform. To meet this challenge, Markov Random Field modelling is selected because it includes low level processing on pixels and has enough flexibility in defining objective functions for the problem at hand [5]. By applying the wavelet transform to the original image, a pyramid of decomposed images at various scales is created. Let Y be the gray levels of this pyramid's pixels and $Y_p(s)$ be the intensity of point s in level p. Similarly, segmentation of image into regions at different resolutions will be denoted by X, where $X_p(s) = i$ means that the pixel s at level p is set to class i. In the MMRF segmentation algorithm, an estimation of X is obtained by maximizing an *a posteriori* probability function $P(X|Y)$. From Bayes' theorem [6]

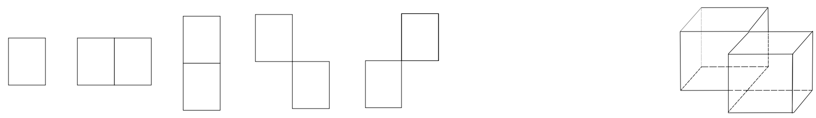
$$P(X|Y) \propto P(Y|X)P(X)$$

where $P(X)$ is the *a priori* probability of the region process and $P(Y|X)$ is the conditional probability of the observed image given the segmentation in different scales. The label field X is modelled as a Markov Random Field (MRF). Spatial continuity is incorporated into the segmentation algorithm because they are inherent to MRFs. It is well known that an MRF is Gibbs distributed and is completely defined by its energy function $U(X)$. This energy function can be written as a sum of potential functions $V_c(X)$ so that

$$P(X) = \frac{1}{Z}\exp\{\sum U(X)\},\ U(X) = \sum_c V_c(X)$$

where Z is a normalization factor. A clique is a set of points that are neighbors of each other. A clique function depends only on the pixels that belong to a clique V_c. In single level segmentation, usually one and two pixels cliques are used as in Figure 2(a), and one assumes that the one pixel clique potentials are zero, which means that all region types are equally likely [3].

As mentioned earlier, with regard to scalability, a pixel and its corresponding pixels on the other levels have the same segmentation class. Therefore they can change only together during segmentation. To change the class of a point, the pixel and all its corresponding pixels on the other levels have to be analyzed together.



(a) (b)

Figure 2: (a) Normal one and two point clique. (b) a clique of two array clique with the arrays' length equal to two pixels.

As a result, multi-pixel or a vectors[2] analysis instead of pixel analysis needs to be used. The cliques are extended to act on the vectors space. A vector is a combination of a pixel and its corresponding pixels on the other levels. The dimension of a vector depends on the index of it's pixels and it can be 1, 2 or more. In this work, we use cliques of two arrays instead of two pixels. Figure 2(a) shows regular one-pixel and two-pixel cliques. In Figure 2(b), the extension of one of these cliques to the array mode with the arrays' dimension equal to two can be seen. The extension of clique functions is achieved through the following steps: equation (1) is used for cliques of two pixels on the level P where pixels s_p and r_p are two neighboring pixels on the resolution P. Equation (2) is defined for multiple levels, where $\{s_p\}$ and $\{r_p\}$ vectors correspond to two neighboring pixels s_p and r_p on level P. The lowest scale of the clique vector's points is M and its dimension denoted as N. The other parameter, β, has a positive value, so that two neighboring pixels on the same scale are more likely to have the same class than two different classes. Increasing the value of β has the effect of decreasing sensitivity to grey level changes [3].

$$V_{c_1}(s_p, r_p) = \begin{cases} -\beta & \text{if } X(s_p) = X(r_p) \quad s_p, r_p \in \mathrm{c}_1 \\ +\beta & \text{if } X(s_p) \neq X(r_p) \quad s_p, r_p \in \mathrm{c}_1 \end{cases} \tag{1}$$

$$V_{c_N}(\{s_p\}, \{r_p\}) = \sum_{k=M}^{M+N-1} (-1)^{L_k}.\beta \,,$$
$$L_k = \begin{cases} 1 & \text{if } X(s_k) = X(q_k) \quad s_k, q_k \in \mathrm{c}_N \\ 0 & \text{if } X(s_k) \neq X(q_k) \quad s_k, q_k \in \mathrm{c}_N \end{cases} \tag{2}$$

Instead of an image in one resolution, we have decomposed the image into different levels and the summation on pixels is over different pixels on different levels. Therefore the objective function can be written as the following

$$\frac{1}{2\sigma^2} \sum_{\{S\}} \{||Y(\{s\}) - \mu_{\{s\}}^{X(\{s\})}(\{s\})||^2 + \sum_{r_p \in \partial s_p} V_{c\{s_p\}}(\{s_p\}, \{r_p\})\} \tag{3}$$

where the first summation is over vectors while the second summation is over all possible vector cliques of vector $\{s\}$. The gray level of points $\{s\}$ form a vector $Y(\{s\})$, and similarly, $\mu(\{s\})$ and $X(\{s\})$ are mean and classification vectors. The probability density function has two terms. One constrains the region intensity to be close to the data; the other imposes spatial continuity. Increasing σ^2 is equivalent to increasing β. Thus, for simplifying the expression, the parameter σ^2 is set to 1, and the segmentation result is controlled by the value of β.

In normal multiresolution segmentation, low level segmentation affects high level segmentation. But the proposed model considers the use of both high and low levels together, and therefore, there is also some effect of high level information on low level segmentation and vice versa. As the results will show, this two-way effect gives better segmentation results especially in low level.

3.2. Algorithm to find the MAP estimation

A method to minimize the probability function formula (3) has to be used. After initial segmentation with the k-means clustering algorithm, the optimization method, iteration condition mode (ICM), improves the accuracy of the segmentation estimation. Details of the used ICM is similar to the single level segmentation algorithm of Pappas [3], but matched to our problem which is a scalable multiresolution segmentation algorithm.

First, The estimation of $\mu_p^i(s)$ is considered. It is estimated by averaging the gray levels of all pixels that belong to the region i and inside a window with width w centered at pixels s in the level p. The window size w for each resolution is twice the window size of lower level and half of the next higher level's window size.

Second, the estimation of the distribution of regions is considered. Given the intensity function $\mu_p^i(s)$, we must maximize the *a posteriory* probability density (3) to obtain the MAP estimation of X. To obtain the local minima, the ICM approach has been used. Originally, that is maximizing the conditional density at each point $X_P(S)$ given the data Y and the current segmentation of all other points in the image [7]. But in this work, extended to a scalable multiresolution segmentation, a vector is analyzed and the conditional density of the vector's points, given the data Y and the current segmentation of all other points in the pyramid, is maximized. If the syntax $\{s_p\}$ de-

[2]Direction is not important and the word "vector" is used for convenience instead of set of points.

termines a set of vector's pixels include s on level p and all it's corresponding pixels in lower and higher levels of the decomposition, we have to maximize the following conditional density related to vector $\{s_P\}$.

$$P(X\{s_p\}|Y,X) = P(X\{s_p\}|Y, X(\{r\}),\, r \in N_{\{s_p\}}\,)$$

$$\propto \exp\{-\frac{1}{2\sigma^2}\sum_{s\in\{s_p\}}[y(s)-\mu^{x(s)}(s)]^2 - \sum_{s\in\{s_p\},r\in\partial s} V_c(s,r)\} \quad (4)$$

Now we consider the overall algorithm. The initial segmentation of the pyramid is obtained by the k-means clustering algorithm. The average of any point S and its correspondence on the other levels $\{S\}$ is used to classify the points of $\{S\}$ to one label. Now given the regions label X we process the pyramid's vectors' points, progressively from low to high resolutions. At each resolution we estimate the intensity $\mu_p^i(s)$ at each pixel s in the frame for all possible classes i with a pre-determined window size w used for estimation. Then we update the estimate of X_p using the ICM approach with a multi level analysis by formula (4). By updating the class pixels of level p the corresponding pixels in the other scales are also updated. The algorithm then moves to the next scale and updates the estimates of μ and X and so on, until all scales are processed. The process is repeated until convergence. The stopping criterion is that the update of X in each resolution changes a number of class indices that is below a pre-defined threshold. To reduce the number of iterations, other convergence criteria can also be used. The whole procedure may be repeated with smaller window size. The algorithm then stops when the minimum window size for the lowest level is reached. We have considered the minimum of 5 as the minimum window size of lowest level.

4. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, the results of processing frame 15 of the CIF sequence Claire and the Lena image with the proposed segmentation algorithm are presented. The results are compared to a regular single and multiresolution segmentation algorithm [3]. In the first step, the image is decomposed into three resolutions, using the (9/7) wavelet filter. Then in each level of the decomposition, the image is segmented and the objects of interest (such as Claire's head and shoulder) are extracted according to the presented algorithm in different resolutions. Scalability between objects/regions in different resolutions, as required for the arbitrary shape wavelet transform, is achieved with the proposed algorithm. The image and its segmentation by scalable segmentation and regular multiresolution segmentation are shown in Figure 3.

Table 1: SPIHT PSNR results for Claire object.

Resolution	Rate (BPP)				
	0.05	0.1	0.25	0.5	1.0
72×88	23.7	27.3	32.2	34.7	42.3
144×176	24.4	28.5	34.2	36.7	46.8
288×352	26.1	31.4	37.3	44.5	57.8

Table 2: Regions number of Claire image seg.

Seg. algorithm	88×72	176×144	352×288
Multi Resolution	46	71	93
Scalable	72	98	116
Single level	–	–	138

As an application of this algorithm, a user, by a graphic user interface, roughly determines the object of interest by a closed contour. Then all the regions with a predetermined percentage of their area inside this closed contour are selected as the region of the extracted object. Joining of all selected regions creates the final object. As an example, a user has roughly determined the objects of interest in Fig. 4(a). The algorithm then determines the exact borders of the object in different resolutions as shown in Figure 4. This algorithm can detect all regions including regions with concave borders, overcoming a shortcoming in some object detection algorithms. And finally we implement an object-based modification of the highly scalable SPIHT (HS-SPIHT)[8] algorithm which allows us to code only the pixels that belong to an arbitrarily shaped object. The extracted object, Claire head and shoulder, is then coded by the object-based HS-SPIHT [8]. Table 1 shows peak signal-to-noise ratio (PSNR) results obtained for three spatial resolutions at different bit rates, all decoded from a single bit-stream.

In lower scales of regular multi-resolution segmentation, brief and compact versions of the image are processed. But in the proposed scalable segmentation algorithm, due to the multi scale analysis in the optimization process, there is also some effect of higher resolution information on lower level segmentation. This improves the results of the proposed algorithm. In the following, we explain some of these improvement on the results of Claire image segmentation.

Table 2 shows the number of detected regions of the Claire image in three spatial resolutions for different segmentation algorithms. The proposed scalable segmentation detects more relevant regions than multiresolution, and nearly the same as single level segmentation. In lower scales of regular multiresolution segmentation, brief and compact versions of images are processed, and therefore some small size or low contrast regions are not detected even in the higher resolutions. But in the introduced algorithm the effects of

(a) (b) (c)

Figure 3: Claire image segmentation with $k = 5$ cluster and $\beta = 50$; (a) the main image; (b) segmentation by the proposed algorithm; (c) regular multiresolution segmentation.

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

Figure 4: Claire object extraction; (a) First selection; (b) object at 288×352; (c) object at 144×176; (d) object at 72×86;

Table 3: Misclassified pixels in noisy images.

Resolution	Multiresolution	Scalable	Single Level
Object.	$\%11.98$	$\%9.8$	$\%15.48$
Image	$\%17.34$	$\%17.8$	$\%33$

Table 4: Number of Misclassified pixels.

Resolution	72×88	144×176	288×352
Down sample.	$\%7.59$	$\%4.85$	0
Scalable	$\%4.95$	$\%2.45$	$\%0.92$

Table 5: Means of curvature estimation.

Resolution	72×88	144×176	288×352
Downsample.	0.5614	0.5408	0.4802
Scalable	0.5214	0.529	0.4936
Difference	$\%4.3$	$\%2.2$	$-\%2.8$

high resolutions on low resolutions results in the detection of more significant regions than regular multiresolution segmentation. In other words, the algorithm leads to a better detection of small or low contrast objects especially in low resolutions. This feature can be useful in applications which require sharp separation of objects/regions from background.

Multi-resolution segmentation algorithm overcomes noisy images better than single level segmentation. To test the scalable segmentation algorithm on noisy images, a uniform noise in the range $(-30\ ,\ +30)$ is added to the Claire image, and the number of misclassified pixels for Claire object include head and shoulder (70553 pixels in high resolution of scalable segmentation) as well as the entire image pixels are counted. The results in Table 3 show that the proposed algorithm can deal with the noisy images the same as multiresolution image segmentation and much better than single level segmentation. This result confirms that the introduced multilevel segmentation algorithm keeps most advantages of multiresolution segmentations over single level segmentations such as better segmentation of noisy images.

To compare the segmentations obtained with different algorithms, the result of a standard multi-level segmentation is accepted as a ground truth. Note that this ground truth does not satisfy the shape constraints inflicted by the arbitrary shape wavelet transform, but at each level, it gives an appealing segmentation. For the proposed algorithm and the standard single-level technique, the segmentations for the different resolution levels are determined by downsampling the highest-level segmentations. The misclassified pixel numbers that occur when comparing the two scalable segmentation algorithms with the ground truth and are given in Table 4. The results confirm that the proposed

(a) (b) (c)

Figure 5: Lena Image segmentation with $k = 6$ clusters and $\beta = 50$; (a) The Lena image (256×256); (b) highest level segmentation of multiresolution segmentation; (c) highest level of scalable segmentation.

algorithm has lower number of pixel misclassifications in lower resolutions.

One criterion to have a good looking shape is objects/regions' border smoothness [9]. Especially in low resolution, the lower number of points intensifies the importance of smoothness on the subjective and objective test of shape quality. In high resolution levels, however, large number of pixels ensure the visual quality of the shapes. The smoothness of segmented regions are measured with an estimation of curvature [10], which is related to the change of tangent's slope in the borders pixels. At each region's border pixels, a low value of curvature is indicative of a smooth edge. Table 5 shows the smoothness of scalable segmentation and single level segmentation downsampled to lower levels. The results show that the proposed algorithm ensures smoother edges than the downsampled single level segmentation.

In the next example with the segmentation of Lena image the effects of scalable segmentation on the optimality of segmentation come to view. The image and it's regular multiresolution segmentation [3] and scalable segmentation both optimized by ICM technique are shown in Figure 5(a),(b) and (c). The comparison of regions above the hat, as a subjective test, shows that the scalable segmentation result is better than multi resolution segmentation. Tying the low level pixels to high level pixels or the multilevel analysis feature of the proposed algorithm leads the optimization process to produce better results than regular multiresolution segmentation especially when the used optimization methods gives local maxima such as the ICM technique [7]. This is more important in complicated images that have a higher number of possible locally optimum segmentation results.

5. CONCLUSIONS

We have presented a scalable image segmentation algorithm that is optimized to extract objects/regions, useful for object-based wavelet coding applications such as interactive multimedia transmission over the internet. In addition to scalability, the presented algorithm gives better shape quality or smoothness and a lower number of misclassified pixels in low resolutions compared to down sampling of regular segmentations. The proposed multi scale analysis improves the sensitivity to grey level variations but still performs well in noisy environment. In the future, we are going to work on the improvement of extracted regions/objects' smoothness especially on low resolution and extend the work to the video domain.

REFERENCES

[1] MPEG-4 Applications, ISO/IEC JTC1/SC29/EG11/N2322, July 1998.

[2] C.Christopoulos, A. SDkordas, and T.Ebrahimi, "The jpeg2000 still image coding system: an overview," *IEEETrans. Comsumer Electronics*, vol. 46, no. 4, pp. 1103–1127, Nov. 2000.

[3] T.N.Pappas, "An adaptive clustering algorithm for image segmentation," *IEEE Trans. Image Processing*, vol. 40, no. 4, pp. 901–914, Apr. 1992.

[4] A.Mertins and S.Singh, "Embedded wavelet coding of arbitrary shaped objects," in *Proc. SPIE 4076-VCIP'00*, Pert, Australia, June 2000, pp. 357–367.

[5] Stan Z.Li, *Markov Random Field Modeling in Image Analysis*, Springer Verlag, Tokyo: Japan, seconed edition, 2001.

[6] P.B. Chou and C.M.Brown, "The theory and practice of baysian image labeling," *International journal of computer vision*, vol. 4, pp. 185–210, 1990.

[7] J.Besag, "On the statistical analysis of lattice systems," *Journal of Royal Statistical Society, Series B*, vol. 48, no. 3, pp. 259–279, 1986.

[8] H.Danyali and A.Mertins, "Highly scalable image compression based on sphit for network applications," in *Proc. IEEE Int. Conf. Image Processing*, Rochester, NY, USA, Sept. 2002, pp. 217–220.

[9] R.M.Haralick and L.G.Shapiro, "image segmentation techniques," *Computer Vision, Graphics and Image Processing*, vol. 29, pp. 100–132, 1985.

[10] M.Kass, A.Witkin, and D.Terzopoulos, "Snakes: active contour models," *International journal of computer vision*, vol. 1, no. 4, pp. 321–331, 1987.