

ACTIVE MESH CODING AND RATE-DISTORTION THEORY

Philippe Salembier, Eva Martí and Montse Pardàs

Dept. of Signal Theory and Communications
Universitat Politècnica de Catalunya
Campus Nord - Mòdul D5
C/ Gran Capità, 08034 Barcelona, Spain
Tel: (343) 401 74 04, Fax: (343) 401 64 47
E-mail: philippe@gps.tsc.upc.es

ABSTRACT

This paper presents a video coding scheme for very low bit rate applications. The coding approach relies on *active meshes* and can be viewed as a particular case of region-based coding. The active mesh is used to efficiently represent and code the various regions of the scene and the motion information. The variation of the mesh topology as well as the strategy for coding the synthesis error are defined by an optimization technique following the Rate-Distortion criterion.

1. INTRODUCTION

In the framework of very low bit rate video coding, there is an increasing interest in second generation image compression techniques [1], in particular region-based coding schemes. For video sequences, these techniques are receiving a particular attention within the *MPEG4* framework because of their potential efficiency for compression and also because of their ability to handle content-based functionalities.

In this paper, we are interested in a video coding algorithm able to efficiently represent the image content and also to follow the temporal evolution of the regions. This approach implies the definition of a signal dependent partition that has to be transmitted to the receiver. Moreover, time tracking of regions is an important feature if one wants to be able to efficiently relate and compensate the information (partition, color, etc) between two successive frames and also to deal with content-based functionalities [2].

A first attempt in defining such an algorithm can be found in [3]. The proposed coding algorithm combines a spatial segmentation of the sequence with a motion compensation of the transmitted information. On the one hand side, the spatial segmentation is used to get a general scheme able to deal with any kind of sequences and scene changes. On the other hand side, motion information is used to increase the coding efficiency by compensation of the spatial information that has to be transmitted (partition and color). An analysis of the results given in [3] shows that 55% of the bit-stream is devoted to the partition information, 35% to the color and the remaining 10% to motion. To improve these results, the number of bits devoted to the partition has to be reduced. Two solutions have been reported:

- A first solution is to reduce the number of regions and to process (that is to compensate), as a single region, sets of regions that follow the same motion. This leads to a *spatio-temporal* segmentation where some regions are homogeneous in gray level and others are homogeneous in motion. To efficiently define the *spatio-temporal* segmentation a careful analysis of the bit allocation problem has to be done. The work reported in [4] shows how to solve the problem of joint optimization of a partition and a set of coding techniques to be used within each region. The approach can be viewed as an optimal solution in the rate-distortion sense of the bit allocation problem. The approach discussed in [4] defines a partition for each frame independently, therefore, there is no time tracking possibility. The work was extended in [5, 6] to relate the partitions of the successive frames and to actually track regions. To this end, the concept of *Partition Tree* was proposed in [6]. The resulting scheme [7] gives very interesting results. However, for very low bit rates, the partition coding technique used in [7] is too expensive in terms of bits and the algorithm has a tendency to merge regions.
- A second solution to improve the results of [3] can be found in [8]. Here, the goal is to define a scheme that follows the same philosophy as the one proposed in [3] but where much more color information is sent to the receiver. To reach this goal, it is proposed to work with simple shapes as triangles. This idea leads to the definition of an active triangular mesh coding scheme.

This paper presents a scheme that takes into account the strong points of the two previous solutions: on the one hand side, active meshes are interesting because they very efficiently handle the partition information and, on the other hand side, optimization in the rate-distortion sense of the partition and of the coding strategy provides an optimum bit allocation. This combination leads to an optimum active mesh coding scheme. The organization of this paper is as follows: Section 2 defines the structure of the algorithm whereas Section 3 describes and illustrates the various steps. Finally, Section 4 is devoted to the conclusions.

2. STRUCTURE OF THE ALGORITHM

The main steps of the algorithm are described in Fig. 1. It is conceptually equivalent to the one proposed in [7]. There are three sets of functions: The *Mesh Functions* are devoted to the definition, the time tracking (*Projection*) and the to-

This work was supported by the European Community through the VIDAS project of the ACTS program and the Spanish government through the TIC95-1022-C05-05

pological modifications (*Mesh tree*) of the mesh. The *Coding Functions* send to the receiver the necessary information to reconstruct the coded image. Finally, the definition of the coding strategy leading to an optimum definition of the active mesh and of the coding techniques to be used within each region is done in the *Bit Allocation Function*. Note that, in general, the optimum mesh is not triangular but made of arbitrary polygons. Let us describe the block of Fig. 1.

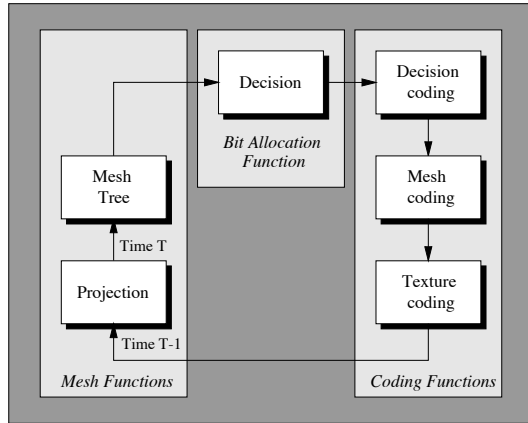


Figure 1: Active triangular mesh coding scheme

3. PROCESSING STEPS

3.1. Mesh functions

The *Projection* defines the time evolution of the mesh. It tracks regions in time: based on the mesh defined for frame $T - 1$, this block makes an estimation of the mesh corresponding to frame T . The approach consists in defining the motion of each node of the mesh as described in [8]. Note that the mesh topology (number of polygons, adjacency relationship) is not modified within this block.

The *Mesh Tree* is the equivalent of the *Partition Tree* of [6] but for polygonal meshes and not for arbitrary regions. The objective is to construct a set of mesh proposals from the projected mesh, as shown in Fig. 2. These mesh proposals define a reduced set of polygons that are candidate to belong to the final mesh.

The *Mesh Tree* is a hierarchical set of meshes. This means that if one node or one edge is present at a given level, this node or edge is also present at the same position at lower levels. The *Mesh Tree* is made of two parts:

- 1 Below the projected mesh (the mesh resulting from the projection), a set of “finer” meshes can be found. They involve smaller polygons that are able to represent finer details of the image. The creation of these lower levels can be viewed as several segmentation steps. The segmentation works as follows: first, a segmentation priority is assigned to each polygon based on its variance so that polygons of high variance are segmented first. Then, each selected polygon is either triangularized if it is not a triangle or segmented by the techniques proposed in [8] (see Fig. 3).
- 2 Above the projected mesh, polygons are constructed by merging polygons of lower levels. They represent large re-

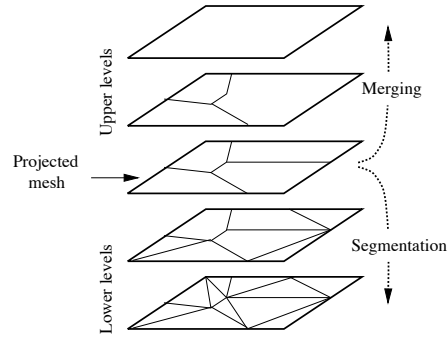


Figure 2: Mesh Tree

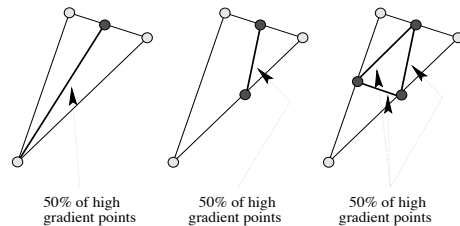


Figure 3: Segmentation of triangles

gions that can be processed (predicted) as a single region. The merging priority is defined by the increase of synthesis error obtained if two neighbor polygons are merged.

An illustration of the *Projection* and of the *Mesh Tree* can be seen in Fig. 4. The *Projection* effects can be seen in the center of the figure and the various meshes of the *Mesh Tree* are shown in the lower part. Note, in particular, how levels 1 and 2 are created by merging regions of the projected mesh (level 3) and how levels 4 and 5 are constructed by segmenting areas that are either textured or corresponding to transitions between objects (around the face).

This approach has been selected because, if the scene content has not been strongly modified, the projected mesh should be a good approximation of the optimal mesh. The *Mesh Tree* proposes to the decision a set of partitions made of polygonal regions that are “fluctuations” with respect to the projected mesh. The decision will construct the optimal mesh by selecting some polygons of the *Mesh Tree*.

3.2. Bit allocation function

The *Decision* block takes the proposals from the *Mesh Tree* and makes a decision on which polygons will belong to the final mesh, and which coding technique will be used within each region. Note that the final mesh is made of polygons coming from various levels of the *Mesh Tree*. Moreover, for each region, several intra-frame as well as inter-frame coding algorithms are proposed to the decision. The inter-frame coding techniques generally involve the synthesis (or compensation) of the current frame based on the previously coded frame and the coding of the prediction error.

The decision is based on an optimization algorithm following the rate-distortion criterion. As illustrated in Fig. 5,

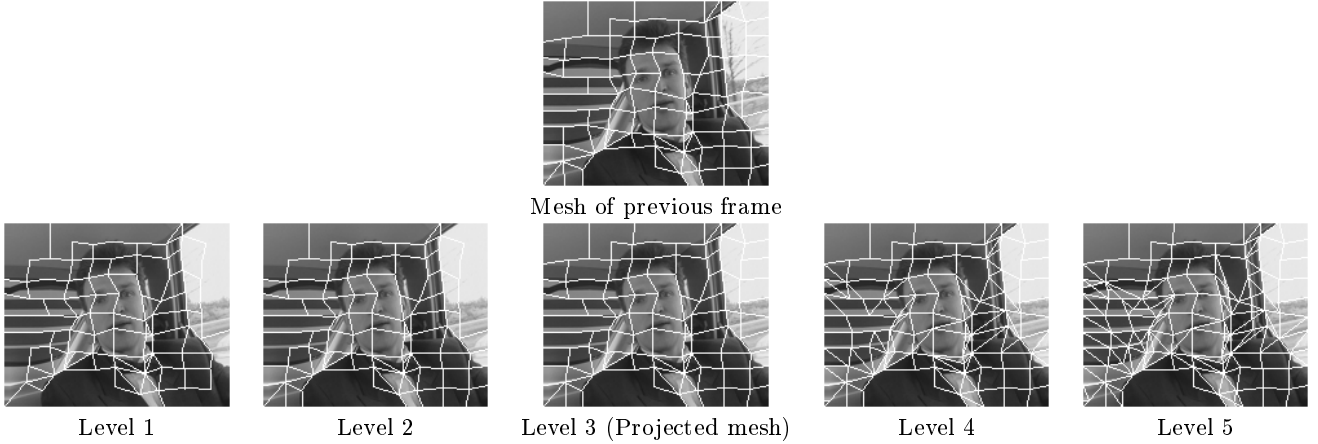


Figure 4: Example of projection and *Mesh Tree*

the algorithm analyzes each polygon of the *Mesh Tree* and, for each polygon, considers a set of possible coding techniques ($\{C_1, \dots, C_n\}$). This analysis results in a set of rates R and distortions D for each combination of polygons and coding techniques. Then, an optimization algorithm [4, 5] is used to find the best coding strategy by minimizing the distortion under the constraint of a limited budget. Note that the hierarchical structure of the proposed polygons is an important feature allowing the use of fast optimization techniques.

The *Decision* process is illustrated in Fig. 6. Based on the polygons proposals of Fig. 4, the *Decision* selects the mesh shown in Fig. 6.A and defines the coding strategy represented in Fig. 6.B. In Fig. 6.B, white areas indicate polygons that are coded in intra-frame mode. They correspond to parts of the image that are difficult to predict (window area). The remaining gray level values represent various coding techniques. A dark value indicates a technique with a very low number of bits per pixel (bpp) whereas a light gray value indicates a technique leading to a better quality but a higher number of bpp (shape adaptive DCT [9] with several levels of quantization was used in this test). See how bits are saved for the car areas with respect to more complex areas such as the face.

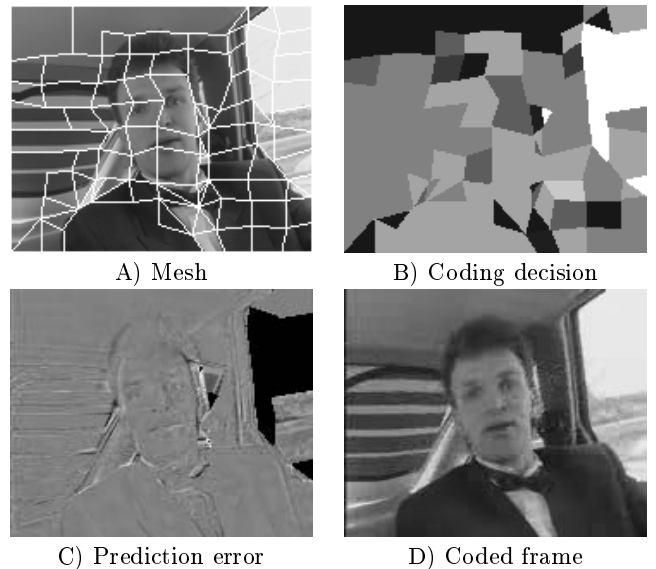


Figure 6: Decision and coding

3.3. Coding functions

The *Coding* blocks take the results of the decision block and send all the necessary information to the receiver. This information is composed of the coding strategy itself (which color coding technique is used in each region), the mesh evolution (motion of the nodes and introduction or elimination of polygons) and the synthesis error (see Fig. 6.C and D).

The synthesis (prediction) is defined by the nodes motion. To limit the complexity of the compensation, each polygon is triangularized and a motion model is defined in each resulting triangle. Since each triangle defines an affine motion field, the motion field of each polygon is piecewise affine.

3.4. Initialization of the mesh

The initialization of the mesh for the first frame is done also by using the concept of *Mesh Tree* and is illustrated in Fig. 7. The algorithm starts by a regular rectangular mesh (Fig. 7.A) and modifies the nodes position to get closer to high gradient points (Fig. 7.B). Based on this first mesh, a complete *Mesh Tree* is constructed and the *Decision* defines the first mesh (Fig. 7.C) and the coding strategy (in this case only intra-frame coding techniques are considered).

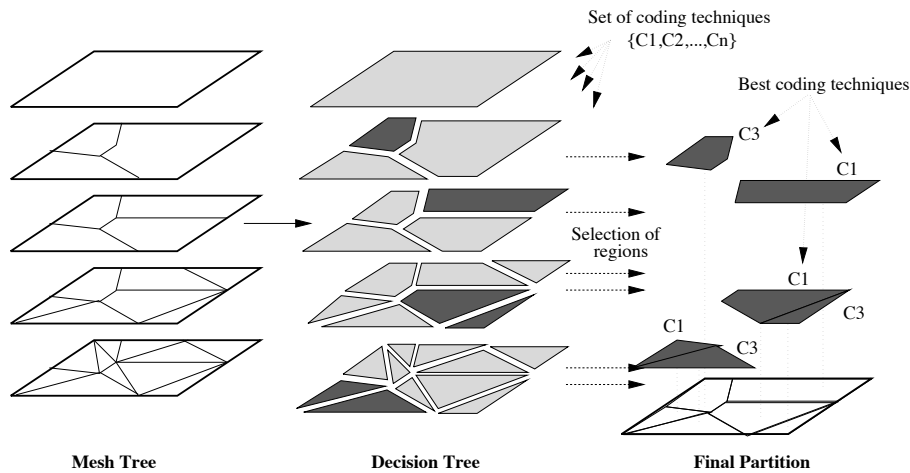


Figure 5: Decision Process

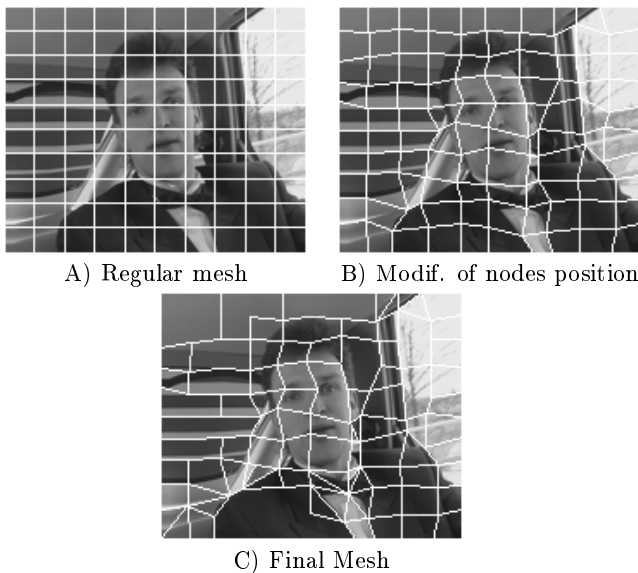


Figure 7: Mesh initialization

4. CONCLUSIONS

This paper has discussed the use of active meshes representations and of Rate-Distortion optimization for video coding. The coding approach can be viewed as a particular case of region-based coding. Active meshes are used to efficiently represent and code the various regions and the motion information. The variation of the mesh topology as well as the strategy for coding the synthesis error are defined by an optimization technique following the Rate-Distortion criterion. The concepts of *Projection* and *Partition Mesh* are efficient ways to track regions and to propose modifications of the mesh topology. Moreover, the proposals are structured in a way allowing the use of fast optimization algorithms.

5. REFERENCES

- [1] M. Kunt, A. Ikononopoulos, and M. Kocher. Second generation image coding techniques. *Proceedings of the IEEE*, 73(4):549–575, April 1985.
- [2] MPEG. MPEG-4 Proposal Package Description (PPD). Technical Report ISO/IEC JTC1/SC29/WG11, MPEG, July 1995.
- [3] P. Salembier, L. Torres, F. Meyer, and C. Gu. Region-based video coding using mathematical morphology. *Proc. of IEEE (Invited paper)*, 83(6):843–857, June 1995.
- [4] E. Reusens. Joint optimization of representation model and frame segmentation for generic video compression. *EURASIP Signal Processing*, 46(11):105–117, Sept. 1995.
- [5] R. Morros, F. Marqués, M. Pardàs, and P. Salembier. Video sequence segmentation based on rate-distortion theory. In *SPIE Visual Communication and Image Processing, VCIP'96*, volume 2727, pages 1185–1196, Orlando (FL), USA, March 1996.
- [6] M. Pardàs, P. Salembier, F. Marqués, and R. Morros. Partition tree for segmentation-based video coding. In *IEEE Int. Conference on Acoustics, Speech & Signal Processing, ICASSP'96*, Atlanta (GA), USA, May 1996.
- [7] 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.
- [8] P. Salembier and X. Ayuso. Very low bit rate video coding using active triangular mesh. In *IEEE International Conference on Acoustics, Speech & Signal Processing, ICASSP'96*, Atlanta (GA), USA, May 1996.
- [9] T. Sikora, S. Bauer, and B. Makai. Efficiency of shape-adaptive 2-D transforms for coding of arbitrarily shaped image segments. *IEEE Transactions on Circuits and Systems for Video Technology*, 5(1):59–62, February 1995.