行政院國家科學委員會專題研究計畫　成果報告

應用 MFRSST 之方法發展視訊分割系統

計畫類別：個別型計畫
計畫編號：NSC92-2213-E-216-024-
執行期間：92 年 08 月 01 日至 93 年 07 月 31 日
執行單位：中華大學資訊工程學系

計畫主持人：連振昌

報告類型：精簡報告

處理方式：本計畫可公開查詢

中　華　民　國 93 年 9 月 13 日

報告內容

一、中英文摘要

近年來，產生了許多視訊多媒體方面新的應用，例如影片快速瀏覽、視訊內容之搜尋、視訊內容之索引、視訊內容之摘要以及視訊內容之檢索。而目前在訂定中之視訊壓縮標準 MPEG-7 即是為了上述的各項視訊多媒體方面的應用而發展，其原理是利用視訊影像中之顏色、紋理、形狀和運動的描述來發展之。基於這些較低層次之影像特徵，我們可由切割後之視訊片段(video shot)中擷視訊中之物件，更進而分類這些物件。藉著分析這些物件之間的發生時序上與空間上的相互關係，我們可以用來辯識出視訊影片中所發生之某事件。目前大部份之視訊內容之搜尋與檢索方面的研究可分成三大類：1)視訊畫面之結構化(syntactic structurization of video)；2)視訊分類(video classification); 3) 視訊內容意義之擷取(extraction of semantics)。現今大部份之研究大概都集中於前二類，本計劃著眼的研究方向在於第三類。藉著擷取視訊內容之意義，我們計劃發展場景事件識別系統，此系統可應用於一些智慧型自動化與監控的系統。場景事件識別系統包含四個子系統，分別為：1) 視訊切割技術(Video Segmentation); 2)視訊畫面分割技術(Video Shot Boundary Detection); 3)視訊畫面內之物件分類技術(Video Object Extraction and Classification); 4)場景事件偵測與分析技術(scene event identification)。

關鍵詞：視訊瀏覽, 視訊索引, 視訊摘要, 視訊檢索, 場景事件識別

Abstract

In the recent years, a lot of research activities are addressed on the video segmentation problems for developing the object-based video processing systems. Based on the extension of the image segmentation methods, the video segmentation systems may be constructed for obtaining the video objects and then developing the object-based video processing systems. However, the conventional image segmentation methods have some problems that make the video segmentation process inaccurate, i.e., they can’t ensure the region connectivity, define the region boundaries accurately, and partition image with an arbitrary number. In this project, the concepts of multiresolution decomposition and flooding process of watershed algorithm are applied to develop the multiresolution and flooding based RSST (MFRSST) image segmentation method that can speed up the image segmentation process and have the same partition quality as the RSST. In the MFRSST segmentation method, the $N + M$ regions are partitioned by the two-step segmentation processes. Firstly, a lower resolution subband image is obtained by the multiresolution decomposition process and then the subband image is partitioned into N regions by using the flooding and FRSST methods. Secondly, the region boundary refinement process is applied to obtain the additional M detailed regions. Based on the MFRSST method, we may improve the COST211 video segmentation system by applying the optical flow motion detecting method such that the rotation and zooming movements of objects can be detected.

Keywords: COST 211 video segmentation system, Video objects, RSST, MFRSST.

二、緣由與目的

Recently, with the great advance in computing and computer network, many new emerging multimedia applications, such as low bit rate video coding, video summarization, video indexing, and video query, enforce the development of new video coding standards: MPEG-4 [1] and MPEG-7 [2]. The objects in both the video coding systems are acquired by applying the video segmentation process, which use the image features such as color, intensity, motion, textures, and shapes to partition the regions of objects. Once the video objects are acquired, the aforementioned multimedia applications can be fulfilled. Hence, the quality and efficiency of image segmentation process determine the performance of the object-based video processing system.

However, the conventional image segmentation methods, such as the histogram-based algorithms [3,4] and split-merge algorithms [5,6], have some problems that make the video segmentation process inaccurate. Firstly, the histogram-based segmentation methods can't ensure the region connectivity. Secondly, the split-merge methods can't define the region boundaries accurately. Finally, both methods are unable to partition the image with an arbitrary number. The method of recursive shortest spanning tree (RSST) [7] is proposed to solve these problems by using the global information and is widely adopted in many video segmentation systems [8-14]. Alatan *et al*. [10, 11] develop the European COST 211 framework by utilizing the RSST algorithm to partition the regions with the same color or motion homogeneity. Tuncle *et al.* [12] utilize the RSST algorithm and affine motion model to develop their video object segmentation system. Doulamis *et al.* [13] propose an efficient RSST algorithm for non-sequential video content representation. Vlachos *et al.* [14] extend the RSST algorithm to partition color images by introducing red-green-blue (RGB) components into the cost function.

The high computation complexity is the main drawback of the RSST algorithm. The sorting procedure is required whenever the two vertices of the least weighted link are merged. Kwork *et al.* [15] propose the fast RSST (FRSST) algorithm to improve the efficiency of the RSST algorithm. Two improvements that make the FRSST algorithm efficient are: (1) the sorting process is removed and (2) the region-growing method is applied.

However, the FRSST method is still too inefficient to meet the requirements for some new visual applications such as video segmentation, region-based video coding, and object-based video retrieval. In order to reduce the computation cost further, the concepts of two image processing methods are adopted to develop the new image segmentation algorithm. The first one is the flooding method in the watershed algorithm [16]. Instead of merging the two vertices of the least weighted link, we merge all the neighboring vertices whose link weights are closed to the lowest weight concurrently. Such a merging process is similar to the flooding method in the watershed algorithm that fills up the image valleys with a specified height. The second one is the concept of multiresolution decomposition. Performing the partition process in the decimated image can reduce the computation cost significantly. However, the interpolation process may introduce the blocky effect in the region boundary. Therefore, the boundary refining process is used to define the region boundary accurately. In this project, the aforementioned two concepts are applied to develop the multiresolution and flooding based RSST (MFRSST) image segmentation method that can speed up the segmentation process.

Furthermore, we improve the COST211 video segmentation system by applying the optical flow motion detection method such that the rotation and zooming movements of objects can be detected. In addition, two new detection rules are proposed such that the regions of zooming and rotating objects can extracted. The new detection rules are established by utilizing the characteristics of rotating and zooming movements. For example, if the object doesn't move but has the motion information, then it must be rotating. In the new proposed segmentation system, the YIQ color signals are used for color segmentation and the optical flow motion information is used for motion segmentation. It is important to define the link values for the color and motion in the segmentation process, because the definition of the link values will affect the quality of segmentation.

三、結果與討論

In section 3.1, the quality and efficiency analyses for the RSST, FRSST, and MFRSST algorithms are illustrated. In section 3.2, the experimental results of the proposed video segmentation system are shown.

3.1 The Experimental Results for the MFRSST algorithm

3.1.1 An Example of MFRSST Segmentation Process

First of all, we illustrate an example of MFRSST segmentation process. Fig. 1-(a) shows the Lena image with size 128×128 pixels and Fig. 1-(b) shows the 98 segmented regions in decimated image. Based on the segmented regions in Fig. 1-(b), the reconstruction process mentioned in section 3.2 is applied to obtain the accurate segmented regions with the original resolution. Fig. 1-(c) shows the watershed-like image after the 2-D interpolation process and Fig. 1-(d) shows the accurate segmented regions after boundary refining process.

3.1.2 The Quality and Efficiency Analyses for the RSST, FRSST, and MFRSST Segmentation Algorithms

Fig. 2 illustrates that Lean image are partitioned into 300 regions using the RSST, FRSST, and MFRSST algorithms. The experimental results show that the three algorithms may define the region boundary accurately. However, less spark regions shown in Fig. 2-(c) imply that more complete regions may be acquired by using the MFRSST method. The computation time of the three methods for partitioning various size of image is shown in Fig. 3. The experimental results show that the MFRSST is faster than the other algorithms. The reason why the MFRSST method may improve the efficiency of the FRSST algorithm is that the partition process is performed in the decimated image.

3.2 The Experimental Results of the Improved Video Segmentation System

In the COST211 segmentation system, two modes of operations are established. In the first mode, a binary mask that distinguishes the moving objects (foreground objects) from the static background is generated. In the second mode, the movements of objects are further classified into several motion types. Here, additional two new rules are proposed to detect the rotation and zooming movements of objects.

3.2.1 Mode I — Foreground and Background Separation

By applying the video segmentation method mentioned in section 4.5.1, the regions of foreground (moving objects) and background are partitioned and shown in Fig. 4.

3.2.2 Mode II — Detection of the various kinds of Moving Objects

In this section, some experimental results of detecting the various kinds of moving objects are presented. All the segmentation processes are performed by using the MFRSST method.

(a) Object tracking. In Fig. 5-(a) and (b), the 15^{th} and 17^{th} frames of Akiyo image sequence are shown respectively. There are 310 color regions (listed in Table 1) for the above two successive frames are partitioned and shown in Fig. 5-(c). The motion segmented regions of

R_t^M and R_t^{MC} are shown in Fig. 5-(d) and (e). By applying the rule 1 of mode II the final segmentation result R_t^F is obtained and shown in Fig. 5-(f). Furthermore, the color segmented regions projected onto the motion and the motion compensation masks are illustrated in Table 1. From Table 1, we know that the set of moving regions is the same as the set of previous moving object, i.e., the moving object is being tracked (rule 1 of mode II).

(b) Detection of newly exposed objects. In Fig. 6-(a) and (b), the 25th and 27th frames of Mother & Daughter image sequence are shown respectively. There are 273 color regions (listed in Table 2) are partitioned and shown in Fig. 6-(c). The motion segmented regions of R_t^M and R_t^{MC} are shown in Fig. 6-(d) and (e). By applying the rule 4 of mode II, the final segmentation result R_t^F is obtained and shown in Fig. 6-(f). Furthermore, the color segmented regions projected onto the motion and the motion compensation masks are illustrated in Table 2. From Table 2, we know that the set 1 of moving regions is the same as the set of previous moving object, i.e., the moving object is being tracked (rule 1 of the mode II). Furthermore, the moving regions in set 2 are detected as the newly exposed object by the rule 4 of mode II.

(c) Detection of the zooming objects. In Fig. 7-(a) and (b), the 5th and 7th frames of basketball image sequence are shown respectively. There are 108 color regions (listed in Table 3) are partitioned and shown in Fig. 7-(c). The motion segmented regions of R_t^M and R_t^{MC} are shown in Fig. 7-(d) and (e). By applying the rule 3 of mode II, the final segmented region R_t^F is obtained and shown in Fig. 7-(f). The color segmented regions projected onto the motion and the motion compensation mask are illustrated in Table 3. From Table 3, we know that the set of moving region is more than the set of the previous moving object, i.e., the object is being zoomed (rule 3 of mode II).

(d) Detection of rotating objects. In Fig. 8-(a) and (b), the 64th and 65th frames of image sequence of rotating globe are shown respectively. There are 260 color regions (listed in Table 4) are partitioned and shown in Fig. 8-(c). The motion segmented regions of R_t^M and R_t^{MC} are shown in Fig. 8-(d) and (e). By applying the rule 2 of mode II, the final segmentation result R_t^F is obtained and shown in Fig. 8-(f). The color segmented regions projected onto the motion and the motion compensation masks are illustrated in Table 4. From Table 4, we know that the set of moving region isn't the same as the set of previous moving object, but the R_{t-1}^F is the same as the current object, i.e., the moving object is being rotated (rule 2 of the mode II).

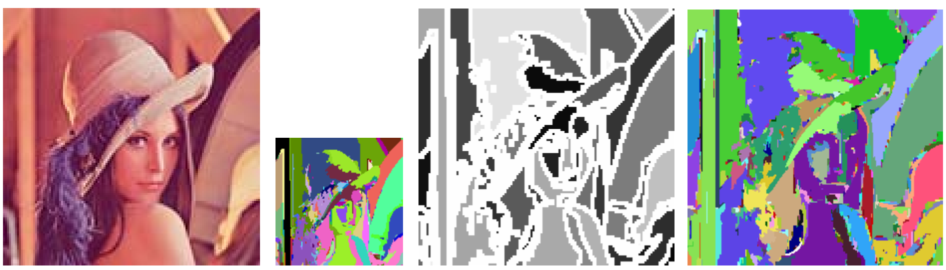

(a) (b) (c) (d)

Fig. 1. (a) The Lena image with size 128×128 pixels. (b) There are 98 partitioned regions in the decimated image with size 64×64 pixels. (c) The interpolated image. (d) There are 200 regions that are partitioned after the boundary refining process.

(a) (b) (c)

Fig. 2. The Lena image is partitioned into 300 regions using (a) RSST, (b) FRSST, and (c) MFRSST algorithms.

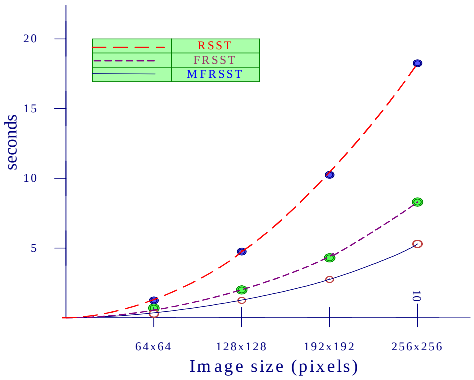
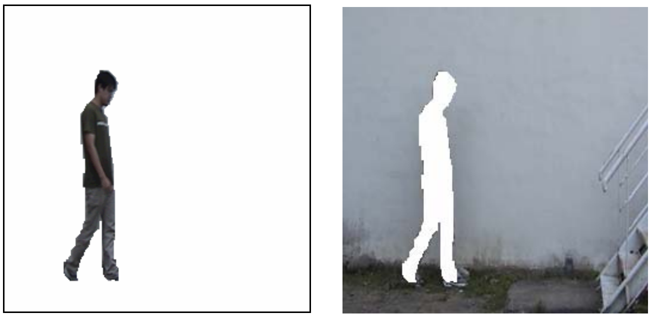


Fig. 3. The efficiency analysis for the RSST, FRSST, and MFRSST algorithms.



(c) (d)

Fig. 4. The segmentation of the foreground and background regions (mode I). (a) The foreground regions and (b) background regions for the walking image.

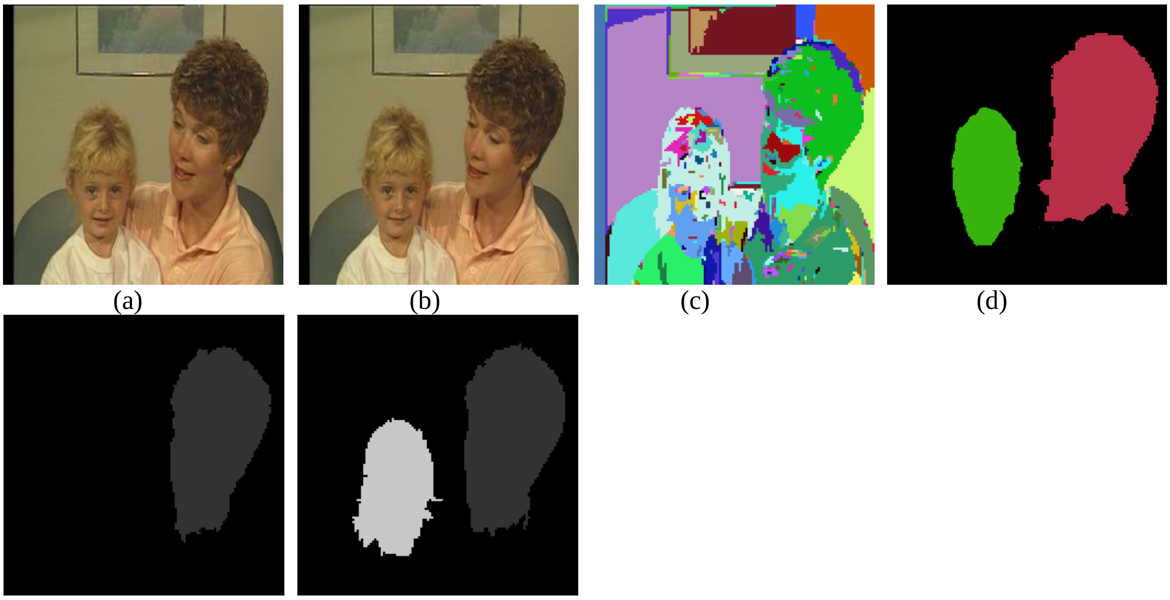
(a) (b) (c) (d)

(e) (f)

Fig. 5. The 15th and 17th frames of Akiyo image sequence are shown in (a) and (b). There are 310 color segmented regions in (c). The motion segmented regions of R_t^M and R_t^{MC} are shown in (d) and (e). The final segmentation result R_t^F is shown in (f).

Table 1. The moving regions and stationary regions in R_t^I.

Color Region R_t^I	{1, 2, 3, …, 310}
Moving region set $G(I_{t,i},R_t^M,x)$	{1, 2, 3, …, 236}
Stationary region set	{237, 238, …, 310}
Previous moving object $G(I_{t,i},R_t^{MC},x)$	{1, 2, 3, …, 236}



(a) (b) (c) (d) (e) (f)

Fig. 6. The 25th and 27th frames of mother & daughter image sequence are shown in (a) and (b). There are 273 color segmented regions in (c). The motion segmented regions of R_t^M and R_t^{MC} are shown in (d) and (e). The final segmentation result R_t^F is shown in (f).

Table 2. The moving region sets and stationary region sets in R_t^I.

Color Region R_t^I	{1, 2, 3, …, 273}
Moving region set 1 $G(I_{t,i},R_t^M,x)$	{1, 2, 3, …, 89}
Moving region set 2 $G(I_{t,i},R_t^M,x)$	{90, 121, , …, 178}
Stationary region	{179, 180, …, 273}
Previous moving object $G(I_{t,i},R_t^{MC},x)$	{1, 2, 3, …, 89}

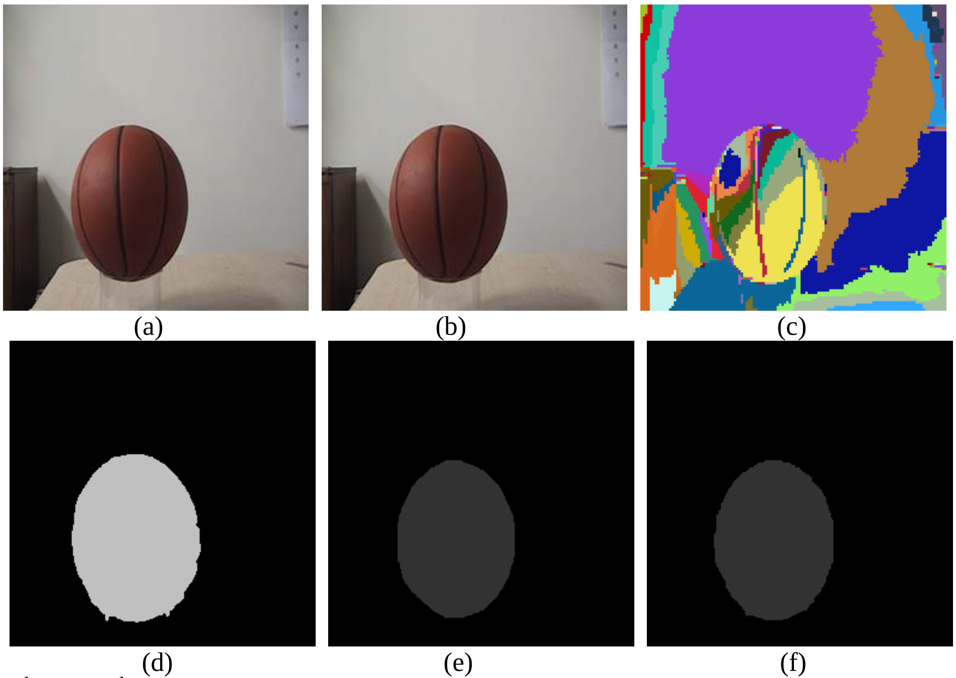


Fig. 7. The 5th and 7th frames of basketball image sequence are shown in (a) and (b). There are 108 color segmented regions in (c). The motion segmented regions of R_t^M and R_t^{MC} are shown in (d) and (e). The final segmentation result R_t^F is shown in (f).

Table 3. The moving region sets and stationary region sets in R_t^I.

Color Region R_t^I	{1, 2, 3, …, 108}
Moving region set $G(I_{t,i},R_t^M,x)$	{1, 2, 3, …, 39}
Stationary region	{39, 40, …, 108}
Previous moving object $G(I_{t,i},R_t^{MC},x)$	{1, 2, 3, …, 39} Except {10, 12, 35}

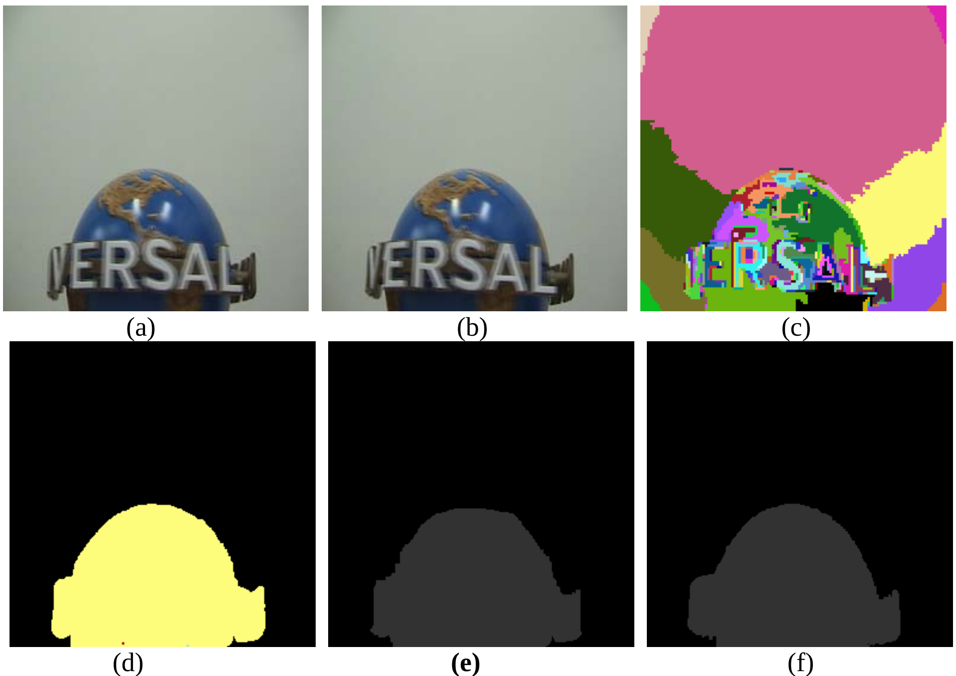


Fig. 8. The 64th and 65th frames of the image sequence of rotating globe are shown in (a) and

(b). There are 260 color segmented regions in (c). The motion segmented regions of R_t^M and R_t^{MC} are shown in (d) and (e). The final segmentation result R_t^F is shown in (f).

Table 4. The moving region sets and stationary region sets in R_t^I.

Color Region R_t^I	{1, 2, 3, …, 260}
Moving region $G(I_{t,i},R_t^M,x)$	{1, 2, 3, …, 250}
Stationary region	{251, 252, …, 260}
Previous moving object $G(I_{t,i},R_t^{MC},x)$	{1, 2, 3, …, 250} Except {7, 12, 22, …}

四、計劃成果自評

In this project, the concepts of multiresolution decomposition and flooding method in the watershed algorithms are applied to develop the multiresolution and flooding based RSST (MFRSST) image segmentation method that can speed up the vide segmentation process and have the satisfied quality. Based on the MFRSST algorithm the efficiency of COST211 video segmentation system is improved. Furthermore, two additional detection rules are proposed to extract the rotating and zooming objects in the video sequence. Once the video objects are acquired efficiently and accurately, many emerging multimedia applications, e.g., intelligent surveillance and video event detection may be developed.

五、參考文獻

[1] MPEG, MPEG-4: Applications document, Tech. Rep. ISO/IEC JTC1/SC29 /WG11/w2724, MPEG, Seoul, Korea, Mar. 1999.

[2] MPEG, MPEG-7: Applications document, Tech. Rep. ISO/IEC JTC1/SC29 /WG11/w2860, MPEG, Vancouver, Canada, July 1999.

[3] R. M. Haralick, and L.G. Shapiro, Image segmentation technique, *Computer Vision, Graphics & Image Processing*, 29, 1985, 100-132.

[4] R. Ohlander, K. Price, and D. R. Reddy, Picture segmentation using a recursive region splitting method, *Computer Graphics & Image Process.*, 8, 1978, 313-333.

[5] S. L. Horowitz, and T. Pavlids, Picture segmentation by a tree traversal algorithm, *J. Assoc. Computer Machine*, 23, 1976, 368-388.

[6] P. Willemin, T. R. Reed, and M. Kunt, Image Sequence Coding by Split and Merge, *IEEE Transactions on Communications*, 39(12), 1991, 1845-1855.

[7] O. J. Morris, M. J. Lee, and A. G Constantinides, Graph theory for image analysis: an approach base on the shortest spanning tree, *Proc. Inst. Elect. Eng.*, 133, 1986, 146-152.

[8] M. M. Chang, A. M Tekalp, and M. I. Sezan, Simultaneous motion estimation and segmentation, *IEEE Transaction on Image Processing*, 6, 1997, 1326-1333.

[9] G. D. Borshukov, G. Bozdagi, Y. Altunbasak, and A. M. Tekalp, Motion segmentation by multistage affine classification, *IEEE Transaction on Image Processing*, 6, 1997, 1591-1594.

[10] A. A. Alatan, E. Tuncel, and L. Onural, A rule-based method for object segmentation in video sequences, *IEEE Proc. Int. Conf. Image Processing ICIP 97*, 1997, 522-525.

[11] A. A. Alatan, L. Onural, M. Wollborn, R. Mech, E. Tuncel, and T. Sikora, Image sequence analysis for emerging interactive multimedia services-the European cost 211 framework, *IEEE Transaction on Circuit and System for Video Technology*, 8(7), 1998, 802-813.

[12] E. Tuncel and L. Onural, Utilization of the recursive shortest spanning tree algorithm for

video-object segmentation by 2-D affine motion modeling, *IEEE Transaction on Circuit and System for Video Technology*, 10(5), 2000, 776-781.
[13] A. D. Doulamis, N. Doulamis, and S. Kollias, Non-sequential video content representation using temporal variation of feature vectors, *IEEE Transactions on Consumer Electronics*, 46(3), 2000, 758-768.
[14] T. Vlachos and A. G. Constantinides, A graph-theoretic approach to color image segmentation and contour classification," *Proc. 4th. Int. Conf. Image Processing and Its Applications*, 1992.
[15] S. H. Kwok and A. G. Constantinides, "A fast recursive shortest spanning tree for image segmentation and edge detection," *IEEE Transactions on Image Processing*, 6(2), 1997, 328-332.
[16] L. Vincent and P. Soille, Watersheds in digital spaces: An efficient algorithm based on immersion simulations, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 13(6), 1991, 583-598.
[17] J. B. Kruskal, On the shortest spanning subtree of a graph and the traveling salesman problem, *Proc. Am. Mach. Soc*., 7, 1956, 48-50.
[18] R. Mech and M. Wollborn, A noise robust method for segmentation of moving objects in video sequences, *Proc*. 1997 *IEEE Int*. Conf. *Acoustics*, *Speech and Signal Processing ICASSP* 97, 1997, 2657-2660.