

Recursive Image Sequence Segmentation by Hierarchical Models

F. Marques, V. Vera, A. Gasull
Dept. Teoria del Senyal i Comunicacions
E.T.S.E.T.B. - U.P.C.
Apdo. 30002, 08080 Barcelona, SPAIN

Abstract

This paper addresses the problem of image sequence segmentation. A technique using a sequence model based on Compound Random Fields is presented. This technique is recursive in the sense that frames are processed in the same cadency as they are produced. New regions appearing in the sequence are detected by a morphological procedure.

1 Introduction

The usual way to face the problem of segmenting image sequences is by extending a segmentation technique for still images. A few methods have been proposed, mainly under the scope of the so-called Second Generation image coding techniques [2]. The basic "region growing" method was extended to the 3D case [7], as well as the "split and merge" approach [9]. In both cases, the main drawbacks of the basic 2D methods lead to cumbersome 3D techniques.

To overcome these drawbacks, different 2D approaches have recently been proposed. Two of them, although based on very different concepts, share the feature of being pure Top-Down approaches. The first one uses morphological tools [8] and has already been successfully extended to the 3D case [6]. The second one [3] relies on Compound Random Fields (CRFs) [1] and deterministic maximization.

This paper deals with the extension of the above commented image segmentation technique [3] to a recursive sequence segmentation method. That is, the segmentation of a given frame takes into account only a small window of frames around it. This improvement is twofold: first, the necessary delay can be drastically reduced and second, the parameters of the image model characterizing the temporal behavior can be locally adapted. The problem of new regions appearing in the sequence has been solved by a morphological step which detects their presence.

The structure of this paper is as follows: the image sequence model is presented in Section 2. Section 3 is devoted to the presentation of the basic segmentation algorithm, while Section 4 describes the proposed improvements: the recursive and adaptive parts as well as the morphological detection of new regions. Some results and conclusions are discussed in Section 5.

2 Hierarchical Image Sequence Model

In this work, image sequences are modeled by 3D CRFs. Sequences are assumed to be composed of a set of regions defined on the space-space-time domain. Each one of these regions is characterized by an independent random field. The union of these random fields forms the observed sequence (upper level of the model). The location of the random fields within the sequence is governed by an underlying random field (lower level of the model).

The chosen 3D CRF relies on that presented in [3] for the 2D case. The upper level is a set of white Gaussian Random Fields (GRFs). Each GRF characterizes the texture information within a region in both the spatial and the time domains. The lower level is a second order Strauss process [1] defined also in both spatial and time domains.

The model extension to the sequence case for the upper level has been performed straightforward. However, the modeling of a sequence partition by a Strauss process has required further study. Strauss processes, being a special case of Markov Random Fields, are governed by a Gibbs Distribution [1] whose general expression is:

$$P(X = x) = 1/Z \exp(-U(x)) \quad (1)$$

where $U(x)$ is an energy function, and Z is a normalizing constant. The energy function $U(x)$ is usually expressed as a combination of potentials $V_c(x)$ associated with the cliques of the neighborhood of x [1].

A Gibbs Distribution can be described by the set of potentials assigned to each one of its possible cliques. In our case, the Strauss process has to model the contour behavior in the sequence partition. This results in a set of constraints [5] which leads to four different types of cliques: two for the spatial domain (type 1s: horizontal and vertical ones, type 2s: diagonal ones) and two for the temporal one (type 1t: horizontal ones, type 2t: diagonal ones).

The energy function characterizing the Gibbs distribution of the 3D Strauss process can be expressed as:

$$U(x) = 1/T [k_{1s}V_{1s} + k_{2s}V_{2s} + k_{1t}V_{1t} + k_{2t}V_{2t}] \quad (2)$$

where k represents the number of cliques of a given type present in the partition, V stands for the potential and s and t denote the space and time domains, respectively. This set of parameters can be reduced to four:

$$U(x) = 1/T^* [k_{1s} + k_{2s}V_s^* + R_{st}^*(k_{1t} + k_{2t}V_t^*)] \quad (3)$$

where T^* controls the relative importance between the lower level and upper level models, V_s^* characterizes the spatial boundaries, R_{st}^* sets the relation between the spatial and temporal parts of the lower level model and V_t^* models the temporal boundaries [5].

3 Basic Segmentation Algorithm

The way to use the above sequence model within a segmentation procedure is by assuming that sequences are realizations of the upper random field $X = x$. The objective is therefore to determine the realization of the partition $Q = q$ which more likely has given rise to x . To carry out this procedure in a feasible amount of time, a deterministic maximization approach is followed. The maximization is performed by seeking iteratively the labeling leading to a maximum likelihood, for all pixels in the sequence laying on the boundaries of the current segmentation. A very simple algorithm, allowing a fast implementation, has been developed to carry out this procedure [4].

In [4], image sequences are treated as large blocks of frames which are separately processed. The correct initial conditions necessary to achieved a good final result are obtained by using a 3D Gaussian pyramid decomposition of the image sequence. Given that decomposition, the segmentation procedure is performed starting from the coarsest level. The segmentation result of a given level l is used as initial condition for segmenting the following finer resolution level $l - 1$. This procedure is iterated down to the finest resolution is segmented. To further improve the quality of the final segmentation results, the value of the model parameter controlling the relative importance between the lower and upper levels T^* is adapted to the data present at each decomposition level.

This method arises some problems when dealing with sequences whose motion varies rapidly. The parameter characterizing the motion in the sequence is estimated from the whole block of frames and, therefore, an average value is assumed. Furthermore, when applying this technique to interactive or real-time applications such as image sequence coding, the joint use of large blocks of frames results in an unfeasible delay.

4 Recursive Segmentation Algorithm

The above drawbacks can be shortcut using a recursive approach. To reduce the delay, the optimum way would be to process the frames at the same cadency at they are produced. Nevertheless, it has been observed that segmentation results are conform better to the actual sequence information if, when segmenting a given frame n , some information about the future frames is incorporated. In practice, it is enough to segment a given frame n relying on the segmentation of the previous one $n - 1$ and on the data contained on the posterior frame $n + 1$.

As above commented, the quality of the final segmentation depends on that of the initial estimation. However, the use of the multiresolution approach is only necessary

in the first frame [3]. The final segmentation of a given frame n is good enough as approximation of the following one $n + 1$. Therefore, only a unique resolution is to be used for segmenting a generic frame n .

The initial segmentation of a given frame n has been obtained in the previous step (segmentation of frame $n - 1$) and it is imposed as initial estimation for frame $n + 1$, as well. The deterministic maximization commented in Section 3 is now applied, allowing only refinements on the partitions of frames n and $n + 1$. Once the maximum is reached (that is, the segmentation for frame n is achieved), the whole procedure is iterated.

This recursive approach allows also an improvement regarding to the estimation of the model parameters. Given that the segmentation procedure is performed iteratively on a small amount of frames, the parameter related to the motion in the sequence can be more locally estimated.

In (3), V_s^* and V_t^* control the behavior of the contours in the spatial and time domain, respectively. These parameters can be set a priori using prior information about the application on which segmentations are to be used; in our case, image sequence coding. For this case, a range of values for both parameters has been given [5].

In addition, the behavior of spatial and temporal contours are related by means of R_{st}^* . This parameter opens the possibility of, for instance, giving priority to temporal contours if a large motion is detected. To conform the hierarchical model to natural motion, the set of possible values of the parameter R_{st}^* has to be constrained. The analysis of the range of possible values is given in [5].

The parameter T^* controls the relative importance between the upper and the lower levels of the model; that is, the relationship between texture and contour information. This relationship has to be adapted in multiresolution segmentation schemes, given that it varies from one resolution to another. However, in sequence segmentations that use always the same resolution, T^* should not vary, provide that the sequence itself does not change drastically. Therefore, T^* is only adapted when segmenting the first frame and remains fixed for the rest of the procedure. When segmenting image sequences, regions may vary their shape, but they may also disappear or appear in a new frame. The above segmentation algorithm (Section 3) is based on successively refinements on the contours of an initial segmentation. In our case, this initial segmentation is that of the previous frame. Therefore, the algorithm can refine the contours of a region from frame $n - 1$ but it is not able to detect new regions appearing in the frame.

To detect such regions, an error image is built. This image contains the difference between a frame and its segmentation, filling each region with its mean value. Error image information is mainly related to textures in the original image and non-detected interior regions. Therefore, the main task is to discriminate between these two types of information. Morphological tools [8] have been applied, since they are well known to efficiently discriminate between these kinds of information. The residue of

the original signal and a morphological center (contrast extraction transform) turns out to be very efficient for detecting interior regions. The center is computed from the open_close, the close_open and the identity operators:

$$\text{Contrast extractor : } I - \text{center}(\gamma\varphi, \varphi\gamma, I) \quad (4)$$

where γ and φ stand for the open and close filters respectively, and the center is defined as

$$\text{center}(\gamma\varphi, \varphi\gamma, I) = \text{Min}(\gamma\varphi, \text{Max}(\varphi\gamma, I)) \quad (5)$$

Once an interior region has been detected, the previous segmentation technique decides whether this new region deserves to be included into the final segmentation. This way, the segmentation algorithm is allowed to decide whether a seed should be expanded, shrunk or removed.

The last improvement that has been introduced is the possibility to reset the procedure. This is important if the sequence changes drastically. Then, the dependency of the current frame with the past ones can be very small or even null. Actually, the algorithm is robust enough to recover for such changes very rapidly. However, depending on the application, this stabilization process can be very negative. For instance, during the stabilization, contours may change very fast and their evolution may be very difficult to predict and very costly to code. After the reset, the whole procedure starts again; that is, the first frame is segmented by the multiresolution algorithm [3] and the parameter T^* is recomputed.

5 Results and Conclusions

Figure 1 shows in the first row four frames of the Table-Tennis sequence, in the second row the spatial contours of its segmentation, in the third row these segmentations filling each region with its mean value and in the fourth row the temporal contours. In this sequence there is mainly motion and no objects appear or disappear from the previous frame. The segmentation of this sequence contains 47 regions which remain the same from the first to the last frame. The algorithm is able to track the regions through the temporal domain. In addition, the morphological step works correctly since no new region has been introduced in spite of coping with a very textured sequence. The sequence segmentation algorithm uses an image sequence model to which the possible prior knowledge of the kind of sequence to be segmented can be easily incorporated. That is, the model can be tuned to satisfy a given sequence behavior. Moreover, the algorithm itself is very simple since avoids any stochastic maximization and performs very fast. As average, it needs 27 seconds for segmenting a frame of size 256x256 in a Sun Sparc Workstation. Finally, it is unsupervised and, being tested on a large set of sequences with different characteristics, has yielded good results.

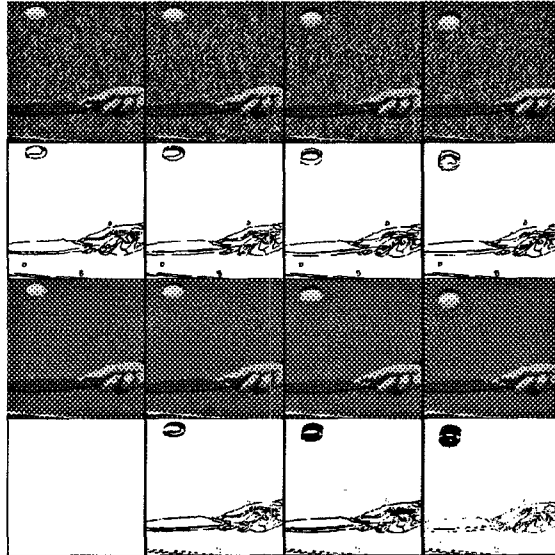


Figure 1: Example of segmentation

References

- [1] S. Geman, D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," *IEEE Trans. PAMI*, Vol. 6, pp. 721-741, 1984.
- [2] M. Kunt, A. Ikonomopoulos, M. Kocher, "Second generation image coding techniques," *Proceedings of the IEEE*, Vol. 73, pp. 549-575, 1985.
- [3] F. Marques, J. Cunillera, A. Gasull "Unsupervised segmentation controlled by morphological contrast extraction," *Proceedings of ICASSP'93*, pp. 5.17-5.20, 1993.
- [4] F. Marques, J. Cunillera, A. Gasull "Multiresolution image sequence segmentation," *Proc. of the IEEE Int. Conf. on Signal Processing'93*, pp. 859-862, 1993.
- [5] F. Marques, V. Vera, A. Gasull "A hierarchical image segmentation model for segmentation: Application to object-based sequence coding," *Proc. of SPIE Visual Communications and Image Processing*, 1994.
- [6] M. Pardas, P. Salembier, "3D Morphological segmentation and motion estimation for images sequences," to be published in *EURASIP Signal Processing*, 1994.
- [7] S. Rajala, M. Civanlar, W. Lee, "Video data compression using three-dimensional segmentation based on HVS properties," *Proceedings of ICASSP'88*, pp. 1902-1905, 1988.
- [8] J. Serra, *Image analysis and mathematical morphology*, Academic Press, London, 1982.
- [9] P. Willemin, T. Reed, M. Kunt, "Image sequences coding by split and merge," *IEEE Transactions on Communications*, Vol. 39, pp. 1845-1855, 1991.