# Scalable Multiresolution Image Segmentation and Its Application in Video Object Extraction Algorithm

Fardin Akhlaghian Tab, Golshah Naghdy
School of Electrical, Computer and Telecommunications Engineering
University of Wollongong
Wollongong, NSW 2522, Australia
Email: {fat98, golshah }@uow.edu.au

Alfred Mertins
Signal Processing Group, Institute of Physics
University of Oldenburg,
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 suitable for the 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. 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. Application of the spatial segmentation in video segmentation, compared to traditional image/video object extraction algorithms, produces more visually pleasing shape masks at different resolutions which is applicable for object-based video wavelet coding. Moreover it allows for larger motion, better noise tolerance and less computational complexity. In addition to spatial scalability, the proposed algorithm outperforms the standard image/video segmentation algorithms, in both objective and subjective tests.**

## I. 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. However the traditional multiresolution image segmentation algorithms extract objects'/regions' shape hierarchically from lower to higher resolutions, and the final objects'/regions' shape are obtained at highest resolution. It means that the lower resolutions segmentation maps to some extent are different with the higher resolutions segmentation map. In other words, the highest resolution segmentation map is more precise than the other resolutions.

In this paper we present a MRF-based multiresolution image segmentation algorithm. It produces similar segmentation maps at different resolutions which are applicable to object-based wavelet coding algorithms. We call the multiresolution segmentation algorithm with similar patterns at different resolutions as scalable segmentation.

The multi scale 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 objects/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 down sampling often result in rough borders/edges, in this work, a multiresolution smoothness criterion is incorporated in the objective function of the segmentation algorithm which results in more natural or visually pleasing objects/regions. By considering bigger smoothness coefficients for the smoothness terms of lower resolutions, the distortion effect of down sampling is reduced and the extracted objects/regions are more visually pleasing.

Extending the scalable image coding to video, in the scalable object-based video coding, it is necessary to extract object's shape at different resolutions. One regularly informally used option is the single level video segmentation where objects in fine resolution are extracted and then down sampled according to the existing relationship between shapes at different resolutions determined by the wavelet filter used [1]. However down sampling distort shapes and cannot preserve topology at lower resolutions for all possible shapes [2]. In other words, a visually pleasing object at higher resolution does not necessarily ensure similar quality at lower resolutions. For example in Figure 1, down sampling of two digital circles are compared where pixels with even indexes are down sampled to lower resolution. It can be seen that better approximation of a digital circle at high resolution results in

worse down sampled circle shape.

As an application of the proposed scalable image segmentation algorithm we present a multiresolution video segmentation algorithm which extracts visually pleasing video object at different scales. The proposed method is noise tolerant, computational simple and allows of larger motion.

## II. 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), spatial and temporal scalabilities. 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 low pass (LL) sub band. In this paper we use an odd length wavelet filter (e.g. 9/7), where all shape pixels with even indices[1] are down sampled for the low pass band [1]. 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 next 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 all the higher resolutions and pixels with indices that are multiples of $2^n$ in both dimensions are down sampled to the 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. In Fig 2 an example of down sampling is shown.

## III. Spatial Segmentation Algorithm

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 modeling is selected as it includes low level processing at pixel level and has enough flexibility in defining objective functions according

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

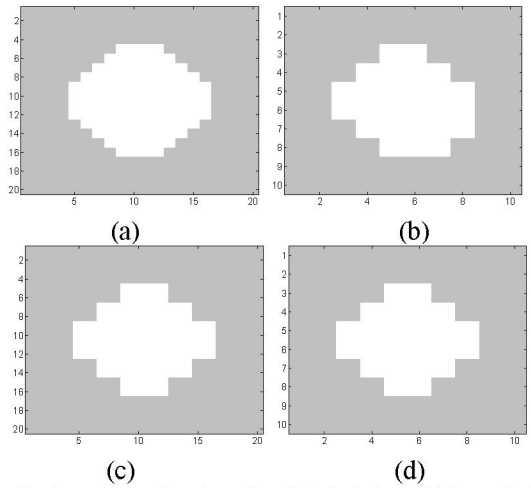



Fig. 1. (a) closer approximation of a digital circle at high resolution; (b) down sampling to lower resolution; (c) worse approximation of a digital circle at high resolution; (d) down sampling to lower resolution.

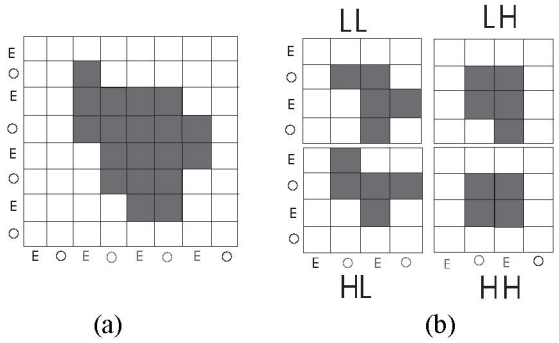



Fig. 2. Decomposition of a non rectangular object with odd-length filters; (a) the object, shown in dark gray; (b) decomposed object after filtering. The letters "E" and "O" indicate the position(even or odd) of a pixel.

to 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 [3], [4] and then extend it to the scalable multiresolution segmentation mode. The desired segmentation is denoted by $X$, and $Y$ is the observed intensity image. This results in the following cost or objective function which has to be minimized with respect to $X(s)$ [4]:

$$E(X) = \sum_s \left( (Y(s) - \mu_{X(s)}(s))^2 + \frac{1}{T} \sum_{r \in \partial_s} V_c(s,r) \right) \quad (1)$$

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

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

Herein $\beta$ 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 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 $Y$ be the intensity 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 is 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 its 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 3(a) shows regular one and two pixels clique sets. In Figure 3(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:

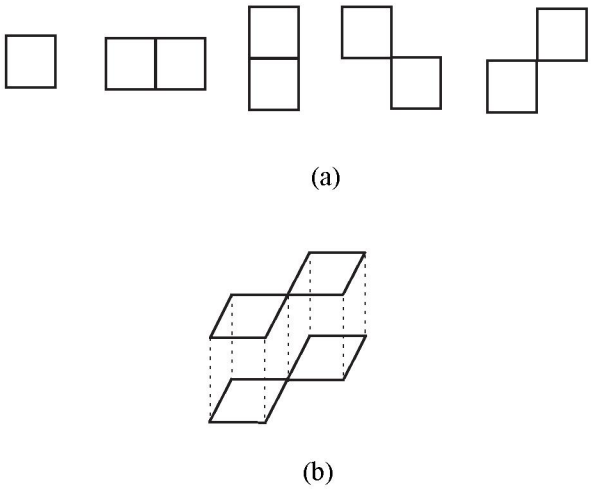
$$V_c(\{s\},\{r\}) = (\frac{1}{N}) \sum_{k=M}^{M+N-1} (-1)^{L_k}.\beta \,,$$

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

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 $\beta$, so that adjacent pixels, of two neighboring vectors, located at the same resolution are more likely to belong to the same region than to different regions.

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. Therefore the objective function is extended to vector space as follows:

$$E(X) = \sum_{\{S\}} ||Y(\{s\}) - \mu_{X(\{s\})}(\{s\})||^2 + \sum_{\{r\}\in\partial\{s\}} V_c(\{s\},\{r\}) \tag{4}$$



(a)

(b)

Fig. 3. (a) Normal one and two pixels cliques sets. (b) A clique of two vectors with the vectors' dimension equal to two.

The intensity of vector of pixels is shown by the $Y(\{s\})$ and $\mu(\{s\})$ is the mean vector. 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 dawn sampling and $\{r\}$ is a vector neighbors of $\{s\}$.

### *A. 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 nature of segmentation problem. 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 [5]. 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 down sampling often creates rough edges, there is a correlation between border smoothness and visual quality. Therefore border smoothness terms corresponding to different resolutions have been added to the objective function to enhance 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 [6]. Minimizing the proposed estimation of smoothness prevents unwanted fluctuations in the border pixels.

To enhance border smoothness in lower resolutions, bigger coefficients are allocated to lower resolutions smoothness. Therefore the objective function is updated according to the following equation:

$$E(X) = \sum_{\{S\}} \{||Y(\{s\}) - \mu_{X(\{s\})}(\{s\})||^2 +$$

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

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 takes part in the segmentation algorithm and changes the segmentation outcome.

Finally the Iterated Condition Mode (ICM) optimization method [3] is used to minimize the objective function at equation 5 which classifies each pixel of the pyramid to obtain the segmentation of the image pyramid.

## IV. Video object Extraction

At the core of most video segmentation algorithm routines is a tracking algorithm. In the backward tracking algorithm the spatial segmentation gives the precise borders of object(s). This also overcomes the problems of non rigid moving objects and uncovered background. Therefore we have proposed a multiresolution backward tracking algorithm.

In the first frame through user's intervention and spatial segmentation, meaningful objects are determined. In the subsequent frames, the object is tracked by an automatic procedure. The scalable multiresolution intra frame segmentation is performed as mentioned in section III. Scalable segmentation ensures similar segmentation patterns at different resolutions [7]. We have used this feature in our proposed tracking algorithm to track some regions in the proper resolution and extend the results to corresponding regions at other resolutions. Regions classification starts from lowest level of decomposition. Regions bigger than a threshold, are processed in this resolution and small size regions are processed in higher resolutions. Each processed region is divided into morphological catchment basins and each watershed basin is classified into object or background. This overcomes the probable short comings of spatial segmentation to separate the entire object from the background. Motion estimation provides information for the backward projection of each basin.

In any projected region, the percentage of pixels projected to the object area is shown by $OPR$. If $OPR$ be more than a threshold, such as $65\%$ of the region's size, it is classified as object. Similarly, for $OPR$ less than a threshold such as $\%35$ it is classified as background. For $OPR$ in the range between $35\%$ to $65\%$ we also consider the smoothness of the shape $\triangle SMT$. If addition of this basin significantly increase the smoothness of the shape, the regions is classified as object region. Therefore we define the following equation

$$S = OPR + \alpha_2 . \triangle SMT,$$

$$\triangle SMT = (SMT2 - SMT1)/MAX(SMT1,\ SMT2)$$

Where $SMT2$ is the smoothness of the object area after adding the basin to the area and $SMT1$ is the shape smoothness when the basin is classified as background[2]. Therefore $SMT$ is the relative increase of the smoothness of the shape after addition of the processed region to the object area compare to when that it is classified as background. The $SMT$ value is between $(-1, 1)$ and final classification of the regions is based on the percent of the pixels projected to object region and smoothness increase/decraese. $\alpha_2$ is a coefficient selected to values such as $0.3$. In Figure 4 the object smoothness could be seen.

[2]The object area smoothness is the mean smoothness of it's border pixels

## V. Experimental Results and discussion

To evaluate the performance of the proposed algorithm, a still image and two different image sequences Clair with CIF format and Table Tennis with SIF format are segmented.

In the first example the tracking algorithm is run over the $75$ frames of the Clair sequence. The extracted objects at frames number $20$, $40$ and $60$ for different resolutions are shown in the Figures 5(a), (b) and (c).

To compare the proposed algorithm with other region based object tracking and extraction, we have used similar tracking algorithm but in single resolution mode which includes regular single level spatial segmentation [3] and tracking only at the finest resolution. To ensure similarity to the existing region based tracking algorithms, which are often morphological based [8], the object areas are extended to fill the morphological catchment basins regions which overlap with the extracted object. The qualitative criterion for comparison is border smoothness of the extracted objects. Object smoothness is averaged over the curvature of border pixels. Although it is not an ideal criterion, but in our experiments it has confirmed performance of our subjective tests. The smoothness comparison for the $75$ frames of the Clair sequence for the $3$ resolution levels are shown in Table I. The smoothness term affects the segmentation in areas of the image that have lower grey level contrast. In the Clair sequences the regions around the head have lower contrast compare to shoulder and body areas. If we only consider the head parts the smoothness improves by $13.17\%$, $11.5\%$ and $10.5\%$ at different resolutions. As a qualitative test, look at the extracted objects of the $23th$ frame of Clair sequence by scalable and regular algorithm in Figure 6. In this Figure, images of different resolutions are shown at the same size to highlight the details . Analyzing both images, shows that our algorithm has extracted the Clair object smoother and more visually pleasing. It looks as if our algorithm has done a nice hair cut to Clair.

As the second example we have processed the standard MPEG-4 Table Tennis sequence, which has textured background with fast moving objects. Frame numbers $10$, $20$ and $32$ with the extracted objects are shown in Figure 7. As an example, the extracted objects in frame number $10$ of table tennis sequence by the single level tracking algorithm

TABLE I

Clair Sequence Smoothness.

| | $88 \times 72$ | $144 \times 176$ | $288 \times 352$ |
|---|---|---|---|
| Scalable Tracking | 54.67 | 54.7 | 53.15 |
| Regular Tracking | 58.95 | 58 | 56.87 |
| improvement | %7.54 | %6.03 | %6.77 |

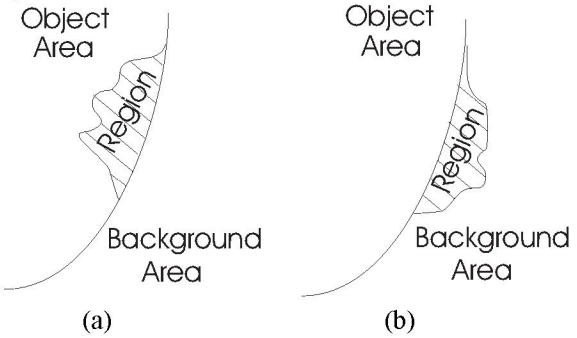
(a) (b) (c)

Fig. 5. Clair sequence tracking; (a) Object extracted in frame number 20; (b) Object extracted in frame number 45; (c) Object extracted in frame number 65;

($a_1$) ($b_1$) ($c_1$)

($a_2$) ($b_2$) ($c_2$)

Fig. 6. Clair object of $23th$ frame; ($a_1$) scalable $288 \times 352$; ($b_1$) scalable $144 \times 176$; ($c_1$) scalable $72 \times 88$; ($a_2$) regular $288 \times 352$; ($b_2$) regular $144 \times 176$; ($c_2$) regular $72 \times 88$;




(a) (b)

Fig. 4. The processed region is shaded. (a) Adding this region to object area increases the smoothness; (b) Adding this region to object area decreases the smoothness.

presented by [9] at 3 different resolutions are shown in Figure 8. For a quantitative comparison we have measured the objects smoothness and the improvements are about 7% in different resolutions. Again if we only consider the hand and fingers with the racket the smoothness improvements are nearly doubled. Also the computational complexity of multiresolution tracking algorithm is reduced to less than 30% of single resolution object tracking.

## VI. Conclusion

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. As well as 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 down sampling distortion, different smoothness coefficients are considered for different resolutions. The multi scale analysis improves the sensitivity to intensity variations


Fig. 7. Table Tennis object extraction; (a) frame 10; (b) frame 23; (c) frame 32;

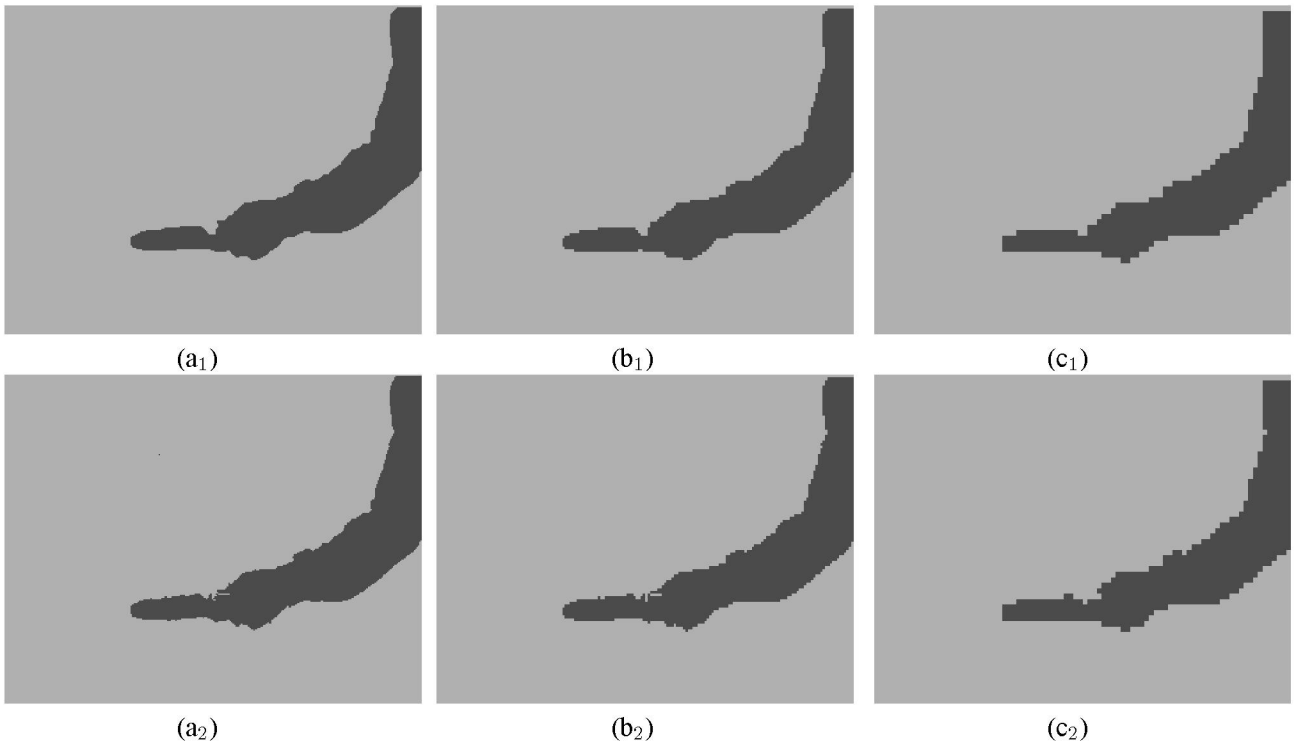



Fig. 8. Table Tennis object $10th$ frame; ($a_1$) scalable $240 \times 352$; ($b_1$) scalable $120 \times 176$; ($c_1$) scalable $60 \times 88$; ($a_2$) regular $240 \times 352$; ($b_2$) regular $120 \times 176$; ($c_2$) regular $60 \times 88$;

while maintaining high performance in noisy environments. The image segmentation algorithm is useful for multiresolution video object extraction algorithms. The extracted objects are visually pleasing and quantitatively smoother than objects detected through regular region based object extraction algorithms. The multiresolution algorithm has less computational complexity and performs well with noisy environments.

## REFERENCES

[1] 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.

[2] G.Borgefors, G.Ramella, G.Sanniti di Baja, and S.Svenson, "On the multiscale representation of 2d and 3d shapes," *Graphical Models and Image Processing*, vol. 61, no. 1, pp. 44–62, 1999.

[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] 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.

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

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

[7] F. Akhlaghian Tab, G. Naghdy, and A. Mertins, "Multi resolution image segmentation with border smoothness for scalable object-based wavelet coding," in *Proc. DICTA*, Sydney, Australia, 2003, pp. 977–986.

[8] Y.Tsaig and A.Averbuch, "Automatic segmentation of moving objects in video sequences: a region labeling approach," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 12, no. 7, pp. 597–612, 2002.

[9] M. Kim, J. Choi, D. Kim, H. Lee, M. Lee, C. Ahn, and Y. Ho, "A vop generation tool: automatic segmentation of moving objects in image sequences based on spatio-temporal information," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 9, no. 8, pp. 144–50, 1999.