SEGMENTATION-BASED VIDEO CODING: TEMPORAL LINKING AND RATE CONTROL

Philippe Salembier, Ferran Marqués and Montse Pardàs
Universitat Politècnica de Catalunya
Campus Nord - Mòdul D5
C/ Gran Capità, 08034 Barcelona, Spain
E-mail: {philippe,ferran,montse}@gps.tsc.upc.es

ABSTRACT

This paper analyzes the main elements that a segmentation-based video coding approach should be based on so that it can address coding efficiency and content-based functionalities. Such elements can be defined as temporal linking and rate control. The basic features of such elements are discussed and, in both cases, a specific implementation is proposed.

1 INTRODUCTION

Coding algorithms relying on a content-based representation are becoming increasingly popular. The main difference between the techniques reported in the literature is the relative importance they assign to the spatial or the motion information. In a first set of examples [17, 5, 15], regions are characterized by their spatial homogeneity, whereas in a second set [8, 2], motion information is used as the main homogeneity criterion. Finally, examples of segmentation techniques involving both criteria can be found in [3, 1].

This last approach offers a large number of advantages: the spatial homogeneity allows a very accurate definition of the regions and the processing of generic sequences because, for example, it can deal with the appearance of new objects. The motion homogeneity criterion is important to limit the number of regions and therefore the coding cost associated to the partition. This limitation is generally obtained by merging spatial regions that can be compensated together. Neighbor regions with homogeneous motion are usually related to the concept of objects. Such mergings can therefore allow an object representation of the scene. This representation opens the door to content-based functionalities [7].

Segmentation techniques combining regions with motion and spatial homogeneity should profit from all the previous advantages. As main goals, segmentation techniques should yield partitions that correctly represent the objects in the scene while allowing the improvement of the coding efficiency. In addition, they have to enable the temporal tracking of objects in order to really address content-based functionalities.

In this paper, the necessary elements of a segmentation-based video coding scheme aiming at the previous goals are discussed; namely, the temporal link and the rate control. Moreover, for each one of them, a specific implementation is proposed. This discussion is structured in the paper as follows. In Section 2, the need for a temporal link in the segmentation procedure is analyzed and a technique based on a double-partition approach is presented. Section 3 studies the necessity of a rate control involving a joint optimization of the segmentation and the coding strategy and proposes a method relying on a set of hierarchical partitions; the so-called *Partition Tree*. Finally, Section 4 gives the conclusions of the paper.

2 TEMPORAL LINKING

Video sequence segmentation for coding purposes should assign to an object appearing in consecutive frames the same label (number associated to a region in the partition). This temporal link allows the improvement of the coding efficiency while enabling content-based functionalities:

- **Coding efficiency:** Texture compensation can be applied more efficiently if the areas of homogeneous motion or texture are related through the time domain. In addition, motion compensation techniques can be used for coding the partition sequence (contour information).
- **Content-based functionalities:** The temporal link of the object representation in the different frames permits to address the object as a single spatio-temporal entity (concept of Video Object Plane -VOP-).

2.1 Partition projection

The temporal link between consecutive partitions can be achieved by using a *Projection* step in the segmentation [9, 5]. This step adapts the partition P_{T-1} of frame $T-1$ to the current frame T. This link is obtained by estimating the motion between consecutive frames and compensating P_{T-1}. After compensation, a set of markers, approximating the position and shape of the previous regions in the current image, is obtained.

Usual *Projection* approaches extend the compensated regions in the current image assuming that regions are spatially homogeneous. However, as stated in the introduction, partitions should contain regions with temporal as well as spatial homogeneity. Therefore, an algorithm for projecting partitions without restrictions on the type of homogeneity is necessary.

2.2 Double-partition approach

This approach extends the work presented in [9] to the case of partitions with regions which are not spatially homogeneous. To solve this problem, the *Projection* block uses a double partition approach [4]. This approach is illustrated in Fig. 1.

The partition of the previous image P_{T-1} is re-segmented in order to achieve a finer partition. This fine partition contains a larger number of regions which are obtained by re-segmenting the regions in P_{T-1} following spatial criteria. The objective is to guarantee the spatial homogeneity of each region.

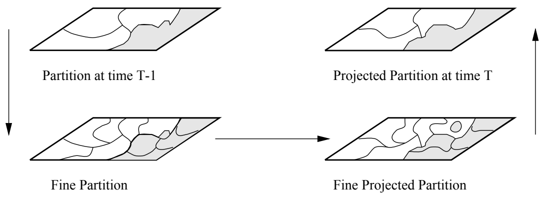

Figure 1: *Projection* based on a double-partition approach

This fine segmentation is then projected in the current frame so that a fine segmentation at time T is achieved. To obtain the partition P_T, only those contours of the fine projected partition associated to the coarse partition have to be kept. Then, the projection of a region from the coarse partition is obtained by merging the projections of the regions that belong to it at time T. However, this procedure does not ensure that an actual partition is obtained. Indeed, unconnected regions with the same label may appear after relabeling. This problem is solved by keeping the largest component for each label and removing the other components with the same label. A pure 2D region-growing such as the watershed algorithm is applied to obtain the final partition.

3 SEGMENTATION AND RATE CONTROL

3.1 A link between the analysis and the coding

Most of the segmentation-based coding schemes discussed in the introduction have the same structure illustrated in Fig. 2 involving an *analysis* or *segmentation* step followed by a *coding* step. The main drawback of this coding strategy is that there is no strong relation between the two steps. One tries to obtain the "best" segmentation, and then, to code the partition as well as the texture with the lowest possible amount of bits. However, high coding efficiency requires a strong interaction between the segmentation and coding steps. Indeed, the segmentation should depend on the set of available coding techniques.

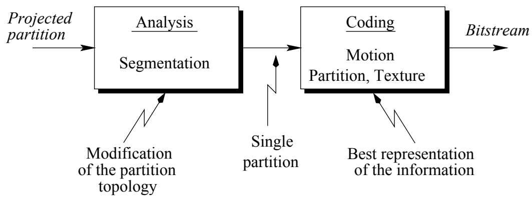

Figure 2: "Segmentation and (then) coding" structure

Based on the work done for bit allocation reported in [16, 11], a solution for an optimal definition of a partition and of the set of coding techniques to be used in each region is proposed in [12]. As illustrated in Fig. 3, the link between the analysis and coding steps can be achieved by introducing a *Decision* block and by modifying the objective of the analysis phase. Now, the goal of the analysis is to create not a single partition but a *universe* of regions out of which the optimal partition will be constructed. Based on this universe of regions and on the definition of a set of possible coding techniques, the objective of the *Decision* is to select the optimum set of regions to form the final partition and to define how each region should be coded. In [12], the universe of regions is defined for each frame independently of the previously transmitted partitions. Therefore, there is no time tracking possibility. In the following section, we present the concept of *Partition Tree* to solve the problem of bit allocation while maintaining the temporal linking defined by the projection.

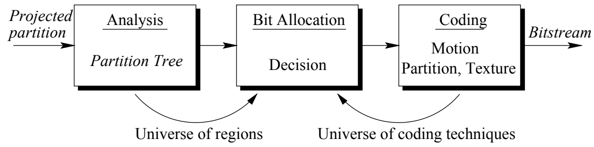

Figure 3: Joint optimization of the segmentation and of the coding strategy

3.2 Partition Tree and optimization

Besides defining the universe of regions for the *Decision*, the *Partition Tree* [10] relates each region to a region belonging to the projected partition and structures this information in a hierarchical way. The first requirement is used in order to maintain the temporal linking between regions in successive frames, whereas the second one is useful to limit the complexity of the *Decision* process.

The creation of the *Partition Tree* is illustrated on the left side of Fig. 4. The idea consists in creating variations with respect to the projected partition, that is to create new regions either by segmentation or by merging of already known regions. The *Partition Tree* is formed by the projected partition plus a set of hierarchical partitions that are "above" and "below".

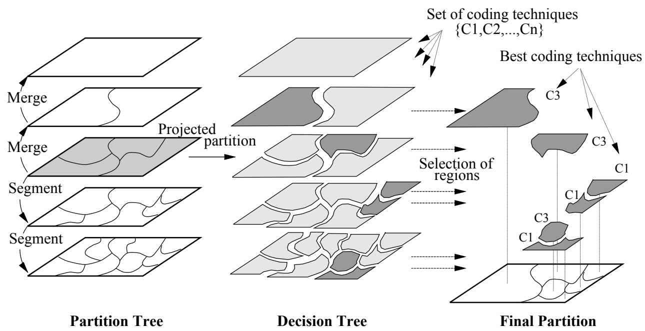

Figure 4: *Partition Tree* and bit allocation

Below the projected partition, several levels of finer partitions are created. The partitions are finer in the sense that they involve a larger number of regions and that the contours present at a given level are also present at lower levels. This procedure can be viewed as a hierarchical segmentation splitting regions of a given level to produce the regions of the next lower level [13, 15]. The segmentation criterion relies on the spatial homogeneity. Indeed, one of the main objectives of these segmentation steps is to introduce in the universe of regions new regions that can represent objects appearing in the scene. This explains why the procedure is purely spatial and does not take into account any motion information.

Above the projected partition, several levels of coarser partitions are created. Here, regions are merged if they can be processed globally, that is, if they can be compensated by the same motion information. Merging regions on the basis of their homogeneity in motion reduces the partition coding cost and goes closer to the notion of objects.

As can be seen, the *Partition Tree* defines the universe of possible regions and relates each region to the projected partition: on upper levels of the tree, regions are obtained by merging of projected regions, whereas on lower levels, regions are obtained by segmentation of projected regions. Note that the regions belonging to the *Partition Tree* can be homogeneous either spatially (lower levels of the tree) or in motion (upper levels of the tree). Finally, no decision is made concerning the actual partition to be coded. The objective is simply to define a reduced set of regions that are likely to be part of the optimum partition.

The decision process itself is illustrated on the right side of Fig. 4. The *Decision* block selects the best strategy in terms of regions and coding techniques. The *Partition Tree* contains all regions that may belong to the final partition; for each of them, a set of possible coding techniques involving several region-based coding schemes with various levels of quality is proposed. Note that techniques can be proposed in intra mode (coding of the original signal) or in inter mode (motion compensation of the region and coding of the prediction error).

To define the coding strategy in the Rate-Distortion sense, the rate and distortion associated to each region of the *Partition Tree* and to each possible coding technique are computed. Based on this information, the optimization can be performed using the technique discussed in [11, 12]. The problem can be formulated as the minimization of the distortion D of the image with the restriction that the total cost R is below a given budget (defined for each frame). It has been shown that this problem can be reformulated as the minimization of the Lagrangian: $D + \lambda R$ where λ is the so-called Lagrange parameter. Both problems have the same solution if we find λ^* such that R is equal (or very close) to the budget. Therefore, the problem consists in finding a set of regions creating a partition and a set of coding techniques minimizing $D + \lambda^* R$. Thanks to the hierarchical structure created by the *Partition Tree* fast optimization algorithms can be used (see [12, 6] for more details).

3.3 Examples

An example of *Partition Tree* creation and decision process is given in Fig. 5. The image on the first row corresponds to the partition of the previous frame. This partition is projected and the projected partition can be seen in the center of the second row. This step defines the time evolution of the previously transmitted regions. Based on the projected partition, the *Partition Tree* is created: in the example of Fig. 5, levels 1 and 2 are obtained by hierarchical segmentation following a spatial homogeneity criterion. Note in particular, how regions representing details of the face or of the background are introduced in the universe of regions. Levels 4 and 5 are created by merging regions with similar motion. Note here how background regions are merged because of their homogeneity in motion. The final partition is shown in the center of the lower row. In this partition, some regions are homogeneous in terms of gray level (regions corresponding to homogeneous part of the building) and others are homogeneous in motion (region corresponding to the face). Finally, the original as well as the resulting coded frames are shown in the lower row. The coding has been achieved using the algorithm described in [14] where more details can be found about the motion estimation/compensation and the coding techniques actually used.

Sequ.	Kbit/s	Dec.	Motion	Part.	Texture
Foreman	42	1.8 %	4.2 %	19.0 %	75.0 %
Children	320	1.0 %	2.1 %	8.9 %	88.0 %

Table 1: Bit allocation for two examples

The bit allocation among the different types of information to be sent for two examples are summarized in Table 1. Note that the bit rates involved in the two examples are very different. As a result, the *Decision* algorithm has selected a quite different strategy for the two bit rates: for low bit rates almost 20 % of the bitstream is devoted to the partition information whereas less than 10% is used for this type of information for higher bit rates. These figures illustrate one of the claims of *second generation coding* techniques stating that, at low bitrates, contour information becomes more important than at high bitrates.

4 CONCLUSIONS

In this paper, we have discussed the importance of temporal linking and joint optimization of the partition and of the coding strategy for segmentation-based video coding schemes. Temporal linking is an important feature allowing motion compensation of the partition information as well as of the texture. Moreover, it opens the door to content-based functionalities where objects have to be tracked in time. We have proposed a double-partition *Projection* that allows the linking of regions in consecutive partitions without any assumption on the type of homogeneity of the regions.

The joint optimization of the partition and of the coding strategy is necessary to reach a high coding efficiency. The notion of *Partition Tree* has been proposed as a technique to create a universe of regions structured in a suitable way for the optimization. It also relates these regions with previously coded regions. This approach provides some flexibility to the coding scheme: one can easily introduce a new coding technique and the decision step will modify the partition to take full benefit of the new coding technique. Content-based functionalities involving the selective coding of some areas is also very easy to implement. One has simply to track the regions of interest with the projection and to assign different weights to the distortions inside and outside the regions of interest.

Acknowledgment

This work has been supported by the European Commission: RACE Projects *MORPHECO* and *MAVT* and by the Spanish government: *CICYT TIC95-1022-C05-05*.

References

[1] J. Benois, L. Wu, and D. Barba. Joint contour-based and motion-based image sequences segmentation for TV image coding at low bit rate. In *VCIP'94*, pages 1074–1085, Chicago, USA, September 1994.

[2] P. Bouthemy and E. François. Motion segmentation and qualitative dynamic scene analysis from an image se-

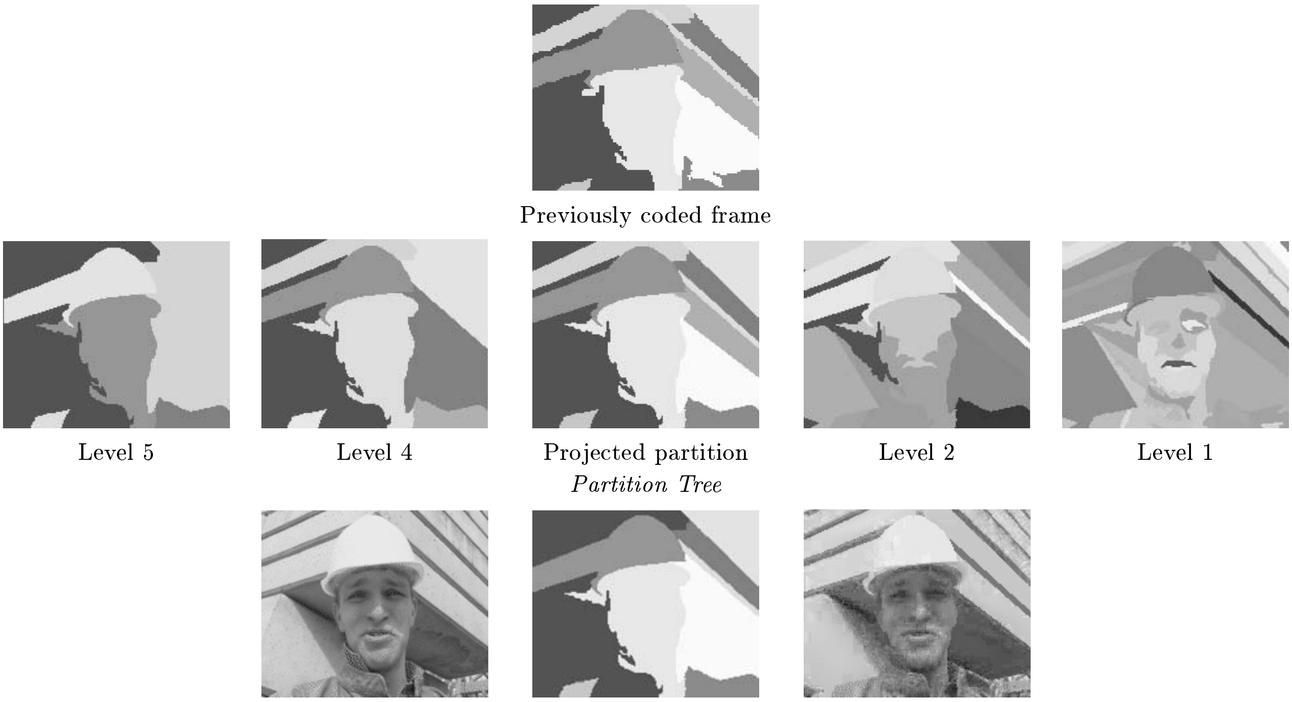

Figure 5: Example of inter-frame segmentation

quence. *Int. Journal of Computer Vision*, 10(2):157–182, 1993.

[3] C. Gu and M. Kunt. Very low bit-rate video coding using multi-criterion segmentation. In *IEEE ICIP'94*, volume II, pages 418–422, Texas, U.S.A., November 1994.

[4] F. Marqués, B. Marcotegui, and F. Meyer. Tracking areas of interest for content-based functionalities in segmentation-based video coding. In *IEEE ICASSP'96*, Atlanta, USA, May 1996.

[5] F. Marqués, V. Vera, and A. Gasull. A hierarchical image sequence model for segmentation: Application to object-based sequence coding. In *Proc. SPIE VCIP'94*, pages 554–563, Chicago, USA, Oct 1994.

[6] R. Morros, F. Marqués, M. Pardàs, and P. Salembier. Video sequence segmentation based on rate-distortion theory. In *SPIE VCIP'96*, volume 2727, pages 1185–1196, Orlando (FL), USA, March 1996.

[7] MPEG. MPEG-4 Proposal Package Description (PPD). Technical Report ISO/IEC JTC1/SC29/WG11, MPEG, July 1995.

[8] H.G. Musmann, M. Hőtter, and J. Ostermann. Object-oriented analysis-synthesis coding of moving images. *Signal Processing, Image Communication*, 1(2):117–138, October 1989.

[9] M. Pardàs and P. Salembier. Joint region and motion estimation with morphological tools, *Second Workshop on Mathematical Morphology and its Applications to Signal Processing*, Fontainebleau, France, pp. 93-100, Sept. 94.

[10] M. Pardàs, P. Salembier, F. Marqués, and R. Morros. Partition tree for segmentation-based video coding. In *IEEE ICASSP'96*, Atlanta (GA), USA, May 1996.

[11] K. Ramchandran and M. Vetterli. Best wavelet packet bases in a rate-distorsion sense. *IEEE Trans. on Image Processing*, 2(2):160–175, April 1993.

[12] E. Reusens. Joint optimization of representation model and frame segmentation for generic video compression. *EURASIP Signal Processing*, 46(11):105–117, September 1995.

[13] P. Salembier. Morphological multiscale segmentation for image coding. *EURASIP Signal Processing*, 38(3):359–386, September 1994.

[14] P. Salembier, F. Marqués, M. Pardàs, R. Morros, I. Corset, S. Jeannin, L. Bouchard, F. Meyer, and B. Marcotegui. Segmentation-based video coding system allowing the manipulation of objects. *IEEE Trans. on Circuits and Systems for Video Technology*, Submitted as an invited paper.

[15] P. Salembier, L. Torres, F. Meyer, and C. Gu. Region-based video coding using mathematical morphology. *Proceedings of IEEE (Invited paper)*, 83(6):843–857, June 1995.

[16] Y. Shoham and A. Gersho. Efficient bit allocation for an arbitrary set of quantizers. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, 36:1445–1453, September 1988.

[17] P. Willemin, T. Reed, and M. Kunt. Image sequence coding by split and merge. *IEEE Trans. on Communications*, 39(12):1845–1855, December 1991.