Time-Recursive Segmentation of Image Sequences

Montse PARDAS, Philippe SALEMBIER
Dept. Teoria del Senyal i Comunicacions,
E.T.S.E.T.B, UPC
Apdo. 30.002, 08080 Barcelona
Tel: (343) - 401 64 54, E-mail: pardas@tsc.upc.es

Abstract. This paper presents a segmentation algorithm for image sequences. The segmentation produces in a recursive way a three dimensional (3D = 2D plus time) partition of the sequence. The main issue of the process is to get a segmentation that is coherent in time and that solves the region correspondence problem. The method is completely based on morphological tools, which are very efficient to deal with object-oriented techniques such as segmentation. In order to extract visually important regions, the segmentation relies on size and contrast criteria. This method is particularly suitable for object-based very low bit rates video coding applications.

1. Introduction

Image sequence segmentation is an important issue in image processing for a large number of applications and, in particular, for second generation video coding techniques. For this purpose, a good segmentation should extract the visually important regions of the scene, be coherent in time, avoid as much as possible random fluctuations of the contour and solve the region correspondence problem. This last requirement indicates that one should be able to follow the time evolution of a given region (a given object). With these requirements, it is very difficult to obtain a good segmentation if the process is only 2D or intra-frame. Indeed, if frames are segmented independently, both the time coherence and the region correspondence problem will be difficult to solve. Somehow, the temporal relationship between frames must be exploited.

In this direction, a first solution consists in dealing with the sequence as a 3D (2D plus time) signal and in performing a 3D segmentation. This approach implies to split the sequence into 3D blocks of a given number of frames and to segment these 3D blocks. Examples of this approach can be found in [1], [2] and [3]. However, these techniques have some drawbacks. Indeed, if the temporal size of the blocks is large, major drawbacks appear: huge memory requirement, high computational load and introduction of a large processing delay discarding any interactive applications. By contrast, if the temporal size of the block is small, the temporal correlation between frames will be ignored at each block transition, that is very often.

However, most of the information contained in a frame is already present in the previous frame. The changes are mainly due to object motion and the appearance of new objects. This point of view leads to a segmentation process that can be called time-recursive. This approach involves two modes of operation: intra-frame and inter-frame segmentation. During the intra-frame mode, the first frame is segmented. It is a purely 2D process. Then, during the inter-frame mode, each frame is recursively segmented using the segmentation of the previous frame. Two different steps can be distinguished in the inter-frame mode: first, define the time evolution of the regions that are present in the segmentation of the previous frame and second, detect the appearance of new regions. Using this methodology, the drawbacks of purely 3D methods no longer apply: the memory requirement and computational load increase are moderate and the processing delay is minimized.

This paper presents a time recursive segmentation algorithm based on morphological tools. Morphological tools are particularly efficient when dealing with object-oriented techniques. Furthermore, they are computationally very efficient. The organization of this paper is as follows: Section 2 presents the intra-frame mode of the segmentation whereas section 3 is devoted to the inter-frame mode. Section 4 describes some results of segmentation. Finally, conclusions are reported in section 5.

2. Intra-frame Segmentation

The intra-frame segmentation relies on the hierarchical method proposed for still images in [4]. The algorithm first produces a simplified segmentation in the sense that it involves a reduced number of regions. Then, the segmentation is progressively improved by introducing more regions. Typically four segmentation steps are used. In the following, the principles of this method are summarized. More details can be found in [4].

Each segmentation step involves four basic transformations shown in figure 1: simplification, marker extraction, decision and modeling.

- *Simplification*: In this step, images are simplified to make them easier to segment. The simplification controls the amount of information that is kept for segmentation at this level of the hierarchy. Different simplification tools can be used depending on the segmentation criterion. For coding applications, two visually important criteria are used: size and contrast. A "size" segmentation means that all regions larger than a given limit are segmented. Using a contrast criterion, regions of high contrast are extracted. Morphological filters by reconstruction or area filters are very efficient for size simplification. Rh-maxima or minima operators are suitable for contrast simplification [5]. These operators are very attractive for segmentation purpose because they simplify the signal by removing small or poorly contrasted regions without corrupting the contour information. Moreover, they belong to the class of "connected operators" [6] which are known to produce flat zones (zones of constant gray level value). The concept of flat zones is very useful for the marker extraction step.

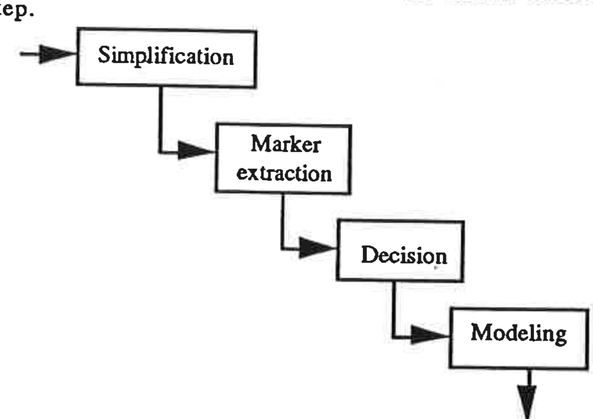

Figure 1: Intra-frame segmentation

- *Marker extraction:* This step is aimed to detect the presence of homogeneous regions. It produces markers identifying the interior of the regions that are going to be segmented. The marker extraction technique depends on the segmentation criterion. As explained in [5], for size oriented segmentation, the marker extraction consists in labeling the interior of large regions after the simplification. This can be easily done thanks to the simplification process which has produced flat zones. For contrast oriented segmentation, the marker extraction relies on the labeling of specific flat zones produced by the Rh-maxima / minima operators [5].

- *Decision:* After marker extraction, the number and interior of regions to be segmented are known. However, a large number of pixels are not assigned to any regions. These pixels belong to the so-called uncertainty areas. Assigning these pixels to a given region can be viewed as a decision process that precisely locates the contours between regions. The classical morphological decision tool is the watershed [7]. Generally it works on the morphological gradient of the image to segment. However, as discussed in [1] the use of the morphological gradient results in a loss of information on the contour position of ±1 pixel which is generally too high for coding application. To solve this problem, it was proposed in [1] to use a different version of the watershed algorithm working on the original image. The idea of using the watershed algorithm directly on the signal to segment was first proposed in [8] to deal with color images. Our approach follows these principles with some slight modifications. The resulting algorithm is a kind of region growing process: the set of markers is extended until they occupy all available space. During the extension pixels of the uncertainty areas are successively assigned to a given marker. A point is assigned to a specific region because it is in the neighborhood of at least one marker and it is more similar (in the sense defined by a specific criterion) to this marker than to any other marker of its neighborhood.

- *Modeling:* The next hierarchical step in the segmentation process will have to deal with the regions that cannot be well represented by the coding process. In order to have information about these regions, each region is texture-coded, that is, filled with a gray level function which represents the coded version of the region the receiver will have. Then, the difference between the coded image and the original one, called the modeling residue, is computed. Since this residue concentrates all the information about the badly coded region, it is used as signal to segment by the next segmentation steps.

Typically four segmentation steps are used for the intra-frame processing. The three first ones deal with the size criterion and the last one with the contrast criterion. This hierarchical procedure has the following main advantages: it produces a good segmentation of the image taking into account the possibilities of the texture coding (it segments the coding residue). It allows the progressive estimation of the segmentation parameters, size and contrast, to get a segmentation result compatible with the coding objective (appropriate number of regions or of contour points, etc.).

3. Inter-frame segmentation

After the intra-frame segmentation, a time-recursive procedure has to be performed in order to segment the successive images in a coherent way. As discussed in the introduction, two main steps can be distinguished: first, extension of the previously segmented regions into the current frame to segment and, second, extraction of new regions that might have appeared in the scene. These two points are discussed in the following:

2.1 Extension of the previous segmented regions into the current frame.

This process is basically a projection problem. Denote F_{t-1} and F_t the frames at time t-1 and t. Assume that the segmentation at time t-1, S_{t-1}, is known. Our goal is to find the segmentation, S_t, at time t without introducing new regions. For this purpose, two 3D signals are going to be constructed. Frames F_{t-1} and F_t are grouped together to form a temporal block $\mathcal{F}$ of size 2 in the time direction. Similarly, the frame S_{t-1} is grouped with an empty frame S_o representing an entire frame of uncertainty. The resulting 3D signal denoted S is considered as the set of markers that should be used to segment the signal $\mathcal{F}$.

The extension itself is achieved by the decision algorithm used in the intra-frame segmentation. The only difference is the nature of the signals that are 3D signals. The watershed extends the markers defined by S_{t-1} into the empty frame S_o. Each pixel of the uncertainty area (that is of frame S_o) is assigned to a region of frame S_{t-1} based on a similarity criterion. The similarity consists mainly of the gray tone difference between the pixel of $\mathcal{F}$ under consideration and the mean of the pixels that have already been assigned to the region.

This basic similarity measure has to be modified to take into account the shape of the contours. Indeed, for video coding, where the gray level transition between two regions is not very strong, it is interesting to have simple contours, even if some precision on the position is lost. Similarly, in the time dimension, it is convenient to have stable regions. This requirement can be also seen as a contour straightness. Therefore, the objective is to have straight contours where the transition between regions is not very strong, both in the spatial and time dimensions. For this aim, the similarity measure is modified. The similarity is the weighted sum of the gray level difference between the pixel and the mean of the region plus a penalty term corresponding to the contour complexity:

$$\text{Similarity} = \alpha \,.\, \text{Difference in gray tone} + (1-\alpha) \,.\, \text{Contour complexity}$$

The measure of contour complexity is made by counting the number of contour points that are added if the pixel is assigned to that region. A 6-neighborhood is used for this aim, 4 pixels in the spatial dimension and 2 for the temporal. This measure is weighted and added to the gray level distance of the pixel to the region under consideration. The weighting factor α allows to give more importance to the gray level measure or to the contour complexity. The whole extension technique is illustrated in figure 2.

2.2 Extraction of new regions.

Once the labels of the previously segmented regions have been propagated in the current image, new regions have to be extracted. From now on, the process is purely intra-frame. First of all, the residue image is computed. As for the intra-frame processing discussed previously, the residue image is obtained by texture-coding of the segmentation obtained from the extension of the past regions. Then, the difference with the original image gives the residue, where new appeared regions can be detected.

The segmentation of new regions is performed with the method used for the intra-frame segmentation. That is, it consists on the basic segmentation steps: simplification, marker extraction and decision. The segmentation can be performed according to a size or a contrast criterion. Our experience has lead us to use a contrast segmentation only since, most of the time, large and non contrasted regions of the *residue* are not visually important.

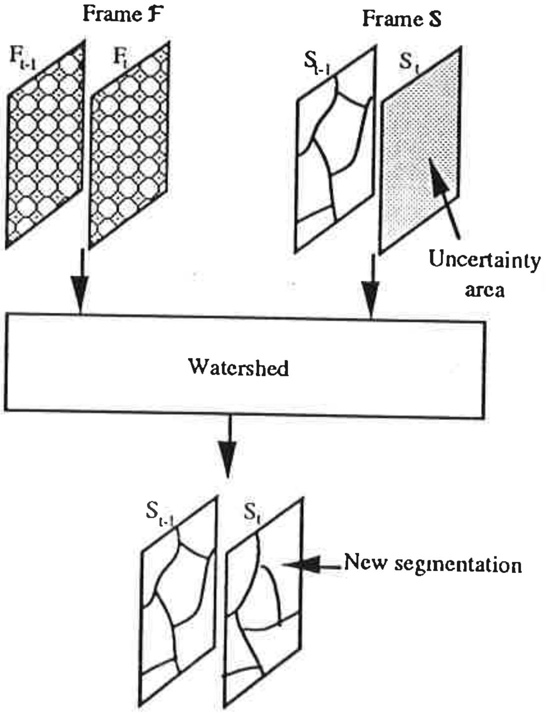

Figure 2: Time extension of regions

4. Results

We present in this section two examples of the segmentation process and its results. In Figure 3 the intra-frame process is shown. The first picture represents the original image and the following ones the successive steps after the modeling process. The first three images correspond to a size criterion for the segmentation, and the last one to a contrast criterion. The parameters used for the simplification at every step are automatically computed from the final number of contour points which is required. In this example the requirement was 4500 contour points and the resulting segmentations contain 9,24,61,and 105 regions successively. The modeling step has been performed using a second order polynomial to model the inside of each region.

Figure 4 presents an example of the inter-frame process. The first row shows 3 original images from the sequence "foreman". The second row shows the segmentation which would be obtained using only the projection step. That is, the regions of the first image are extended into the new frames, but no new region is introduced. This segmentation is modeled with a first order polynomial and the residue is computed (that is, the difference of each image with the corresponding image in the upper row). From this residue the contrasted regions can be extracted. The result after this extraction of new regions, coded with a second order polynomial is shown in the third row. Only one new region has appeared in the third frame. It corresponds to the teeth. The good quality of this segmentation for coding purposes can be observed in these images, even using simple models for the texture as the second order polynomials. Finally, in the forth row we have represented every region in the final segmentation with a different gray color to make it easier to follow the behavior of the regions along the time.

5. Conclusions

We have presented in this paper a time-recursive method based on morphological tools. It works in two different modes: intra and inter frame. After an initial segmentation, a time recursive procedure is iterated which first extends the previously segmented regions and then extracts new regions that might have appeared in the actual frame. The extension of old regions, which is the only 3D process in the segmentation, allows to exploit the temporal information present in the image sequences. The watershed algorithm has proved to be very efficient to extend the segmentation of the previous frame into the current one. Furthermore, it allows to have a control on the shape of the contours and on the temporal stability.

It is shown in this way that it is not necessary to use purely 3D methods to exploit the temporal information, which have different drawbacks and usually become more complicated as they segment the entire 3D block from scratch instead of strongly relying on previous segmentation information.

This method is being used in a segmentation-based coding scheme designed for very low bit rates applications [9]. Bit rates of the order of 10-20 Kbits are achieved within this framework.

Acknowledgements

This work has been performed within the framework of the RACE MORPHECO project and has been partially supported by a grant of the Comissionat per Universitats i Recerca, Generalitat de Catalunya.

References

[1] **M. Pardàs, P. Salembier (1994)**, "3D Morphological Segmentation and Motion Estimation for Image Sequences", in EURASIP *Signal Processing*, Special Issue on Mathematical Morphology, Vol. 38, N. 2, September 1994.

[2] **S. Rajala, M. Civanlar and W. Lee (1988)**, "Video data compression using three-dimensional segmentation based on HVS properties", *Proc. of IEEE Int. Conf. on Acoustics Speech and Signal Processing'88, ICASSP'88*, New York, USA, pp.1092-1095.

[3] **P. Willemin, T. Reed and M. Kunt (1991)**, "Image sequences coding by split and merge", *IEEE Trans. on Communications*, Vol. 39, No12, pp 1845-1855.

[4] **P. Salembier (1994)**, "Morphological multiscale segmentation for image coding", in EURASIP *Signal Processing* (Special issue on Nonlinear Signal Processing), Vol 39, No 2, December 1994.

[5] **P. Salembier (1993)**, "Multi-criterion segmentation for image coding", *Proc. of International Workshop on Mathematical Morphology and Its Applications to Signal Processing*, Barcelona, pp.40-45.

[6] **J. Serra and P. Salembier (1993)**, "Connected operators and Pyramids", *Proc. of SPIE Image Algenra and Mathematical Morpyhlolgy*, San Diego, Vol. 2030, pp. 65-76, 1993.

[7] **F. Meyer and S. Beucher (1990)**, "Morphological Segmentation", *Journal of Visual Communication and Image Representation*, Vol. 1, No.1 pp. 21-46.

[8] **F. Meyer (1992)**, "Color Image Segmentation", *Proc. 4th International Conference on Image Processing and its Applications*, Maastricht, 1992, pp. 303-307.

[9] **P. Salembier et al. (1994)**, "Very low bit rate video coding using morphological segmentation and contour/texture motion compensation", *Proc. of the 12th Int. Conf. on Pattern Recognition, Jerusalem*, Oct. 1994.

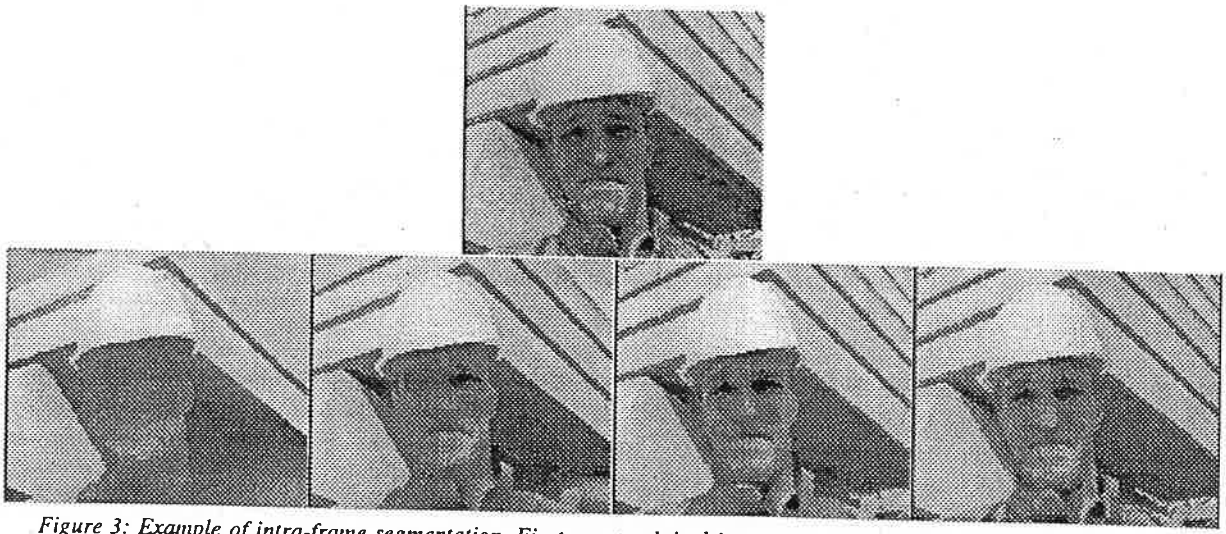

Figure 3: Example of intra-frame segmentation. First row : original image. Second row: four hierarchical steps.

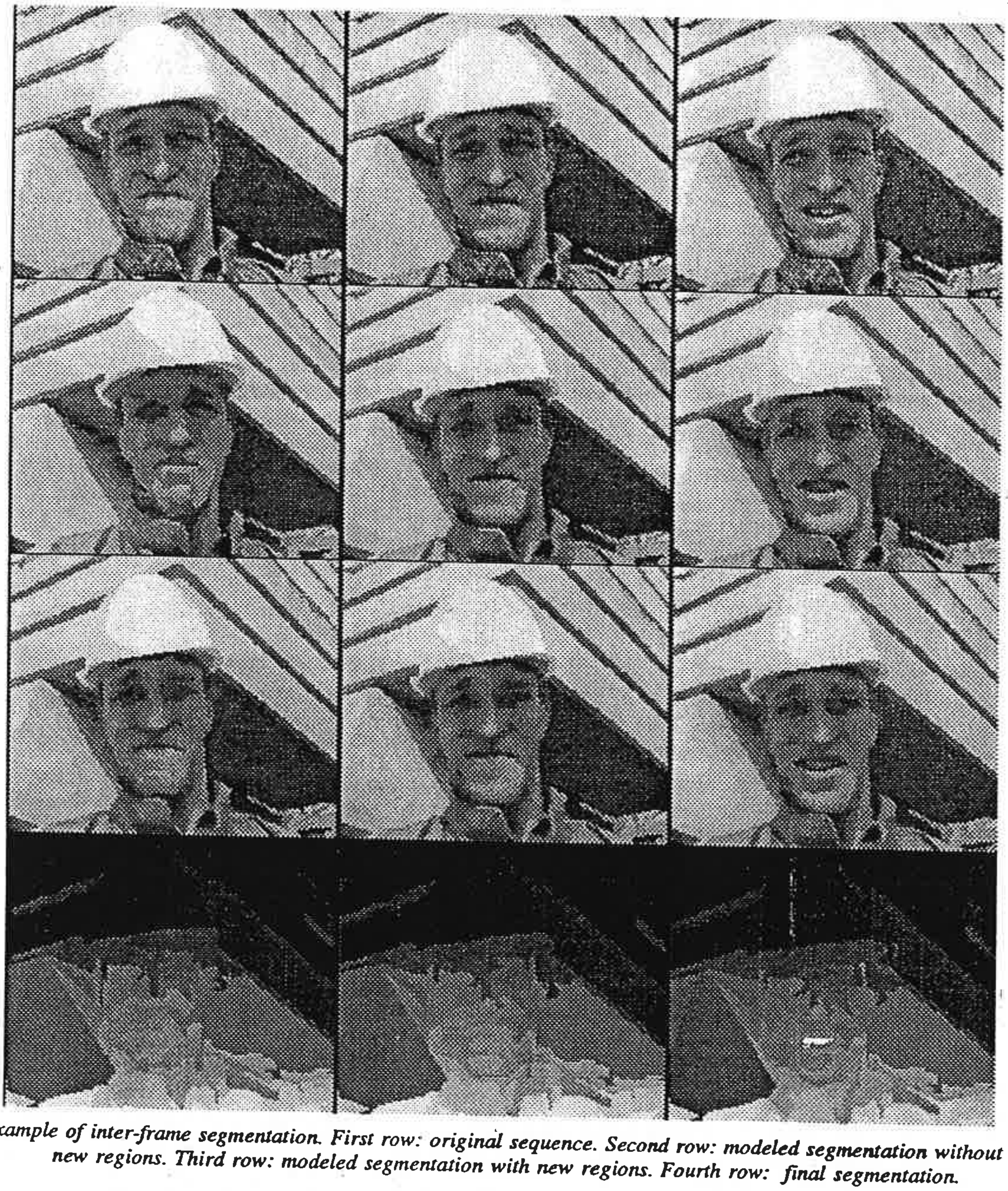

Figure 4: Example of inter-frame segmentation. First row: original sequence. Second row: modeled segmentation without extraction of new regions. Third row: modeled segmentation with new regions. Fourth row: final segmentation.