Proc. VIIth Digital Image Computing: Techniques and Applications, Sun C., Talbot H., Ourselin S. and Adriaansen T. (Eds.), 10-12 Dec. 2003, Sydney

Multiresolution Image Segmentation with Border Smoothness for Scalable Object-Based Wavelet Coding

Fardin Akhlaghian Tab[1], Golshah Naghdy[1], and Alfred Mertins[2]

[1] University of Wollongong,
Wollongong, Australia
{fat98, g.naghdy}@uow.edu.au
[2] University of Oldenburg
Oldenburg, Germany
alfred.mertins@uni-oldenburg.de

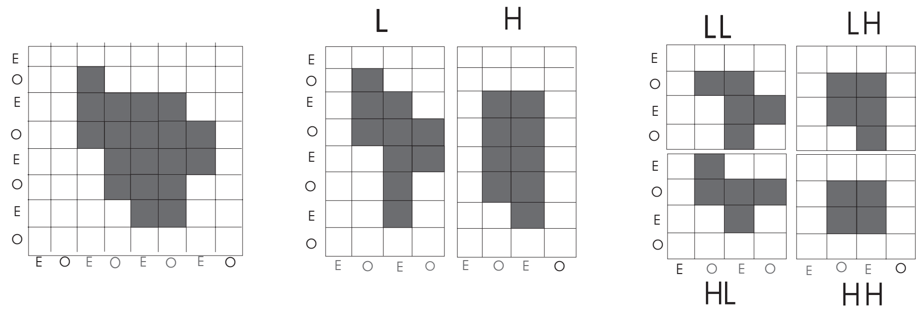
Abstract. This paper introduces a multiresolution image segmentation algorithm for scalable object-based wavelet coding applications. This algorithm is based on discrete wavelet transform and multiresolution Markov random field (MMRF) modelling. The major contribution of this work is to add spatial scalability and border smoothness in the segmentation algorithm usable for object-based wavelet coding algorithm. To optimize the segmentation/extraction of objects/regions of interest in all scales of the wavelet pyramid, with scalability constraint, a multiresolution analysis is incorporated into the objective function of MMRF segmentation algorithm. The proposed algorithm improves border smoothness in all regions, particularly in lower resolutions. In addition to scalability between objects/regions in different levels, the proposed algorithm outperforms the standard multiresolution segmentation algorithms, in both objective and subjective tests, in yielding an effective segmentation that supports scalable object-based wavelet coding.

1 Introduction

Object based coding has provided a large degree of flexibility in digital image and video processing and is expected to play a major role in future multimedia, computer games and related applications. It facilitates object interactivity, manipulations and scalability in a highly flexible manner [1]. Meanwhile, wavelet transform has become increasingly important and gained widespread acceptance in object-based coding. An example is the new JPEG-2000 still image compression standard [2].

One major pre-processing objective for any object-based coding is image segmentation and shape extraction. Therefore, bearing in mind spatial scalability, it is necessary to extract the image objects in multiple resolutions in a way that is useful for scalable wavelet based 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 multiresolution analysis is incorporated into the objective function of the MMRF segmentation algorithm, in order to align the segmentation with the wavelet scalability constraints. To extract more enhanced shapes/regions, border smoothness, as a criterion of shape analysis and image segmentation [3], has also been included in the objective function. In order to improve the borders in the lower level, where border smoothness suffers in the conventional segmentation, augmented smoothness coefficients are used in lower resolutions. For optimization, the iterated condition mode (ICM) algorithm according to [4] , adapted to the scalable multiresolution analysis, has been used.

This paper is organized as follows. Section 2 refers to the scalability in wavelet coding. In Section 3 we discuss about MMRF segmentation, including the statistical image modelling and optimization processes. Some experimental results and discussion are presented in Section 4, and finally, conclusions are drawn in Section 5.



(a) (b) (c)

Fig. 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.

2 Object-Based Wavelet Coding Scalability

Scalability means the capability of partially decoding a compressed bitstream to achieve various resolutions and/or quality of the original sequences. 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[3] are downsampled for the lowpass band [5]. Fig. 1 further illustrates the wavelet decomposition of arbitrarily shaped objects when using an odd-length filter. The final four-band decomposition is depicted in Fig. 1(c). By considering the self similarity of the wavelet transform, it is straightforward to suppose that the points of a shape with even index have the same segmentation classification as the corresponding points on the it’s lower level.

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

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 [6].

3.1 Statistical image model

The main challenge in multiresolution segmentation for wavelet-based object coding, is to keep the same relation between extracted objects in different levels as it exists between the decomposed objects in different resolutions in the arbitrarily shaped wavelet transform. The other constraint is border smoothness particularly in lower resolutions. To meet these challenges, Markov Random Field modelling is selected because it includes low level processing on pixel levels and has enough flexibility in defining objective functions matched with the problem at hand [6]. In a regular single level image segmentation based on Bayes' theorem the objective or energy function is [4] :

$$E(X) = \sum_{s} \{ \frac{1}{2\sigma^2} (Y(s) - \mu_s^{X(s)})^2 - \sum_{r \in \partial_s} V_c(s,r) \} \tag{1}$$

where The main summation is over all pixels, $Y(s)$ is the grey level of pixel s, X is the pixel classification, $\mu_s^{X(s)}$ is the mean of grey level values of pixels with class $X(s)$ in a window centered at s, $V_c(s,r)$ is a clique of two pixels defined at s and the internal summation is over all neighboring pixels around s. It is clear that the objective function has two terms. One constrains the region intensity to be close to the data; the other imposes spatial continuity. Increasing s is equivalent to increasing V_c. Thus, to simplify the expression, the parameter $2\sigma^2$ is set to one, and the segmentation result is controlled by the value of V_c. To obtain the final segmentation, this objective function is minimized by one of the several MRF objective minimization methods [6].

To tailor this objective function to our application, the wavelet transform is applied to the original image and a pyramid of decomposed images at various scales is created. Let Y be the grey levels of this pyramid's pixels and $Y_p(s)$ be the intensity of pixel s in scale 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 scale p is set to i.

A clique is a set of neighboring pixels. A clique function depends only on the pixels that belong to the clique. 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 alike [4].

As mentioned, 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 pixel, the pixel and all its corresponding pixels on the other levels have to be analyzed together. As a result, a set of pixels or a vector[4] analysis instead of a single pixel analysis needs to be used. The cliques are extended to act on the vector space. A vector is a combination of a pixel and its corresponding pixels on all other levels. The dimension of a vector depends on the index of its pixels and it can be 1, 2 or more. In this work, we use cliques of two arrays instead of two pixels. Fig. 2(a) shows regular one and two pixel clique sets. In Fig. 2(b), the extension of one of these cliques to the array mode can be seen.

The extension of clique functions is achieved through the following steps: equation (2) is used for cliques with length one on the level P where pixels s_p and r_p are two neighboring pixels on the scale P. Equation (3) 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, the clique vector is M with its dimension denoted as N. β is a positive parameter, 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 [4].

$$V_{c_1}(s_p, r_p) = \begin{cases} -\beta & \text{if } X(s_p) = X(r_p) \quad (s_p, r_p) \in \mathsf{c}_1 \\ +\beta & \text{if } X(s_p) \neq X(r_p) \quad (s_p, r_p) \in \mathsf{c}_1 \end{cases} \tag{2}$$

$$V_{c_N}(\{s_p\}, \{r_p\}) = \sum_{k=M}^{M+N-1} (-1)^{L_k}.\beta\,, \quad L_k = \begin{cases} 1 \text{ if } X(s_k) = X(q_k) \ (s_k, q_k) \in \mathsf{c}_N \\ 0 \text{ if } X(s_k) \neq X(q_k) \ (s_k, q_k) \in \mathsf{c}_N \end{cases} \tag{3}$$

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

$$\sum_{\{S\}} \{ ||Y(\{s\}) - \mu_{\{s\}}^{X(\{s\})}(\{s\})||^2 + \sum_{r_p \in \partial s_p} V_{c\{s_p\}}(\{s_p\}, \{r_p\}) \} \tag{4}$$

The first summation is over vectors while the second summation is over all possible cliques of vector $\{s\}$. The gray level of points $\{s\}$ form a vector $Y(\{s\})$ similarly $\mu(\{s\})$ and $X(\{s\})$ are mean and classification vectors.

Smoothness of the regions' border is one of the main features considered for the region extraction [3]. Therefore in the proposed algorithm, objects'/regions' border smoothness constraint has also been considered. Ideally these borders are the edges in the image. Edge is one of the most important properties for visual perception. Most natural objects exhibit smooth edges. Edge smoothness is enhanced in high resolution and that's why some objects/regions are visually pleasing only at high resolution and

[4] Direction is not important and the word "vector" is used for convenience.

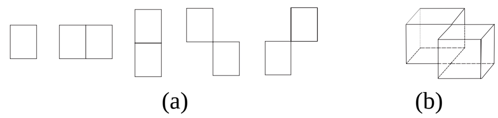


Fig. 2. (a) Normal one and two pixels cliques sets. (b) A clique of two arrays with the arrays' dimension equal to two.

the visual quality suffers at low resolution. To reduce this effect, we have enhanced the smoothness at low resolution more rigorously than high resolutions. The priority is considered by the bigger smoothness coefficients for lower resolution. The smoothness is measured by curvature value [7] at each pixel. Therefore the objective function is updated to the following equation:

$$E(X) = \sum_{\{S\}} \{ ||Y(\{s\}) - \mu_{\{s\}}^{X(\{s\})}(\{s\})||^2 + \sum_{r_p \in \partial s_p} V_{c\{s_p\}}(\{s_p\},\{r_p\}) + \sum_{q \in \{s\}} l_p * \nu(q)\} \tag{5}$$

where $\nu(q)$ shows curvature of pixel q and l_p is a coefficient which decreases with resolutions.

3.2 Algorithm to find the MAP estimation

A method to minimize the probability function equation (4) has to be used. After initial segmentation with k-means clustering algorithm, the optimization method, iteration condition mode (ICM) [8], improves the accuracy of the segmentation estimation. ICM optimizes the objective function pixel by pixel until convergence. At each pixel, segmentation of the processed pixel is optimized given the current X at all other pixels. Therefore the only term of objective function related to the current pixel needs to be minimized which are

$$E_s(X) = (Y(s) - \mu_s^{X(s)})^2 - \sum_{r \in \partial_s} V_c(s,r) \tag{6}$$

In multiresolution scalable mode the objective function term corresponding to a vector of pixels is optimized given the segmentation of all other vectors on the pyramid. And with smoothness constraint the term is:

$$E_{\{s\}}(X) = \{ ||Y(\{s\}) - \mu_{\{s\}}^{X(\{s\})}(\{s\})||^2 + \sum_{r_p \in \partial s_p} V_{c\{s_p\}}(\{s_p\},\{r_p\}) + \sum_{q \in \{s\}} l_q.\nu(q)\} \tag{7}$$

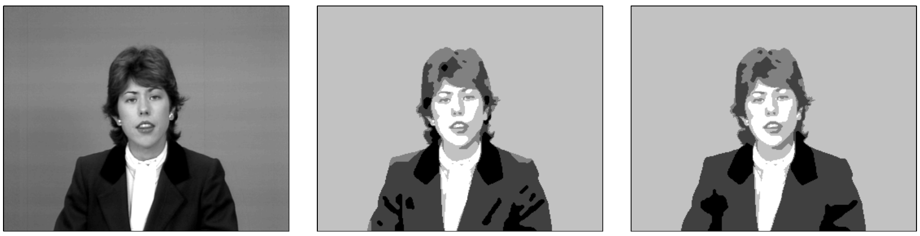
Details of the used ICM is similar to the single level segmentation algorithm of Pappas [4], but it is changed to adapt to scalable multiresolution segmentation algorithm.

First, the estimation of $\mu_p^i(s)$ is explained. It is estimated by averaging the gray 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 is halved when we move to the next lower level.

Now we consider the overall algorithm. The initial segmentation of the pyramid is obtained by 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 class 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 minimizing the energy equation (7). By updating the class points at level p the corresponding points in the other scales are also updated. The algorithm then moves to the next level and updates the estimates of μ_{p+1} and X_{p+1} and so on, until all resolutions are processed. Then, the process is repeated until convergence. The stopping criterion is that the update of X in each resolution changes a number of class indexes that are below a pre-defined threshold. To reduce the number of iterations, other convergence criterion can also be used. The whole procedure may be repeated with smaller window size until the minimum window size of the lowest level is reached.

4 Experimental Results and Discussion

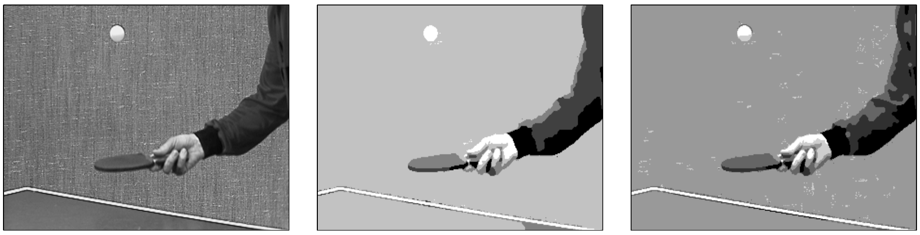
The proposed algorithm is tested using frame 15 of the CIF sequence Claire and frame 5 of SIF sequence Table tennis and CIF sequence Miss America. The results are compared with a regular single and multi-resolution segmentation algorithm [4]. At the first step, in each level of decomposition, the image is segmented and the objects of interest (such as Claire's head and shoulders) are extracted according to the proposed algorithm. Scalability between objects/regions in different levels, as required by the arbitrary shape wavelet transform, is achieved in the proposed algorithm. Figs. 3 - 5 represent the results achieved by the proposed algorithm compared with a well known single and multiresolution segmentation algorithms [4] using standard test images Claire and Table tennis.



(a) (b) (c)

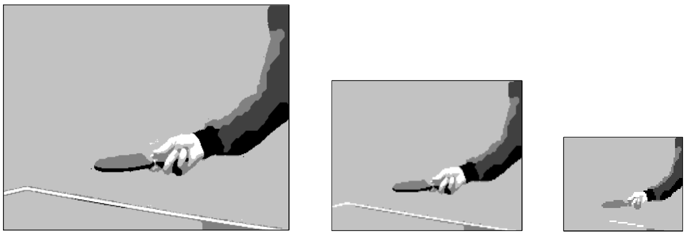
Fig. 3. Claire image segmentation with $k = 5$ clusters and $\beta = 50$; (a) the main image; (b) segmentation by the proposed algorithm; (c) regular multiresolution segmentation;

One of the applications of the proposed segmentation algorithm is object based image coding. To facilitate this, the user determine the rough boundary of the object of interest through a graphic user interface (GUI) program. 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 create the final object.



(a) (b) (c)

Fig. 4. Table tennis image segmentation with $k = 6$ clusters and $\beta = 100$; (a) the main image; (b) segmentation by the proposed algorithm; (c) regular single level segmentation;



(a) (b) (c)

Fig. 5. Normal multiresolution segmentation of Table tennis image with $k = 6$ clusters and $\beta = 100$; (a) 240×352 segmentation ; (b) 120×176 segmentation; (c) 60×88 segmentation;

As an example, a user has roughly determined the objects of interest in Fig. 6(a). The algorithm then determines the exact borders of the object in different resolutions as shown in Fig. 6. And finally the extracted image object, Claire head and shoulder, is then coded by a highly scalable object-based SPIHT algorithm [9]. 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.

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

Fig. 6. Claire object extraction; (a) First selection; (b) object at 288×352; (c) object at 144×176; (d) object at 72×88;

In lower scales of regular multiresolution segmentation, brief and compact versions of the image are processed, therefore, some small size or low contrast regions are not detected. But in the proposed algorithm the effects of high resolutions on low resolutions results in the detection of more significant number of regions than regular multi-

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

resolution segmentation. In other words, sensitivity to grey level changes are increased resulting in a better detection of small or low contrast objects especially in low resolutions. Table 2 shows the number of detected regions of Claire and Table tennis images in three spatial resolutions for different segmentation algorithms. The proposed scalable segmentation detects more relevant regions than the regular multiresolution algorithm. For example consider the segmentation of the textured wall and the detection of the ball in the table tennis image as presented in Figs. 4 and 5 by scalable segmentation, multiresolution and single level segmentation algorithms.

Table 2. Number of regions detected in Claire and table tennis images segmentation.

Frame	Claire			Table Tennis		
Seg. algorithm	72×88	144×176	288×352	60×120	120×176	240×352
Multi Resolution	46	71	93	19	55	164
Scalable	72	98	108	42	83	184
Single level	–	–	138	–	–	314

To compare the segmentations obtained with different algorithms, the result of a standard single-level segmentation at each level of wavelet decomposition 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 multi resolution technique, the segmentations at different resolution levels are determined by down sampling the highest-level segmentations. The misclassified pixel numbers that occur when comparing the two scalable segmentation algorithms with the ground truth are given in Table 3. The results confirm that the proposed algorithm has lower number of pixel misclassifications in lower resolutions.

Table 3. Number of Misclassified pixels.

image	Claire			Table Tennis		
Resolution	72×88	144×176	288×352	60×88	120×176	240×352
Down sample.	%7.59	%4.85	%1.57	%4.9	%5.26	%5.34
Scalable	%4.95	%2.45	%0.92	%3.3	%2.87	%2.72

Border smoothness is one of the contributory factors in presenting a visually pleasing extracted objects/regions. Specially in low resolutions, the lower number of pixels intensifies the importance of smoothness in 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 introduced algorithm increases the smoothness in lower levels by sacrificing it in higher levels. This is a direct result of tying the pixels of different level together. The smoothness of segmented regions' borders are measured with an estimation of curvature [7]. A low curvature value is indicative of a smooth edge. Table 4 shows the average smoothness for all borders pixels in scalable segmentation and multi resolution segmentation down sampled to lower levels at 3 different resolutions. The results show that, the proposed algorithm ensures smoother edges than the down sampled multi resolution segmentation.

Table 4. Means of curvature estimation.

Image	Claire			Table Tennis		
Resolution	72×88	144×176	288×352	60×88	120×176	240×352
Downsample.	0.52	0.54	0.51	0.63	0.51	0.35
Scalable	0.48	0.47	0.45	0.47	0.37	0.38
Improvement	$\%7.7$	$\%16.13$	$\%12.8$	$\%25$	$\%26$	$-\%9.9$

Segmentation optimization algorithms often find locally optimum results. Multi scale analysis and smoothness term improves the segmentation results. Classification of the low contrast border pixels with grey level close to both surrounding regions/classes is affected by smoothness term resulting in softer borders. This is interesting if we consider that the edge of most objects exhibits smoothness. Therefore, smoothness term helps the borders to resemble the normal objects/regions edges more closely. This will overcome the shortcomings of some region based segmentation algorithms in terms of border quality. In the next example with the segmentation of Miss America image the positive effect of smoothness term on the optimality of segmentation comes into view. The image and its regular Baysian-based segmentation [4] are shown in Fig. 7(a,b). Spurious edges can be seen in the left side of image where the intensity contrast is very low. Fig. 7(c) shows the final scalable segmentation with smoothness effect resulting in eliminating the spurious borders. Fig. 7(c) indicates superior performance of the proposed algorithm in subjective test.

(a) (b) (c)

Fig. 7. Miss America image segmentation with $k = 5$ clusters and $\beta = 50$; (a) Miss America image 288×352; (b) scalable segmentation; (c) scalable segmentation with smoothing constraint.

To test the scalable segmentation algorithm on noisy images, first a uniform noise in the range $(-30\ ,\ +30)$ is added to the Claire and table Tennis images then different segmentation algorithms are performed. The number of misclassified pixels for Claire object including head and shoulder (70553 pixels in high resolution of scalable segmentation) and Table tennis object including arm, racket and ball(11033 pixels) as well as the entire images pixels are counted. The results in Table 5 shows that the proposed algorithm can deal with noisy images as effectively as multi resolution segmentation and much better than single level segmentation. It is significant to note while maintaining noise tolerance, this algorithm has improved sensitivity to grey level variation.

Table 5. Misclassified pixels in noisy images.

Image	Claire			Table Tennis		
Algorithm	Multiresolution	Scalable	Single Level	Multiresolution	Scalable	Single Level
Object.	$\%11.98$	$\%9.8$	$\%15.48$	$\%2.2$	$\%1.74$	$\%3.67$
Image	$\%17.34$	$\%17.8$	$\%33$	$\%3.94$	$\%4.0$	$\%14.64$

5 Conclusions

We have presented a scalable image segmentation algorithm that is optimized to extract regions/objects, useful for object-based wavelet coding applications such as multimedia transmission over the heterogenous networks. In addition to scalability, the multi scale analysis gives better results than standard single and multi resolution segmentation in objective and subjective tests, especially at low resolutions. Smoothness constraints results in softer edges and better localization of regions borders. The presented algorithm gives better shape quality or smoothness and less number of misclassified pixels in low resolutions compared to down sampling of regular segmentations. The proposed multiresolution analysis improves sensitivity to grey level variation but still performs well in noisy environment. In the future, we are going to work on the effect of different wavelet filters on the scalable segmentation algorithm and extend the work to the video domain.

References

1. ISO/IEC JTC1/SC29/EG11/N2322: MPEG-4 Applications. (1998)
2. Christopoulos, C., SDkordas, A., Ebrahimi, T.: The jpeg2000 still image coding system: an overview. **46** (2000) 1103–1127
3. Haralick, R.M., Shapiro, L.G.: image segmentation techniques. Computer Vision, Graphics and Image Processing **29** (1985) 100–132
4. Pappas, T.N.: An adaptive clustering algorithm for image segmentation. IEEE Trans. Image Processing **40** (1992) 901–914
5. Mertins, A., Singh, S.: Embedded wavelet coding of arbitrary shaped objects. In: Proc. SPIE 4076-VCIP'00, Pert, Australia (2000) 357–367
6. Li, S.Z.: Markov Random Field Modeling in Image Analysis. seconed edn. Springer Verlag, Tokyo: Japan (2001)
7. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: active contour models. International journal of computer vision **1** (1987) 321–331
8. Besag, J.: On the statistical analysis of lattice systems. Journal of Royal Statistical Society, Series B **48** (1986) 259–279
9. Danyali, H., Mertins, A.: Highly scalable image compression based on sphit for network applications. In: Proc. IEEE Int. Conf. Image Processing, Rochester, NY, USA (2002) 217–220