OBJECT-SCALABLE DYNAMIC CODING OF VISUAL INFORMATION

Corinne Le Buhan, Emmanuel Reusens and Touradj Ebrahimi
Signal Processing Laboratory
Swiss Federal Institute of Technology
CH-1015 Lausanne, Switzerland
e-mail: lebuhan@ltssg4.epfl.ch

ABSTRACT

This paper describes an extension of a dynamic video coding scheme to provide object scalable functionalities. As a particular instance of the dynamic coding concept, the coding scheme considered here jointly optimizes video data partition and representation modes. Indeed, as there exists no universal video coding method, dynamic coding insures the choice of the most efficient technique (for instance DCT, fractal or motion compensation) for each data segment. These data segments are themselves optimally partitioned within the original frame (respectively region of interest). An optimization algorithm achieves this joint data partition/representation modes selection to yield the best rate/distortion compromise within the available set of possible solutions, under a rate or distortion constraint. Such a dynamic coding algorithm designed for low bitrates was proposed to MPEG-4 first set of tests in November 1995. This paper describes the corresponding object scalable coding scheme.

1 DYNAMIC CODING OF VISUAL INFORMATION

1.1 Basic Principle

The dynamic coding concept [1][3][5] relies on the fact that no universal coding technique is available. While the current multimedia trend is rapidly developing, video compression has more and more to deal with heterogeneous environments and data, ranging from text areas to head and shoulders or complex textures. In order to insure the most efficient coding of such a variety of data, the dynamic coding approach is based on a competition among various coding techniques together with different possible image subdivisions (from simple 8*8 square blocks as in MPEG-2 to arbitrary shapes). The set of coding methods as well as the set of possible partitions are a priori selected as the best fitting ones with regards to application requirements (complexity, image quality, bandwidth, scalability...) and video contents (natural images, videotelephony, medical images, synthetic images...).

1.2 Implementation of a Dynamic Codec

As an illustrative application of the dynamic coding approach, EPFL has proposed a dynamic video codec to MPEG-4 first round of tests [2]. This codec is designed for low bitrates applications, from 10kbit/s to 112kbit/s[3][4]. Five main coding techniques are competing, namely DCT, fractal coding and bilevel clustering as intra modes, motion compensation possibly with DCT-based residual error coding and background model as temporal modes. Five different quantization schemes are investigated for DCT. The set of possible partitions is restricted to a generalized quadtree segmentation in order to meet the coding complexity/coding efficiency compromise.

For each frame, the first step consists in building the set of possible solutions. The frame is recursively divided into quadtree blocks. Each preselected technique is applied to each block and the corresponding parameters are stored: technique identifier, coding parameters, rate and distortion estimations. Once the set of solutions is built, the optimal solution selection is done in a rate-distortion sense. Either a maximum rate or a maximum distortion target is fixed. The problem is then to find the solution *B* in the set of possible solutions *S* that minimizes the distortion *D* (resp. the rate) corresponding to that rate (resp. distortion) budget R_{budget} :

$$\min_{B \in S} D(B) \textit{ subject to } R(B) \le R_{budget} \qquad \text{(Eq.1)}$$

This optimization problem is represented by Lagrange multipliers and solved as a two-step unconstrained problem consisting in (1) pruning the generalized quadtree from leaves to root by minimizing the Lagrangian cost function $J(\lambda)$ for a given Lagrange multiplier λ :

$$J(\lambda) = D(B) + \lambda R(B) \qquad \text{(Eq.2)}$$

and (2) finding the optimal Lagrange multiplier λ_c yielding the best rate-distortion compromise for the targeted rate (resp. distortion):

$$R(B^*) \le R_{budget},\ B^* \textit{ min. of } D(B) + \lambda_c R(B) \qquad \text{(Eq.3)}$$

The selected coding techniques data are extracted from the set of solutions and entropy encoded by an adaptive arithmetic coding scheme together with the pruned generalized quadtree description corresponding to the optimal frame partition.

The implemented codec uses the square error as distortion estimation for its additive properties. This square error is color-weighted in order to process YUV color frames:

$$D_{yuv} = SE_y + 0.3SE_u + 0.3SE_v \qquad \text{(Eq.4)}$$

The rate is estimated as the sum of information carried by the coding parameters according to their probability of occurrence. This probability is updated frame after frame once the selected frame parameters are arithmetic encoded to build the bitstream.

1.3 MPEG-4 Requirements

MPEG-4 video first round of tests [2] in November 1995 addressed three main issues: compression, object scalability and error robustness. The EPFL dynamic codec has been proposed to compress low bitrate sequences and to provide object scalable functionalities at 48kbit/s for sequences with low motion and low detail [3][4]. MPEG-4 requirements for object scalability are the following: given the segmented video sequences where different objects are distinguished (at very low bitrates, typically an head and shoulders and the associated background), the codec must be able to provide a scalable bitstream made of as many independent parts as there are predefined regions. The aim is to be able to transmit or to decode only a given region of interest, possibly to replace the background with another one, and to allow all sorts of object manipulations as required by multimedia applications. Moreover, it should be possible to code the various regions with different parameters according to their importance. For instance, spatial and temporal object scalability issues may be addressed by allocating some additional bitrate to enhance the spatial or temporal resolution of the main object of interest.

2 EXTENSION TO OBJECT SCALABLE COMPRESSION

2.1 Shape Coding

Dynamic coding has been extended to be able to code an arbitrarily shaped object instead of the whole scene. Basically, the segmentation which is provided by MPEG-4 together with the test sequences is used to mask the region to be coded. Shape information is encoded in order to provide an accurate frame reconstruction at the decoder side; if lossy shape coding is applied, the reconstructed shape is used to mask the original frame before building the set of solutions.

In the proposed codec, chain coding has been used to represent a label image [6]. A four-connected contour image is first built by considering as a contour pixel any pixel whose north, north-west or west neighbour has a different label. The corresponding chains are represented by four symbols and entropy coded by means of arithmetic coding combined with an improved statistical model: a high order Prediction by Partial Matching Markov model. Temporal redundancy is exploited by not resetting the model from frame to frame. On large shapes, quasi-lossless compression is performed by pre-filtering the shape with an open-close-close-open morphological operation, allowing segmentation noise partial removal. A performance down to 0.6bit/contour pixel is achieved on smooth head and shoulders such as *'Akiyo'* . In any case, the shape coding overhead is less than 5% (Table 1) for type A sequences (2 regions) coded at 48kbit/s in QCIF/5Hz.

2.2 Set of Solutions Building

The set of solutions is built by allowing the recursive quadtree descent within the frame only on blocks that overlap the preselected region. This procedure results in a generalized quadtree where mixed parent/children configurations represent the object boundaries. Exterior (parent) blocks do not require any information from the bitstream, since the decoder is able to identify them from the available shape data. Lastly, the distortion estimation calculated for each block in the set of solutions is adapted to take into account only pixels within the region of interest.

2.3 Adaptation of Block-Based Techniques

Within the object of interest, block-based techniques are applied as usual to the variable sized quadtree blocks. At the boundaries however, these techniques need to be adapted to deal with arbitrary shape when possible, in order to gain efficiency and avoid reference to outside regions.

Block matching based techniques such as fractal and motion estimation must be forced to refer only to the region of interest interior, in order to insure the possible independent decoding of the different regions. Search windows partially overlapping the object under coding are also considered, by calculating the error measure only on interior pixels. This restricts the corresponding search space, but on the other hand makes it more significant with regards to the data under coding: for instance, in the case where the object of interest has a uniform motion, block matching motion vectors at the edges are not biased by the still background.

The block-based bilevel clustering process partitions the block data into two centroids, corresponding to a YUV color, and approximate each pixel with the closest centroid. Run-length coding is then used to code the pixels distribution. The technique is adapted to deal with arbitrary shapes by taking into account only pixels belonging to the region under coding in the clustering process. This avoids biased centroids. Exterior pixels are associated to the strongest centroid, in order to maximize the correlation. Since the shape information is available at the decoder side, a possible improvement, at the cost of a higher complexity, would be to only run-length encode the interior pixels and use the shape information at the decoder to retrieve their original location in the block.

The block-based DCT is directly performed on the gray-masked blocks. A padding operation may be applied prior to DCT coding in order to make the exterior pixels more correlated with the interior data, such as in the current MPEG-4 VM [7]; however, the corresponding technique only uses one-order linear prediction from boundary pixels, which often are very noisy in natural images due to imperfect segmentation. This may yield an insufficient correlation gain. Other padding schemes are currently investigated in the framework of MPEG-4 as well as an arbitrarily shaped DCT scheme [8]. Related Core Experiments in MPEG-4 will help determining which method is the most efficient. Its introduction in the object scalable dynamic coding scheme may then bring another improvement.

2.4 Rate-Distortion Optimization

The optimization process allows to allocate a different rate (resp. distortion) target to the object according to its interest. It is also possible to use different quantizing parameters or preselected coding techniques depending on which object is coded. The optimization algorithm itself remains the same, but operates now on a generalized (pre-pruned) quadtree instead of a simple quadtree. It is also possible to code the main region(s) under a constant quality constraint, and then allocate the remaining bitrate to the background on a frame by frame basis.

2.5 Object-Scalable Bitstream

The object scalable bitstream is the concatenation of the shape data and the independent object bitstreams, so that it is possible to transmit or to extract only a part of it (Fig.1).

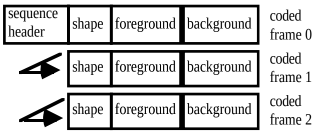

Fig.1. Object-scalable bitstream structure

3 RESULTS

Simulations were performed under the test conditions proposed by MPEG-4 [2]: a total bitrate of 48kbit/s is achieved, with a free rate distribution among the regions. Images from test sequences 'Akiyo' and 'Hall Monitor' were coded in QCIF at 5Hz, and the rate-distortion optimization was performed under a distortion constraint for the foreground and background separately in order to obtain the targeted total bitrate. This introduces coding delay but insures a stable quality along the sequence.

On Fig.2 some object-scalable decoded images are presented and compared with the full frame compression scheme results. A redundancy on the object contours is necessarily introduced to provide separate object representations while keeping the quadtree partition, as shown on the pruned quadtree pictures from Fig.4. Because of this redundancy, the non-scalable version of the codec is globally more efficient than the scalable one when considering the full frame PSNR. However, with the full frame compression scheme, there is no mean to control the rate-distortion distribution among the different regions of interest; hence, when separately calculating the PSNR over the background and the foreground regions, it appears that the latter always suffers from a higher distortion, although it is the main area of interest (Table 2, Fig.3). Since the rate-distortion optimization is done globally, less complex regions tend to be coded with a better quality.

Results could be improved by further adapting either the segmentation or the coding models to object-based data. The extra-cost is mainly due to the redundancy inherent to the quadtree-based object representation, as is shown on Fig.4. A deep quadtree going down to the accurate contours is too expensive to code; another model needs to be designed, probably at the price of a higher complexity. On the other hand, as stated in Section 2.3, coding models may be improved in order to deal more efficiently with arbitrary-shape data on the object boundary blocks.

sequence rate	AKIYO	HALL
full compression	46.9 kbit/s	48.0 kbit/s
obj.-scal. compr.	48.3 kbit/s	48.3 kbit/s
shape coding	1.3 kbit/s	0.98 kbit/s
foreground	38.6 kbit/s	18.5 kbit/s
background	8.7 kbit/s	28.7 kbit/s

Table 1. Rate distribution

sequence Y PSNR	akiyo full frame	akiyo obj. scal.	hall full frame	hall obj. scal.
full frame	37.3 dB	34.2 dB	37.5 dB	35.1 dB
fgd only	34.9 dB	33.7 dB	34.8 dB	35.2 dB
bgd only	39.7 dB	34.5 dB	37.5 dB	35.1 dB

Table 2. Distortion distribution

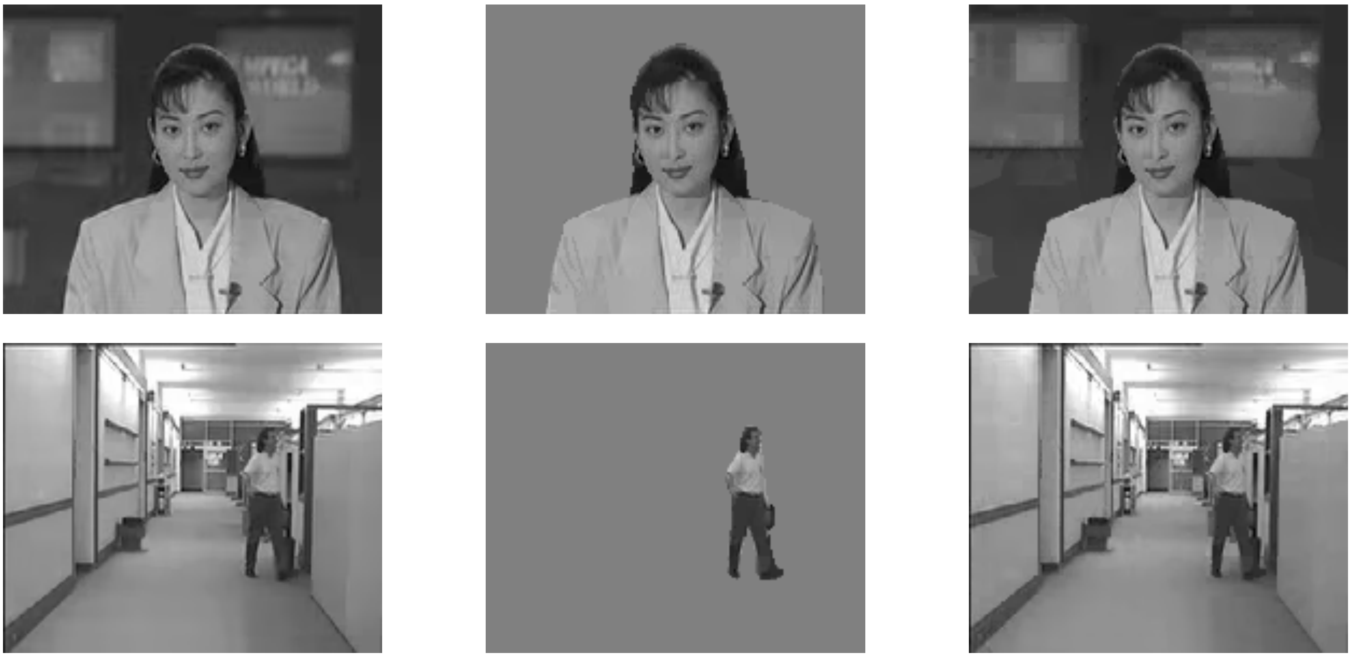

Fig.2. 'Akiyo' (#0) and 'Hall' (#294) sequences coded at 48kbit/s. Left image: full frame compression. Center image: foreground only. Right image: object-scalable compression, reconstructed frame.

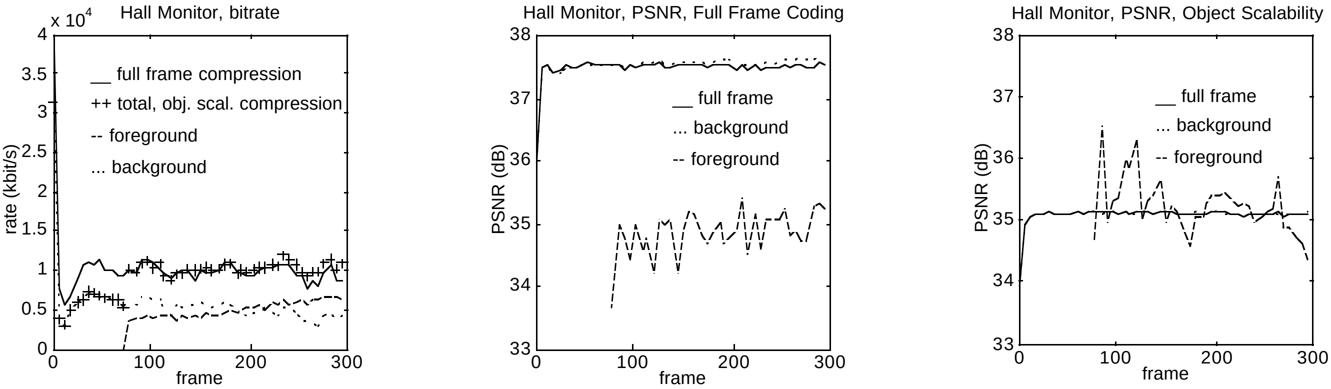

Fig.3. Rate and PSNR measures for 'Hall' sequence coded at 48kbit/s.

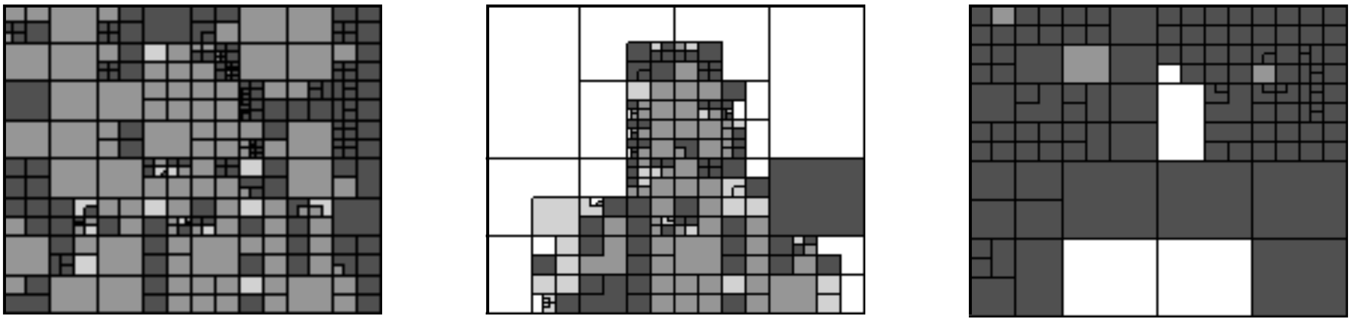

Fig.4. Pruned quadtrees corresponding to optimal solution, 'Akiyo', frame#0. Left: full frame compression; Center: foreground only; Right: background only. White blocks: not coded. Light gray blocks: bilevel clustering selected. Medium gray blocks: DCT selected. Dark gray blocks: fractal compression selected.

4 CONCLUSION

In this paper an object-scalable extension of the dynamic codec proposed by EPFL to MPEG-4 first round of tests has been presented. The dynamic coding efficient way of representing data is exploited to code only a given object within the sequences instead of the full frames. By providing an object scalable bitstream, this codec allows multimedia functionalities such as object of interest editing, archiving, retrieval, transmission and manipulation.

5 REFERENCES

[1] E.Reusens, "Joint optimization of representation model and frame segmentation for generic video compression". *Signal Processing,* Vol.46, No.1, pp. 105-117, September 1995.

[2] "MPEG-4 Testing and Evaluation Procedures Document". ISO/IEC JTC1/SC29/WG11/N999, Tokyo, July 1995.

[3] Touradj Ebrahimi et al., "Dynamic coding of visual information". ISO/IEC JTC1/SC29/WG11/M0320, Dallas, November 1995.

[4] Touradj Ebrahimi et al., "Dynamic coding of visual information: improved implementation". ISO/IEC JTC1/SC29/WG11/M0573, Munich, January 1996.

[5] Emmanuel Reusens, Touradj Ebrahimi, Roberto Castagno, Corinne Le Buhan and Murat Kunt, "Dynamic coding for visual communications". In *Proc. of VIII European Signal Processing Conference,* Trieste, Italy, September 1996.

[6] Frank Bossen and Touradj Ebrahimi, "Region shape coding". ISO/IEC JTC1/SC29/WG11/M0318, Dallas, November 1995.

[7] "MPEG-4 Video Verification Model v.2.0". ISO/IEC JTC1/SC29/WG11/N1260, Firenze, Italy, March 1996

[8] "Description of Core Experiments on object- or region oriented texture coding in MPEG-4 video". ISO/IEC JTC1/SC29/WG11/N1259, Firenze, Italy, March 1996