TRACKING AREAS OF INTEREST FOR CONTENT-BASED FUNCTIONALITIES IN SEGMENTATION-BASED VIDEO CODING

Ferran Marqués[1] *Beatriz Marcotegui*[2] *Fernand Meyer*[2]

[1]Universitat Politècnica de Catalunya
Campus Nord UPC – Edifici D5
C/ Gran Capità s/n. 08034 Barcelona, Spain
E-mail: ferran@gps.tsc.upc.es
[2]Centre de Morphologie Mathématique
35, rue Saint Honoré
77305 Fontainebleau Cedex, France

ABSTRACT

This paper presents a technique for tracking areas of interest in the framework of segmentation-based video coding. The technique is independent of the type of segmentation technique used in the coding approach. Therefore, it can also be used in block-based coding schemes. The algorithm relies on a double segmentation of the image based on morphological tools. This double segmentation permits to obtain the position and shape of the previous area of interest in the current image. In order to demonstrate the potentialities of this algorithm, it is applied in a specific coding scheme so that content-based selective coding is allowed.

1. INTRODUCTION

In the framework of video coding, new coding schemes allowing functionalities such as content-based multimedia data access or content-based scalability are a very active research field [3]. These functionalities demand a description of image sequences in terms of objects or groups of objects that are referred to in the sequel as Areas of Interest (AI). AI should be defined automatically or by the user, even on the receiver side if an interactive application is being used. Once the AI have been defined, the algorithm should make possible their tracking, within the coding scheme, through the time domain. Such a possibility opens the door to content-based functionalities. For instance, it permits to encode the AI with a higher quality than the other parts of the image.

Classical object tracking techniques [13, 12] do not completely fulfill the requirements of this application since they do not account for the special characteristics of video coding functionalities. That is, classical object tracking techniques do not need to interact with coding algorithms. In the framework of video coding functionalities, zones of the image resulting from the AI tracking algorithm should not only correspond to the objects to be tracked, but also they should be suitable for coding purposes. In addition, tracking algorithms have to be able to cope with special constraints of the coding algorithm. For instance, they must not loss track of the AI when a frame in the sequence is coded in intra-mode.

In this work, a technique for tracking areas of interest for content-based functionalities in segmentation-based video coding is presented. Segmentation-based video coding schemes can use different homogeneous criteria to perform the segmentation [6] (e.g.: motion homogeneity, gray level homogeneity, special type of texture homogeneity, etc) or even a combination of some criteria [1]. In order to have an AI tracking technique independent of the type of segmentation criteria used in the coding approach, a double segmentation is applied, using the watershed algorithm [7].

*This work has been partially supported by the RACE Project 2072 (MAVT) of the European Union

This paper is structured as follows. In Section 2, the technique for tracking areas of interest within a sequence is proposed and some results are presented. In order to further validate this technique, Section 3 describes the way to introduce it in a specific segmentation-based coding scheme in order to enable content-based selective coding. In addition, some results of this application are given. Finally, some conclusions are derived in Section 4.

2. DOUBLE PARTITION APPROACH FOR TRACKING AREAS OF INTEREST

Usual segmentation-based coding approaches utilize motion or/and gray level homogeneity in order to compute the image partitions. Such partitions may be in contradiction with the AI obtained by the tracking technique. Let's illustrate this concept with an example. In Figure 1, three frames of the sequence *Foreman* are shown. Let us assume that the AI marked by the user is the head of the person. In some frames, a segmentation technique relying on gray level criteria will merge the helmet with the building, since they have very close gray level values. In turn, if the segmentation is based on motion information, some zones of the background may be jointly segmented with some regions of the face, in the case of global motion of the camera.

To solve this problem, a tracking technique relying on a double partition approach is proposed. This technique uses two different levels of partition: a coarse level partition which is related to the coding scheme and a finest level partition which contains a rich description of the current image and allows the tracking of the AI. Both segmentations are carried out in parallel and, for each image, the coarse partition constrains the finest one. That is, all contours in the coarse partition are also present in the finest one. The ultimate goal is to obtain a final partition containing the necessary regions to efficiently code the image, as well as the necessary regions to correctly track the AI.

2.1. Creation of the finest partition

Let us assume that frame n in a sequence has been already coded and, therefore, its partition is available. This partition should contain a region or a set of regions that defines the AI to be tracked. The first step to obtain the AI in frame $(n+p)$ is to produce a rich partition of the previously coded frame n. This rich partition is obtained by splitting the coded partition into several regions. The segmentation procedure is constrained by the coarse partition already obtained for frame n in the coding procedure. This is done by using a constrained watershed algorithm [11].

The process of creating this rich partition is purely intramode and, therefore, no motion information is used at this stage. The homogeneity criteria that are used only deal with size and contrast information [11]. This reach partition is created in order to make easier the tracking of the AI. That is, this procedure allows obtaining a partition with

Figure 1. Original frames number 80, 105 and 185 from the sequence *Foreman*

good features for AI tracking purposes, independently of the type of partition used at the coding level.

2.2. Projection of the finest partition

The finest partition of frame n is projected onto the current frame $n+p$, so that a rich partition of frame $n+p$ is achieved. Several techniques can be used in order to carry out the partition projection [4, 6, 9]. In this work, the projection step uses the segmentation algorithm presented in [9]. That is, the motion between both frames is roughly estimated using a block-matching algorithm [2] and applied for compensating the partition of frame n. After compensation, a set of markers is obtained which approximate the position and shape of the previous regions in the current image.

In order to obtain the actual rich partition of frame $n+p$, a watershed algorithm is applied using information from both the previous and the current images. This algorithm introduces the concept of contour complexity in order to obtain regions that are easy to code while representing correctly the objects in the scene [9]. In addition, the segmentation algorithm presented in [9] has been modified in order to improve the temporal coherence of the regions through the sequence partition. This improvement utilizes a new cost function in the watershed algorithm that prevents projected regions to grow too far from the previous markers [5]. This increases the stability of the labels through the time domain and, therefore, allows a better tracking of the AI while decreasing the coding cost [5].

2.3. Obtaining the projected areas of interest

From the finest partition obtained for frame $n{+}p$, only those contours related to the coarse level partition of frame n have to be kept. This can be easily done since the watershed algorithm used in the projection step kepts track of the labels of each region. Therefore, an AI in the current frame $n{+}p$ can be obtained just merging those regions that belong to this AI in the previous frame n. The complete procedure is illustrated in Figure 2, where only an area of interest has been selected in the coded frame.

2.4. Frames in intra-mode

The previous technique can also be applied in the case of using an intra-frame mode to code the current frame, since coarse partitions are only used to constrain finest partitions. That is, label coherence through the time domain on the coarse partitions is not mandatory for computing the fine partition and, therefore, for tracking AI.

2.5. Results

An example of the results obtained with this technique is presented in Figure 3. In this example, the result achieved in the case previously presented is shown; that is, the tracking of the head of the man in the sequence *Foreman*.

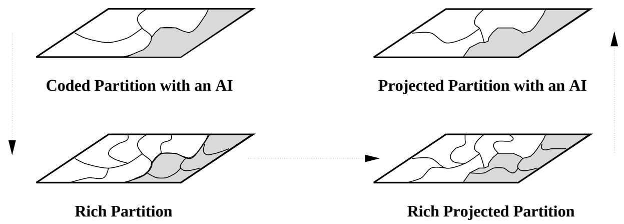

Figure 2. Example of tracking of an area of interest

Results are presented for frames 80, 105 and 185 which correspond with some of the frames with strongest motion in the sequence. Note that the algorithm is able to correctly track the head of the man, even if it is formed by two very different objects: the face and the helmet.

3. CONTENT-BASED SELECTIVE CODING

The above technique for tracking areas of interest has been introduced and tested in the coding scheme presented in [1]. The application that has been addressed is content-based selective coding; that is, given an initial selection of AI, the coding algorithm should be able to track them and to code them with better quality than the rest of the image.

The coding scheme of [1] performs, for each frame, a segmentation relying in both motion and gray level criteria. That is, it selects for each area of the image the coding technique that better fits the kind of data present in the image, tacking into account the data that have been previously transmitted. Therefore, the final partition for each image combines regions with motion homogeneity and regions with gray level homogeneity.

In order to address this functionality, the basic blocks of the coding algorithm have to be modified. In this section, a brief summary of the coding algorithm proposed in [1] is given in order to, afterwards, present the necessary modifications to allow content-based selective coding.

3.1. The SESAME coding algorithm

This coding scheme is composed of four main blocks: *Projection*, *Partition Tree*, *Decision*, *Coding*. In the sequel, a brief description of each block is given.

- **Projection [9]:** This block ensures the time evolution of the regions in the partition. This is a pure region tracking step. The coded partition of the previous frame n is projected onto the current frame $n+p$ in order to define the evolution of the partition in this frame.

Figure 3. The tracking of the head of the man in frames number 80, 105 and 185 of the sequence *Foreman*

- **Partition Tree [10]:** This block 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 contains below the projected partition a set of new partitions introducing new regions. In addition, above the projected partition, there is a set of partitions that is built by merging regions with similar motion.
- **Decision [8]:** This block 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.
- **Coding [1]:** This block performs the coding of the necessary information in order to be able to reconstruct the encoded image in the receiver side. The information is mainly related to motion, texture and partition features.

Figure 4 illustrates with a simple example the coding process. In this example, the *Partition Tree* contains 2 partitions above and below the projected partition. Moreover, the final partition is obtained by combining regions from almost all the levels of the *Partition Tree*.

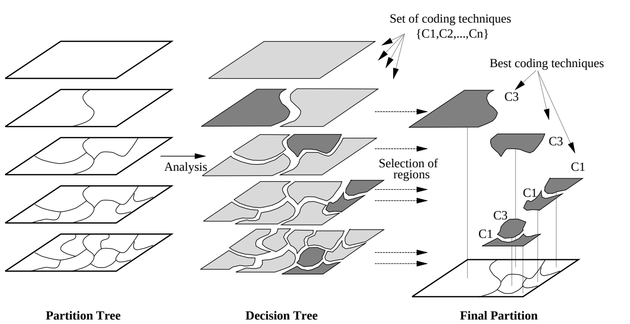

Figure 4. Example of tracking of an area of interest

3.2. Modifications on SESAME to allow content-based selective coding

The basic steps of the SESAME algorithm have to be modified in order to allow the algorithm to track areas of interest as well as to address content-based selective coding. The main modifications are the following ones:

- **Projection:** The partition of the previous frame, which should already contain a region or set of regions describing the AI, is projected following the double partition approach presented in the previous section.
- **Partition Tree:** The set of partitions that forms the *Partition Tree* should be created having in mind the constraint introduced by the projected AI. The different proposals of regions contain in the *Partition Tree* must not be in contradiction with the task of tracking the AI. 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. Such a merging would make impossible the separate tracking of these AI. In addition, the *Partition Tree* should help the task of coding with better quality the AI. This way, partitions below the projected partition, which are achieved by resegmenting the projected partition, try to refine the segmentation inside the AI rather than in those areas where a higher quality in the decoded image is not mandatory.
- **Decision:** This block should yield coding strategy leading to a lower distorsion within the selected AI than in the other areas of the image. In order 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 AI by a given factor.
- **Coding:** This block does not need to be changed since the coding techniques used in this functionality are the same as those used in the general case.

3.3. Results

Figure 5 presents the results of applying the previous algorithm to the sequence *Mother and Daughter*. The first row in Figure 5 presents three original frames of the sequences, whereas the second row shows the decoded frames. In this example, the head of the mother has been selected as area of interest to be tracked and selectively coded. The selective coding has been carried out by multiplying by a factor 10 the distorsion inside this area of interest. The whole sequence has been coded at 30 Kbits and 5 frames/sec.

The different quality obtained for the heads of both person in the scene should be highlighted. In addition, note that the evolution of the mother's face is correctly tracked. This tracking is done in spite of the fact that, between frames 150 and 200, the mother moves up and down her head which results in a partial hidding of her face. The presence of noise in the decoded image in the zones around the mother's face are due to the fact that these areas are related to

Figure 5. The tracking and selective coding of the head of the mother in frames number 0, 84 and 246 of the sequence *Mother and Daughter*

the background which is uncovered by the movement of the mother's head.

4. CONCLUSIONS

In this paper, a technique for tracking areas of interest in a segmentation-based video coding scheme has been proposed. The technique does not assume any type of specific partition and, therefore, it can also be applied to block-based coding schemes that use additional contour information such as masks or alpha planes.

In addition, the algorithm gives the possibility to define or refine the set of regions forming an area of interest. Since the segmentation involved in the tracking of AI relies on the concept of markers (that is, uses the watershed algorithm), user's interaction is feasible. Users can mark the areas of interest or refine the existing ones simply by roughly drawing dots or lines inside or around the areas of interest. Such draws can be introduced as markers of regions in the watershed algorithm so that the areas of interest are defined or improved [1].

Once a technique for tracking areas of interest is available, several content-based functionalities can be addressed. In this paper, the basic SESAME coding scheme has been modified in order to enable content-based selective coding. The necessary modifications for such a functionality only involve, in addition to the application of the algorithm for tracking AI previously presented, slight variations of the *Partition Tree* and *Decision* blocks.

The current work aims at the extension of this algorithm so that functionalities such as content-based scalability, both in the time and the space domains, can be addressed.

REFERENCES

[1] I. Corset et al. Technical description of SESAME. Technical Report MPEG 95/408, ISO/IEC JTC1/SC29/WG11, November 1995.

[2] R. Jain and A.K. Jain. Displacement measurement and its application in interframe coding. *IEEE Transactions on Communications*, 29(12):1799–1808, 1981.

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

[4] B. Marcotegui and F. Meyer. Morphological segmentation of image sequences. In J. Serra and P. Soille, editors, *Mathematical morphology and its applications to image processing*, pages 101–108. Kluwer Academic Publishers, 1994.

[5] F. Marqués. Motion stability in image sequence segmentation using the watershed algorithm. In *Mathematical morphology and its applications to image processing*. Kluwer Academic Publishers, May 1996.

[6] F. Marqués, M. Pardàs, and P. Salembier. Coding-oriented segmentation of video sequences. In L. Torres and M. Kunt, editors, *Video Coding: The second generation approach*, pages 79–124. Kluwer Academic Publishers, 1995.

[7] F. Meyer and S. Beucher. Morphological segmentation. *Journal of Visual Communication and Image Representation*, 1(1):21–46, September 1990.

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

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

[10] 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*, May 1996.

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

[12] I. Sezan and R. L. Lagendijk. *Motion Analysis and Image Sequence Processing*. Kluwer Academic Publishers, Boston, 1993.

[13] G. Tziritas and C. Labit. *Motion Analysis for Image Sequence Coding*. Elsevier Science B. V., The Netherlands, 1994.

TRACKING AREAS OF INTEREST FOR CONTENT-BASED FUNCTIONALITIES IN SEGMENTATION-BASED VIDEO CODING

Ferran Marqués[1] , *Beatriz Marcotegui*[2] *and Fernand Meyer*[2]

[1]Universitat Politècnica de Catalunya
Campus Nord UPC – Edifici D5
C/ Gran Capità s/n. 08034 Barcelona, Spain
E-mail: ferran@gps.tsc.upc.es

[2]Centre de Morphologie Mathématique
35, rue Saint Honoré
77305 Fontainebleau Cedex, France

This paper presents a technique for tracking areas of interest in the framework of segmentation-based video coding. The technique is independent of the type of segmentation technique used in the coding approach. Therefore, it can also be used in block-based coding schemes. The algorithm relies on a double segmentation of the image based on morphological tools. This double segmentation permits to obtain the position and shape of the previous area of interest in the current image. In order to demonstrate the potentialities of this algorithm, it is applied in a specific coding scheme so that content-based selective coding is allowed.