A SEGMENTATION-BASED CODING SYSTEM ALLOWING MANIPULATION OF OBJECTS (SESAME)

F. Marqués [†], *P. Salembier* [†], *M. Pardàs* [†], *R. Morros* [†],
I. Corset [‡], S. Jeannin [‡], B. Marcotegui [††], F. Meyer [††]

[†] Universitat Politècnica de Catalunya, Barcelona, Spain
[‡] Laboratoires d'Electronique Philips, Limeil-Brevannes, France
[††] Ecole des Mines de Paris, Fontainebleau, France

ABSTRACT

In this paper, we present a coding scheme that achieves, for each image in the sequence, the best segmentation in terms of Rate-Distortion theory. It is obtained from a set of initial regions and a set of available coding techniques. The segmentation combines spatial and motion criteria. It selects at each area of the image the most adequate criterion for defining a partition in order to obtain the best compromise between cost and quality. In addition, the proposed scheme is very suitable for addressing content-based functionalities.

1. INTRODUCTION

Among the so-called second generation coding techniques [6], there is an increasing interest in segmentation-based video coding approaches. This interest is due to the fact that these techniques have a strong potential for increasing the coding efficiency as well as an inherent capability for handling content-based functionalities.

Techniques reported in the literature mainly differ in the relative importance they assign to spatial or motion information. In [19, 9, 17], segmentation uses spatial homogeneity criteria, whereas in [11, 2], motion information is used as the main homogeneity criterion. Examples of segmentation techniques involving both criteria can be found in [4, 1].

The combination of both criteria offers a large number of advantages. The spatial homogeneity allows a very accurate definition of the regions and handles the appearance of new objects. The motion homogeneity is important to limit the number of regions and therefore the coding cost associated to the partition. This limitation is generally achieved 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.

Segmentation techniques combining regions with motion and spatial homogeneity should profit from all the previous advantages. As main goals, segmentation techniques should:

- correctly represent the objects in the scene while allowing the improvement of the coding efficiency.
- enable the temporal tracking of objects in order to really address content-based functionalities.

In this paper, we present a segmentation-based coding scheme that allows manipulation of objects in the scene [16]. This scheme has been developed in the MPEG4 framework [3] and supported by the European Community through the MAVT and MoMuSys projects of the RACE and ACTS programs, respectively. In the context of MPEG4, the work carried out in the MAVT project aimed at the development of tools and complete schemes for video data coding. It resulted in several proposals that were submitted to the MPEG4 community in November 1995. In the same context, the work that is being performed in MoMuSys deals with the definition and improvement of the MPEG4 Verification Model.

The system takes advantage of homogeneity in both gray level and motion. It selects at each area of the image the most adequate criterion for defining a partition in order to obtain the best compromise between cost and quality. To do so, multiple partitions are generated for every image in a hierarchical way. Then, a decision step choses the best combination of regions from the different partitions as well as which coding techniques should be used in each region in order to obtain the most efficient coding.

The proposed scheme is very suitable for addressing content-based functionalities [8]. This capability is based on the fact that there is a tracking step which defines the time evolution of the regions. Moreover, the various parts of the algorithm can be adapted to aim at such functionalities.

The presentation of this coding scheme is structured as follows. In Section 2, the main parts of the algorithm are analyzed. Section 3 deals with the modifications that have to be introduced in the algorithm to allow content-based functionalities. Finally, results are presented in Section 4.

2. THE SESAME CODING ALGORITHM

Most of the segmentation-based coding schemes discussed in the introduction involve an analysis or segmentation step followed by a coding step. With this coding strategy there is no strong relation between the two steps. However, high coding efficiency requires a strong interaction between the segmentation and coding steps. In [14], a solution for an optimal definition of a partition and of the set of coding techniques to be used in each region is proposed. The link between the analysis and coding steps is achieved by using a *Decision* block. The goal of the analysis is now to create a universe of regions out of which the optimal partition will be built.

To allow content-based functionalities, this universe of regions has to enable object tracking. In this paper, we present the concept of *Partition Tree* to solve the problem of bit allocation while maintaining the temporal link defined by the projection. A block diagram of the coding scheme is

0-7803-3258-X/96/$5.00

shown in Figure 1. It is composed of three different types of functions: *Partition functions*, *Bit allocation functions*, and *Coding functions*. Here, a summarized description of these blocks is given. For a complete description of every block, the reader is referred to the references that are given herein.

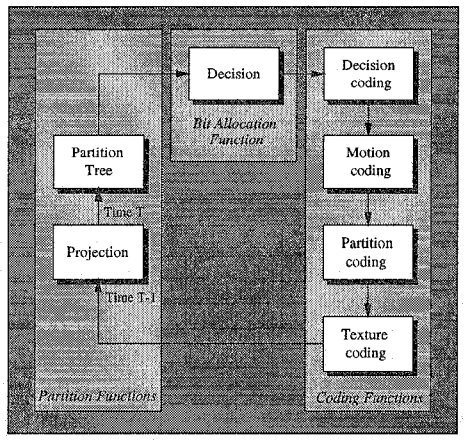

Figure 1: Block diagram of the encoding algorithm

2.1. Partition functions

This set of function produces the universe of regions from which the final partition will be constructed. This final partition has to ensure a temporal link with the previous partition. Such a temporal link is made by the *Projection* whereas the universe of regions is created by the *Partition Tree*.

The *Projection* block [12] ensures the time evolution of the regions in the partition. The coded partition of the previous frame P_{T-1} is projected onto the current frame T in order to define the partition in this frame. The approach presented in [12] has been further improved so that temporal stability is increased [7]. In addition, a double-partition projection approach has been proposed which is able to cope with partitions combining regions with motion and spatial homogeneity [8]. The double-partition projection is based on the re-segmentation of P_{T-1} using spatial criteria. This way, a fine partition, where all regions have spatial homogeneity, is obtained. This fine partition is then projected and, by re-labeling, the final projected partition is defined. Figure 2 illustrates the concept of the double-partition approach.

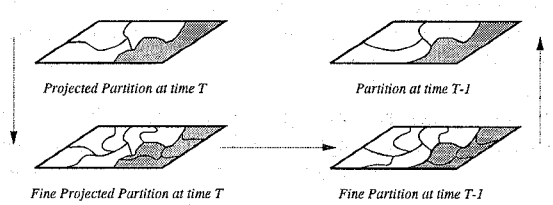

Figure 2: *Projection* using a double-partition approach

The *Partition Tree* block [13] deals with all possible modifications of the partition topology; that is, the detection of new regions appearing in the scene, as well as the merging of regions whose evolution is similar. This is implemented by defining the so-called *Partition Tree* which is a hierarchy of partitions constructed from the projected partition.

It contains below the projected partition a set of finer partitions. 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 that splits the regions of a given level to produce the regions of the next lower level [15]. Note that the procedure is purely spatial: it does not take into account any motion information.

Above the projected partition, there are several levels of coarser partitions that are built by merging regions with similar motion. This process reduces the partition coding cost and goes closer to the notion of objects.

The *Partition Tree* block also includes the estimation of the motion parameters for all regions in the tree [5]. The motion of each region is represented by an affine model, i.e. defined by two polynomials of order one.

The estimation itself is done by differential methods [18]. The objective function is the MSE between the original region and its current prediction. The Gauss-Newton method is used as minimization algorithm. For each region, the initialization tests several sets of motion parameters and selects the one that produces the lowest MSE. In particular the set of parameters estimated for this region in the previous frame, as well as the set of parameters of regions located at the same position but on other levels of the *Partition Tree* are tested. Finally, the estimation is done following a multi-scale coarse-to-fine strategy based on a Gaussian pyramid.

2.2. Bit allocation functions

The *Decision* block [10] has as objective the definition of the coding strategy; that is, the choice of the final partition, by combining the different regions obtained in the previous step, as well as the assignment of a coding technique to each one of these regions. Thus, relying on the Rate-Distorsion theory, the *Decision* has to select the best strategy in terms of regions and coding techniques among the possibilities provided by the *Partition Tree* and the available coding techniques.

The *Decision* process relies on the concept of *Decision Tree* [14, 10]. This tree lists in a compact and hierarchical structure all the possible coding choices. The *Partition Tree* defines the choices in terms of regions. The list of coding techniques deals with the actual coding of these regions.

To define the coding strategy in the Rate-Distortion sense, the *Decision Tree* should convey the information about the coding cost (Rate measured in number of bits) and quality (Distortion assessed by the Squared Error) of all possible coding techniques. In practice, each region of the *Partition Tree* is coded by all techniques (with various quality levels and either in intra-frame or inter-frame mode) and the rate and distortion values are stored in the *Decision Tree*.

The optimization relies on the technique discussed in [14]. 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. The definition of the optimum λ parameter can be done with a gradient search algorithm.

Figure 3 illustrates with a simple example the *Decision* process. In this example, the *Partition Tree* contains 2 partitions above and below the projected partition. In turn, the *Decision Tree* is built by testing the set of coding techniques $\{C1, C2, ..., Cn\}$. The final partition is obtained by combining regions from almost all the levels of the *Partition Tree*.

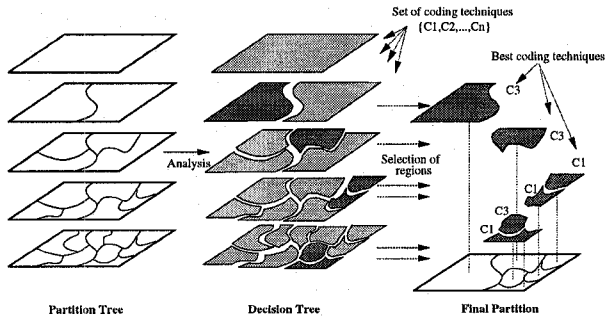

Figure 3: Example of *Decision* process

2.3. Coding functions

The *Coding* block [16] performs the coding of the necessary information in order to be able to reconstruct the encoded image in the receiver side. It deals with the encoding of the coding strategy (*Decision coding*), the motion information (*Motion coding*), the partition (*Partition coding*) and the pixel values (Gray level & color) called Texture. To have an efficient representation, both the partition and the texture are motion compensated. This explains why the *Motion coding* block is located before the *Partition* and *Texture coding* blocks. The complete description of the various coding algorithms that are used in this technique can be found in [16].

3. CONTENT-BASED FUNCTIONALITIES

The basic steps of the SESAME algorithm have to be modified to allow the algorithm to track areas of interest as well as to address content-based functionalities [8]. Each functionality will demand specific modifications. Here, we present the main modifications for content-based selective coding.

- *Projection:* In order to allow content-based functionalities, it has to be done following the double partition approach commented above.
- *Partition Tree:* It should be created having in mind the constraint introduced by the projected areas of interest. This translates into preventing partitions above the projected partition to merge regions with similar motion if they do not belong to the same area of interest.
- *Decision:* It should yield a coding strategy leading to a lower distorsion within the selected areas of interest than in the other areas of the image. To obtain this selective coding, the bit allocation strategy used in the basic coding algorithm is slightly modified. In this case, if a target bit rate has to be reached, selective coding is implemented by simply multiplying the distorsion in the regions forming the areas of interest by a given factor.
- *Coding:* It does not need to be modified since the coding techniques used in this functionality are the same as those used in the general case.

4. RESULTS

In Table 1, the results of coding the sequences *Foreman* and *Children* with two different bit rates are presented. The *Decision* step has selected a quite different strategy for the two bit rates: for low bit rates almost 20 % of the bit stream is devoted to the partition information whereas less than 10% is used for this type of information for higher bit rates.

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

In Figure 4, the procedure for obtaining an inter-frame segmented and coded image is illustrated. In the first row, the previous partition is presented, whereas the second row shows the *Partition Tree*. The third row presents the original image, the final segmentation and the coded images.

Note the presence of regions from different levels of the *Partition Tree* in the final segmentation. For example, the region in the upper-left corner corresponds to Level 5 in the *Partition Tree*, whereas the regions forming the lowest part of the building in the right-hand side belongs to Level 4. Finally, there is a region of the building in the right-hand side that comes from Level 2.

The temporal link can be observed when comparing the labels present in the previous and final partitions. Moreover, the final partition contains regions with spatial homogeneity (e.g.: the helmet is jointly segmented with a part of the building) as well as motion homogeneity (e.g.: the face is gathered in a single region).

5. 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 *Visual Commun. and Image Processing*, pages 1074–1085, USA, Sept. 1994.

[2] P. Bouthemy and E. François. Motion segmentation and qualitative dynamic scene analysis from an image sequence. *International Journal of Computer Vision*, 10(2):157–182, 1993.

[3] LEP, UPC, CMM. Segmentation-based coding system allowing the manipulation of objects (SESAME). Technical Report ISO/IECJTC1/SC29/WG11/MPEG95/408, Nov. 1995.

[4] C. Gu and M. Kunt. Very low bit-rate video coding using multi-criterion segmentation. In *First IEEE International Conference on Image Processing*, volume II, pages 418–422, Texas, U.S.A., November 1994.

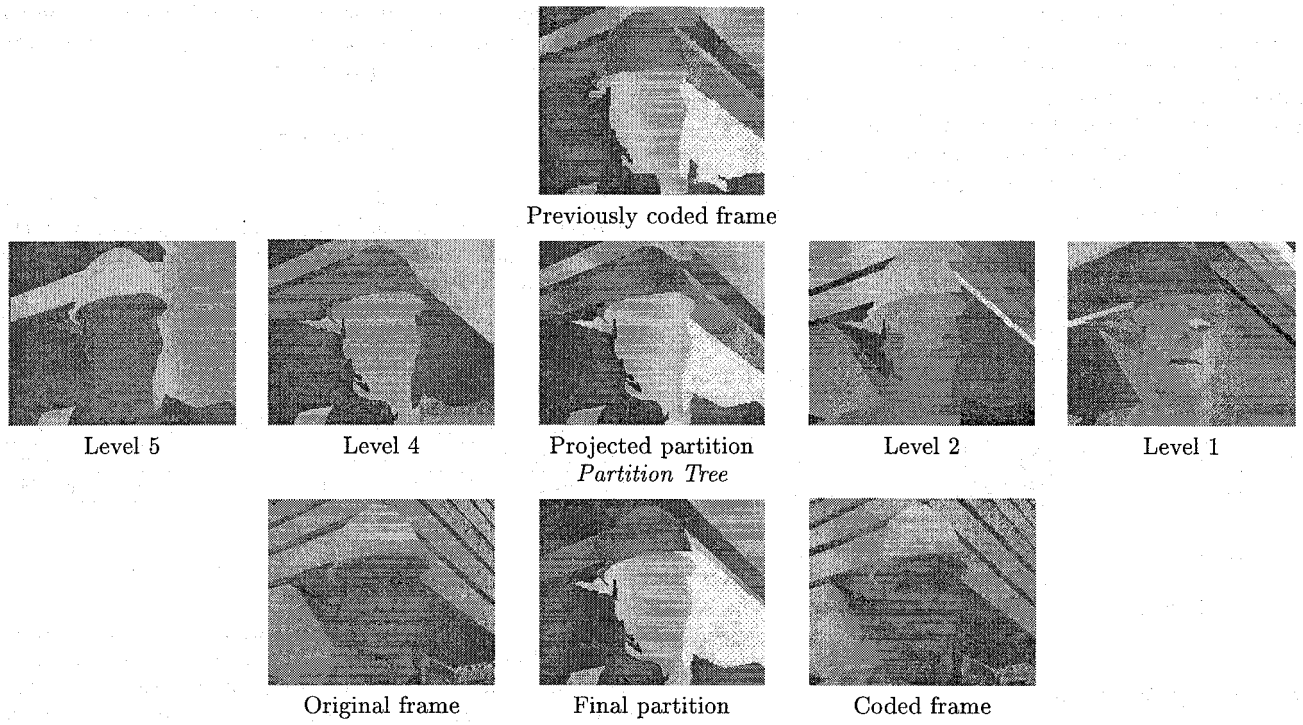

Figure 4: Example of inter-frame segmentation

[5] S. Jeannin. On the combination of a polynomial motion estimation with a hierarchical segmentation based video coding scheme. In *IEEE Int. Conference on Image Processing*, Lausanne, Switzerland, September 1996.

[6] M. Kunt, A. Ikonomopoulos, and M. Kocher. Second generation image coding techniques. *Proceedings of the IEEE*, 73(4):549–575, April 1985.

[7] F. Marqués. Motion stability in image sequence segmentation using the watershed algorithm. In P. Maragos, R. Schafer, and M. Butt, editors, *Mathematical morphology and its applications to image and signal processing*, pages 321–328. Kluwer Academic Publishers, May 1996.

[8] F. Marqués, B. Marcotegui, and F. Meyer. Tracking areas of interest for content-based functionalities in segmentation-based video coding. In *Int. Conference on Acoustics, Speech and Signal Processing, ICASSP'96*, pages II.1224–II1227, Atlanta, USA, May 1996.

[9] F. Marqués, V. Vera, and A. Gasull. A top-down 3D image sequence segmentation technique controlled by morphological tools. In *Proc. 7th European Signal Processing Conf. EUSIPCO-94*, pages 415–418, Oct 1994.

[10] R. Morros, F. Marqués, M. Pardàs, and P. Salembier. Video sequence segmentation based on rate-distorsion theory. In *Visual Communication and Image Processing*, pages 1185–1196, Orlando, USA, April 1996.

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

[12] M. Pardàs and P. Salembier. 3D morphological segmentation and motion estimation for image sequences. *EURASIP Signal Processing*, 38(1):31–43, September 1994.

[13] M. Pardàs, P. Salembier, F. Marqués, and R. Morros. Partition tree for segmentation-based video coding. In *International Conference on Acoustics, Speech and Signal Processing*, volume IV, pages 1982–1985, May 1996.

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

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

[16] 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*, Invited paper.

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

[18] H. Sanson. Toward a robust parametric identification of motion on regions of arbitrary shape by non-linear optimization. In *Proceedings of ICIP'95*, volume I, pages 203–206, October 1995.

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