Scalable multiresolution color image segmentation with smoothness constraint

F. Akhlaghian Tab
University of Wollongong

G. Naghdy
University of Wollongong, golshah@uow.edu.au

Alfred Mertins
University of Oldenburg, Germany, mertins@uow.edu.au

Recommended Citation

Scalable multiresolution color image segmentation with smoothness constraint

Abstract

This paper presents a multiresolution image segmentation method based on the discrete wavelet transform and Markov random field (MRF) modeling. A major contribution of this work is to add spatial scalability to the segmentation algorithm producing the same segmentation pattern at different resolutions. This property makes it applicable for scalable object-based wavelet coding. The correlation between different resolutions of pyramid is considered by a multire solution analysis which is incorporated into the objective function of the MRF segmentation algorithm. Examining the corresponding pixels at different resolutions simultaneously enables the algorithm to directly segment the images in the YUV or similar color spaces where luminance is in full resolution and chrominance components are at half resolution. Allowing for smoothness terms in the objective function at different resolutions improves border smoothness and creates visually more pleasing objects/regions, particularly at lower resolutions where downsampling distortions are more visible. In addition to spatial scalability, the proposed algorithm outperforms the standard single and multire solution segmentation algorithms, in both objective and subjective tests.

Disciplines

Physical Sciences and Mathematics

Publication Details

SCALABLE MULTIRESOLUTION COLOR IMAGE SEGMENTATION WITH SMOOTHNESS CONSTRAINT

Fardin Akhlaghian Tab*, Golshah Naghdy* and Alfred Mertins**

*School of Electrical, Computer and Telecommunications Engineering
University of Wollongong, Wollongong, NSW 2522, Australia
Emails: {fat98, golshah }@uow.edu.au

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

ABSTRACT

This paper presents a novel multiresolution image segmentation method based on the discrete wavelet transform and Markov Random Field (MRF) modelling. A major contribution of this work is to add spatial scalability to the segmentation algorithm producing the same segmentation pattern at different resolutions. This property makes it applicable for scalable object-based wavelet coding. The correlation between different resolutions of pyramid is considered by a multiresolution analysis which is incorporated into the objective function of the MRF segmentation algorithm. Examining the corresponding pixels at different resolutions simultaneously enables the algorithm to directly segment the images in the YUV or similar color spaces where luminance is in full resolution and chrominance components are at half resolution. Allowing for smoothness terms in the objective function at different resolutions improves border smoothness and creates visually more pleasing objects/regions, particularly at lower resolutions where downsampling distortions are more visible. In addition to spatial scalability, the proposed algorithm outperforms the standard single and multiresolution segmentation algorithms, in both objective and subjective tests.

1. INTRODUCTION

Effective segmentation is crucial for the emerging object-based image/video standards such as object-based coding standards defined by MPEG-4. In scalable object-based coding, a single codestream can be sent to different users with different processing capabilities and network bandwidths by selectively transmitting and decoding the related parts of the codestream. A scalable bitstream includes embedded parts that offer increasingly better SNR, greater spatial resolution or higher frame rates. Therefore considering the spatial scalability, it is necessary to extract and present objects' shape at different resolutions for the scalable object-based encoder/decoder systems. For an effective scalable object based coding algorithm, it is required that the shapes of the extracted objects at different resolutions be similar or equivalently, the pattern of segmented regions should be similar at different resolutions. We call the multi resolution segmentation algorithm with the similar patterns at different resolutions scalable segmentation.

In the most straight case, multi resolution segmentation algorithms consider the inter scale correlation by projection of the lower resolution segmentation result to the next higher resolution as an initial segmentation estimation. The segmentation is further refined at the current higher resolution by a single resolution segmentation. This procedure continues progressively until the highest resolution is segmented [1], [2]. These algorithms don't consider the inter scale correlation effectively and cannot detect small objects/regions if they are not detected in the lowest resolutions. In the second group of algorithms, the inter scale correlations are considered in the statistical models and decision at each pixel/block is based on the information of the different resolutions [3], [4], [5]. However, often only the causal inter scale correlation with the last lower resolution [3], [5], or the next higher resolution is considered [4]. None of the known works in the literature have considered inter-scale correlation between all pyramid resolutions. Considering the other resolutions results in a very complex model and increases the time complexity. In addition their extension to scalable segmentation, producing the same segmentation patterns at different resolution, is nearly impossible or results in an algorithm with large computational complexity.

In this paper we present a novel MRF based multiresolution grey/color image segmentation algorithm which extends the statistical model in order to consider the correlation between all the resolutions without overly increasing the computational complexity. It produces the same segmentation patterns at different resolutions and is applicable to object-based wavelet coding algorithms.

The multiscale analysis, incorporated in the objective function of the MRF-based segmentation algorithm, combines good features of both single and multiresolution segmentations. While it is noise resistant, it detects ob-

jects/regions better than regular multiresolution segmentation and also results in a lower number of regions than single-level segmentation.

Natural objects exhibit smooth borders/edges. Hence, to some extent there is correlation between visually pleasing objects and object's border smoothness. Since distortions such as downsampling often result in rough borders/edges, therefore in this work, a multi resolution smoothness criterion is incorporated in the objective function of the segmentation algorithm which results in more normal or visually pleasing objects/regions. By considering bigger smoothness coefficients for the smoothness terms of different resolutions, the distortion effect of downsampling is reduced.

Color images segmentation in YUV or LUV spaces produces more favorable results than the RGB space [6], [7]. Many of the images and image sequences in the databases are in YUV format where Y is in full resolution while U and V components are in half resolution. The proposed algorithm has enough flexibility to directly segment this kind of color and also grey level images.

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 a high degree of flexibility from the coding system. Some of the desirable scalable functionalities are signal-to-noise ratio (SNR) scalability, spatial scalability and temporal scalability. In particular spatial scalability means that, depending on the end user's capabilities (bandwidth, display resolution etc.), a resolution is selected and all the shape and texture information is sent and decoded at the appropriate resolution.

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. In this paper we use an odd length wavelet filter (e.g. 9/7), where all shape pixels with even indices[1] are downsampled for the lowpass band [8]. 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(b). As a result, every shape pixel with even indices has a corresponding pixel on the lower resolution. By considering the self similarity of the wavelet transform, it is straightforward to suppose that the pixels of a shape with even indices have the same segmentation classifications as the corresponding pixels 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 at different scales. Each pixels has a corresponding pixels at the any higher resolutions and pixels with indices that are multiples of 2^n in both dimensions are down sampled to the

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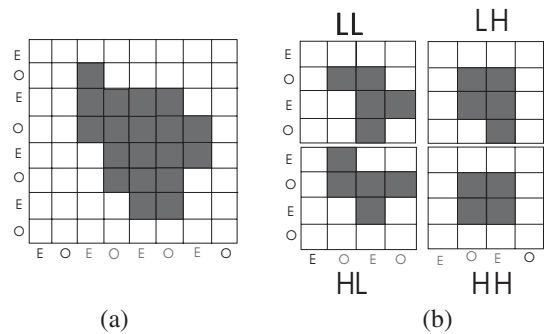
[1]Suppose indices start from zero or an even number.

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

next n lower scales. A pixel and it's corresponding pixels at the lower and higher resolutions form a set called corresponding pixels. Corresponding pixels at different resolutions have the same segmentation class.

3. SPATIAL SEGMENTATION ALGORITHM

3.1. Statistical color image model

The main challenge in multiresolution image segmentation for scalable object-based wavelet coding is to keep the same relation between extracted objects/regions at different resolutions as it exists between the decomposed objects/regions at different resolutions in a shape adaptive wavelet transform. The other constraint is border smoothness particularly in lower resolutions. Different smoothness coefficients defined at different resolutions give some degree of freedom to put more emphasis on the low-resolution smoothness. To meet these challenges, Markov random field modelling is selected as it includes low level processing at pixel level and has enough flexibility in defining objective functions matched with the problem at hand. In a regular MRF-based image segmentation the problem is formulated using a criterion such as the maximum a posteriori (MAP) probability. We first explain the objective function of single resolution grey/color image segmentation [1], [6] and then extend it to the scalable multiresolution segmentation mode. The desired segmentation is denoted by X, and Y is the observed color image with three channels shown by a three dimensional vector $Y = [Y_1, Y_2, Y_3]$. This results in the following cost or objective function which has to be minimized with respect to $X(s)$ [6]:

$$E(X) = \sum_s \left(\sum_{i=1}^{3} (Y_i(s) - \mu^i_{X(s)}(s))^2 + \frac{1}{T} \sum_{r \in \partial_s} V_c(s, r) \right)$$

(1)

Where the clique function in single resolution modes is defined by the following equation:

$$V_c(s,r) = \begin{cases} -\beta & \text{if} \quad X(s) = X(r) \\ +\beta & \text{if} \quad X(s) \neq X(r) \end{cases}, \ (s,r) \in \text{C} \quad (2)$$

Herein β is a positive number and s and r are a pair of neighboring pixels. Note that a low potential or energy corresponds to a higher probability for pixel pairs with identical labels and lower probability for pairs with different labels, which automatically encourages spatially connected regions.

To tailor this objective function to scalable multiresolution color image segmentation, initially, the wavelet transform is applied to the original image and a pyramid of decomposed images at various resolutions is created. Let $Y = [Y_1, Y_2, Y_3]$ where $Y_i, i = 1, 2, 3$ is the intensity of channel i of the pyramid's pixels. The segmentation of the image into regions at different resolutions will be denoted by X.

As mentioned earlier, considering scalability, a pixel and its corresponding pixels at all other pyramid levels have the same segmentation label. Therefore they change together during the segmentation process. To change the segmentation label of a pixel, the pixel and all its corresponding pixels at all other levels have to be analyzed together. As a result, an analysis of a set of pixels in a multidimensional space instead of a single resolution analysis needs to be used. The term "vector" is used to refer to multidimensional space. A vector includes corresponding pixels at different resolutions of the pyramid. A symbol $\{s\}$ shows a vector which includes pixel s. The dimension of a vector is equal to the number of it's pixels which are located at different resolutions. Using these primary definitions, the clique concept is extended to vector space. The extended cliques act on two vectors instead of two pixels. Figure 2(a) shows regular one and two pixels clique sets. In Figure 2(b), the extension of one of these cliques to the array mode in two dimensional space can be seen.

The extension of clique functions is achieved through the following steps: equation (2) is used for cliques of length two at a resolution where pixels s and r are two neighboring pixels at the same resolution level. Equation (3) below is defined for multiple levels, where $\{s\}$ and $\{r\}$ are vectors corresponding to two neighboring pixels s and r. The neighboring pixels of the two vectors $\{s\}$ and $\{r\}$ at level k are denoted as s_k and r_k. The lowest resolution which include a pixel of vector $\{s\}$ is denoted as M and N is the dimension of vectors $\{s\}$ and $\{r\}$. A positive value is assigned to the parameter β, so that adjacent pixels, of two neighboring vectors, located at a same resolution are more likely to belong to a same region than to different regions.

$$V_c(\{s\}, \{r\}) = (\frac{1}{N}) \sum_{k=M}^{M+N-1} (-1)^{L_k} . \beta \,,$$

$$L_k = \begin{cases} 1 \text{ if } X(s_k) = X(r_k) \\ 0 \text{ if } X(s_k) \neq X(r_k) \end{cases} \ s_k \in \{s\}, r_k \in \{r\}, r_k \in \partial s_k \quad (3)$$

It is notable that the equation (3) extends the clique definition to multiresolution mode. Intensity average and segmentation label functions are also extended to vector space. The intensity of pixels at different channels in set $\{s\}$ form a vector $Y(\{s\}) = [Y_1(\{S\}), Y_2(\{S\}), Y_3(\{S\})]$, and similarly, $\mu(\{s\}) = [\mu_1(\{S\}), \mu_2(\{S\}), \mu_3(\{S\})]$ is the mean vector. Therefore the objective function is extended to vector space as follows

$$E(X) = \sum_{\{S\}} \{\sum_{i=1}^{3} ||Y_i(\{s\}) - \mu_{X(\{s\})}^{i}(\{s\})||^2 +$$

$$\sum_{\{r\} \in \partial\{s\}} V_c(\{s\}, \{r\})\} \quad (4)$$

The outer summation is over vectors, while the first inner summation is related to the distances of the pixel's intensities from the estimated average for each channel of color images. The second inner summation is over all neighborhood vectors of vector $\{s\}$. The pixels of $\{s\}$ are corresponding according to the wavelet dawnsampeling and $\{r\}$ is a vector neighbors of $\{s\}$ which means that pixels of $\{s\}$ and $\{r\}$ located at the same resolution are also neighbors.

3.2. Smoothness criterion

Traditionally, in region-based image/video segmentation algorithms, emphasis is put on the accuracy of segmentation. However objects/regions shape delineation, and producing a well-pleasing objects'/regions' shape has not attracted enough attention due to the ill-posed problem nature of segmentation and non existing an ideal segmentation algorithm for the segmentation task [9]. In contour/edge-based segmentation algorithms, another important criterion, related to the appearance of the extracted objects/regions, has been considered. In these algorithms, the extracted objects/regions borders are smoothed [10]. Ideally, borders are edges in the image, which are one of the most important properties for visual perception. Because most natural objects exhibit smooth edges and distortions such as downsampling often creates rough edges, there is a correlation between border smoothness and visually pleasing objects. Therefore border smoothness terms corresponding to different resolutions have been added to the objective function to contribute in our MRF-based segmentation approach.

The proposed smoothness definition is based on the border's curvature. In a digital environment an estimation of curvature is used [11]. Minimizing the proposed estimation of smoothness prevents unwanted fluctuations in the border pixels.

A large number of pixels ensures border smoothness at high resolutions, however, at lower resolutions the visual
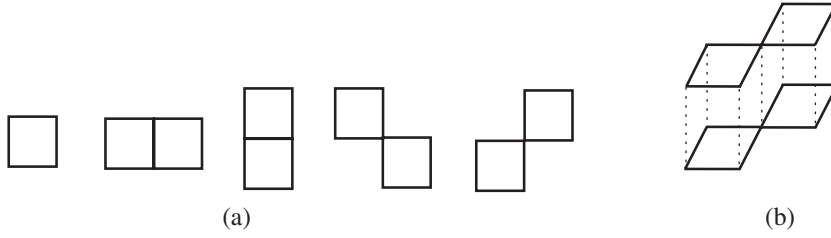
Fig. 2. (a) Normal one and two pixels cliques sets. (b) A clique of two vectors with the vectors' dimension equal to two.

quality can suffer due to downsampling distortion and lack of sufficient information. To reduce this effect, the smoothness could be enhanced at lower resolutions more rigorously than higher resolutions. The priority is realized by bigger coefficients for lower resolution smoothness. Therefore the objective function is updated according to the following equation:

$$E(X) = \sum_{\{S\}} \{ \sum_{i=1}^{3} ||Y_i(\{s\}) - \mu^i_{X(\{s\})}(\{s\})||^2 +$$

$$\sum_{\{r\} \in \partial\{s\}} V_c(\{s\}, \{r\}) + \sum_{q \in \{s\}} l_{res(q)}.\nu(q) \} \quad (5)$$

where $\nu(q)$ shows the curvature estimation of pixel q (which is a pixel of vector $\{s\}$), and $l_{res(q)}$ is a coefficient which is resolution dependent.

The proposed smooth object extraction is different from a simple objects' border smoothness as has been done in [12] which is a filtering of the extracted video object shape to remove the small elongation introduced during the segmentation process, in the following areas. (1) Our smoothing process takes part in the segmentation algorithm and changes the segmentation outcome. (2) With sufficient contrast, the proposed algorithm produces borders that are more faithful to the regions shape. (3) On some occasions, some background pixels are added to the foreground regions to produce better looking shapes especially at different resolution. (4) The smoothness factor could be adjusted for different resolutions to produce visually pleasing shapes at different resolutions with scalability as a constraint.

3.3. MAP estimation

The segmentation is initialized with the k-means clustering algorithm for each channel separately. Then neighboring pixels with equal labels at all three channels form a region. The segmentation estimation is improved using ICM optimization [13]. The objective function term corresponding to the current vector is optimized given the segmentation at all other vectors of the pyramid. The resulting objective function terms related to the current vector is:

$$E(X\{s\}) = \sum_{i=1}^{3} ||Y_i(\{s\}) - \mu^i_{X(\{s\})}(\{s\})||^2 +$$

$$\sum_{\{r\} \in \partial\{s\}} V_c(\{s\}, \{r\}) + \sum_{q \in \{s\}} l_{res(q)}.\nu(q) \quad (6)$$

For YUV color images where U and V are in half resolution, we delete the terms corresponding to U and V only at full resolution, and keep the other terms in the above equation. For grey-level images only the intensity channel or Y_1 in the above formula is kept.

4. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, experimental results obtained using the algorithm introduced in Section 3 are presented. The results are compared with regular single-level and multiresolution segmentation algorithms [1], [6]. Due to limited space, more examples and discussion will be presented in the conference.

As a first example, the proposed algorithm is tested using frame 5 of the SIF sequence Table Tennis. Figure 3 represent the results achieved by the proposed multiresolution scalable, regular single and multiresolution segmentation algorithms. The result of the proposed scalable multiresolution segmentation algorithm is presented only at the finest resolution, because the lower resolutions results have the same patterns and figures.

In the proposed algorithm the effects of high resolutions on low-resolution decreases under segmentation of multi resolution segmentation algorithms and results in detection of objects/regions, which are not detected using regular multiresolution segmentation. In other words, the sensitivity to grey-level changes is increased, resulting in a better detection of small or low-contrast objects especially in low resolutions. The number of detected regions at three resolutions 60×120, 120×176 and 240×352 are 19, 55 and 164 in regular multiresolution while it is increased to 42, 83 and 184 by the proposed algorithm 19, 68 and 314 by single resolution segmentation algorithm [1]. The proposed scalable segmentation detects more relevant regions than the regular multiresolution algorithm. For example, consider the segmentation of the textured wall and detection of the ball in the Table Tennis image as presented in Figure 3. The single-level segmentation detects the ball, but it also detects a number of spurious regions due to the textured background as the number of regions shows. This drawback is called over-
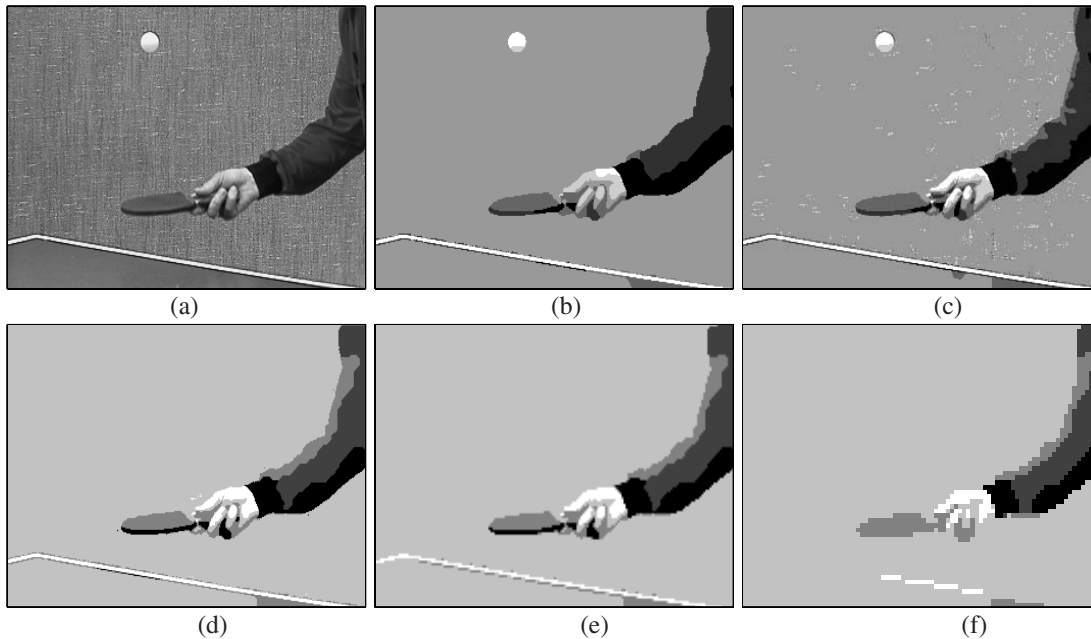
Fig. 3. Table tennis image segmentation with $k = 6$ clusters and $\beta = 100$; (a) the main image; (b) segmentation by the proposed scalable algorithm at 240×352; (c) regular single level segmentation; (d) regular multiresolution segmentation 240×352; (e) 120×176 segmentation; (f) 60×88 segmentation;

segmentation. The regular multiresolution algorithm misses the ball at different resolutions altogether. The proposed algorithm, however, detects the ball as well as avoiding unsightly segmentation of the textured background.

While our algorithm has improved sensitivity to grey-level variation it still maintains noise tolerance. To test the scalable segmentation algorithm on noisy images, first a uniform noise in the range $(-30, +30)$ was added to the Table Tennis images, and then different segmentation algorithms were performed. The number of misclassified pixels for the the entire image pixels were counted. It was $\%7.96$, $\%6.20$ and $\%18.74$ for the regular multiresolution, scalable and single resolution segmentation algorithms. The results confirm that the proposed algorithm can deal with noisy images as effectively as multiresolution segmentation and much better than single-level segmentation.

One of the applications of the proposed segmentation algorithm is in object extraction for object-based image/video processing. In many of the semi automatic video object extraction algorithms, in the first frame, the object of interest is determined by user intervention and is tracked in the subsequent frames. To facilitate the first frame selection, the user can 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 the selected regions and deleting the small object's regions creates the final object. As an example, in the first frame of the Table tennis sequence, after user rough determination of interested object, the exact

borders of the object at different resolutions are determined. The extracted objects by different segmentation algorithms at finest resolutions are shown in Figure 4. Comparison of the extracted objects confirms the superiority of the scalable segmentation algorithm in a subjective test. As a complementary step, using a quantitative test, the border smoothness is measured at all pyramids scales. The average curvature of object's border is decreased from about 23.68 to 16.5 which is about $\%35$ improvement in the object smoothness.

In the second example, frame 34 of the Mother and Daughter sequence was segmented. The image is in qcif format and is given in the YUV color space. Since U and V are in half resolution, in the first solution, the image was segmented in grey-level space by a single-resolution image segmentation algorithm [1]. The result is shown in Figure 5(b). The left area of the daughter's face has not been well separated from the background because there is not enough grey-level contrast between face and background. The same shortcoming happens for the other grey-level segmentation algorithms except when there is oversegmentation with a large number of detected regions, which is not desired for segmentation applications. To successfully separate object's regions from background, color segmentation was performed as an alternative solution. The proposed scalable segmentation algorithm can perform color segmentation using half-resolution chrominance components. The result is shown in Figure 5(d). The number of regions in grey-level segmentation is 273 while in color segmentation it is 112, which shows a reasonable color image segmentation algorithm.
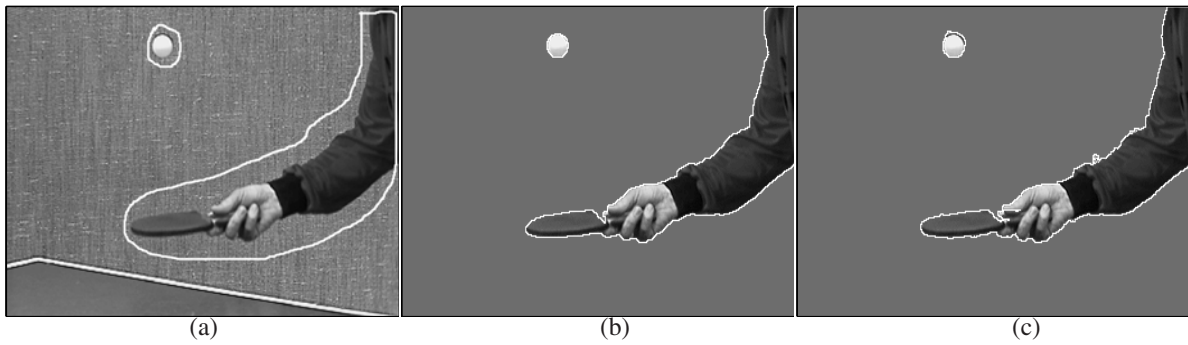
Fig. 4. Table Tennis object extraction at 240×352; (a) User object selection; (b) scalable; (c) single resolution segmentation;

Fig. 5. Frame 34 of qcif size Mother and daughter Image sequence segmentation with $\beta = 40$; (a) original grey-level image; (b) regular grey-level single-resolution segmentation with $k = 7$ clusters; (c) color image where U and V are in half resolution; (d) Proposed scalable segmentation with $k = 7, 2, 2$ clusters.

5. CONCLUSIONS

We have presented a multiresolution scalable image segmentation algorithm which extracts regions with similar segmentation pattern at different resolutions, The proposed segmentation is useful for object-based wavelet coding applications. In addition to scalability, a new quantitative criterion is added to the segmentation algorithm. This criterion, a smoothness function based on the segmentation labels, represents the visual quality of the objects/regions at different resolutions. To reduce the downsampling distortion, different smoothness coefficients are considered for different resolutions. The multi scale analysis improves the sensitivity to grey/color variations while maintaining high performance in noisy environments. The algorithm has flexibility to segment YUV color images where U and V are in half resolution.

REFERENCES

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

[2] L.Zheng, J.C.Liu, A.K.Chan, and W.Smith, "Object-based image segmentation using dwt/rdwt multiresolution markov random field," in *Acoustics, Speech, and Signal Processing, 1999. ICASSP '99. Proceedings., 1999 IEEE International Conference on*, vol. 6, pp. 3485–3488.

[3] M.L. Comer and E.J.Delp, "Segmentation of textured images using a multiresolution gaussian autoregressive model," *Image Processing, IEEE Transactions on*, vol. 8, no. 3, pp. 408–420, 1999.

[4] Z.Kato, J.Zerubia, and M.Berthod, "Unsupervised parallel image classification using markovian models," *Pattern Recognition*, vol. 32, 1999.

[5] R.Wilson and C.T.Li, "A class of discrete multiresolution random fields and its application to image segmentation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, 2003.

[6] M.M. Chang, M.I. Sezan, and A.M. Tekalp, "Adaptive bayesian segmentation of color images," *Journal of Electronic Imaging*, vol. 3, no. 4, pp. 404–414, 1994.

[7] J. Gao, J. Zhang, and M.G. Fleming, "A novel multiresolution color image segmentation technique and its application to dermatoscopic image segmentation," in *Image Processing, 2000. Proceedings. 2000 International Conference on*, 1997, vol. 3, pp. 408–411.

[8] 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.

[9] M.Bertero, T.A.Poggio, and V.Torre, "Ill-posed problems in early vision," *Proceedings of the IEEE*, vol. 76, no. 8, pp. 869–889, 1988.

[10] J.Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–998, 1986.

[11] M.Worring and A.W.M Smeulders, "Digital curvature estimation," *CVGIP-Image Understanding*, vol. 58, 1993.

[12] F.Marques and J.Llach, "Tracking of generic objects for video object generation," in *International Conference on Image Processing (ICIP)*, 1998, vol. 3, pp. 628–632.

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