DYNAMIC CODING FOR VISUAL COMMUNICATIONS

Emmanuel Reusens, Touradj Ebrahimi, Roberto Castagno, Corinne Le Buhan, Murat Kunt
Signal Processing Laboratory
Swiss Federal Institute of Technology
CH-1015 Lausanne, Switzerland

ABSTRACT

In this paper, a new approach to the problem of visual data representation in the framework of multimedia is introduced. The approach, named *dynamic coding*, consists in a dynamic combination of multiple representation models and segmentation strategies. Given an application, these two degrees of freedom are assembled so as to yield a specific profile which meets the specifications dictated by the application. The data is represented as the union of data segments, each described with a locally appropriate representation model. In order to illustrate this approach, a video compression system, based on the principles of dynamic coding, is proposed in the context of video-telephone/conference applications.

1 INTRODUCTION

The current trend in communications is the convergence of the worlds of computers and consumer audio-visual products, together with the diversification of data sources. Multimedia is a platform for exchanging information coming from various sources of different nature. Accordingly, it involves a plethora of applications with constraints or conditions imposed by their specific environment, which results in a crisis in data management. The critical point is that it is acknowledged that no universal coding technique is available, that is, no coding technique on its own will be able to cover effectively the entire range of applications.

This paper presents a unified approach to visual data representation named *dynamic coding*. This denomination stands for dynamic combination of multiple compression tools. The basic principles and the major characteristics of this approach are reviewed in Sec. 2. As illustrative example of dynamic coding, Sec. 3 presents an image sequence coding algorithm in the context of video-telephone/conference applications. In Sec. 4, we show that the proposed system is equivalent to a multi-criteria data segmentation. Section 5 presents a collection of experimental results while Sec. 6 concludes this publication.

2 DYNAMIC CODING PRINCIPLES

The basic principle of dynamic coding stems from the observation that no coding technique alone is able to properly handle the complete range of applications but rather that each technique is more suited to one particular environment. The dynamic coding approach consists therefore of a combination of multiple compression techniques dynamically activated or discarded according to the environment defined by the application. Compression techniques are therefore seen as coding tools which can be combined in several ways to form a system adapted to the application under consideration.

The data is partitioned into several regions, each of them encoded using one representation model chosen from a set of compression techniques. This approach allows to locally adapt the compression model according to the region characteristics. From the above lines, it becomes clear that dynamic coding is based on two main degrees of freedom, namely (1) the data segmentation and (2) the representation model associated with every resulting segment. Fig. 1 illustrates symbolically these two degrees of freedom in a graph. The vertical axis corresponds to the data segmentation. Any segmentation algorithm can be applied ranging from fixed partition into regular blocks up to arbitrary shape segmentation through quadtree or polygonal based division. The horizontal axis corresponds to the set of coding tools used to model each region resulting from a given segmentation. These two axes span a plane which reflects all possible coding strategies available within the dynamic coding framework. Each point in the plane represents a very particular coding strategy in which a given scene is segmented into regions of given shape and size, each approximated using a particular coding technique.

Dynamic coding is therefore not a particular compression system but a *communication language* ensuring segmentation and coding tools to inter-work. In the context of dynamic coding, a scene description is made up of three components: (1) the segmentation, (2) a technique identifier associated with each region, and (3) the coding parameters describing each region within the selected representation model. These three components

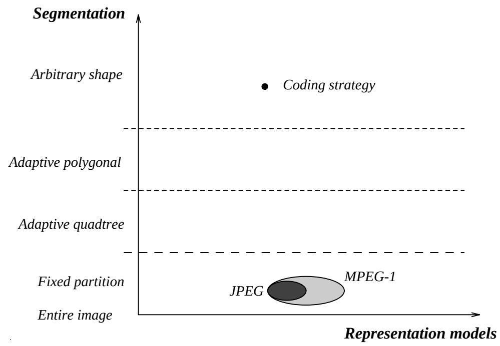

Figure 1: Symbolic description of the data representation system related to the dynamic coding approach

must respect normative requirements to ensure correct interpretation at the decoding side.

As illustrated in Fig. 2, dynamic coding follows a two-step procedure. The two degrees of freedom (segmentation and representation models) generate a space of coding strategies among which many infringe the requirements imposed by the application. The first issue consists therefore in delimiting a set of admissible solutions where the term *admissible* is defined with respect to the considered application. This task is performed by directly acting on the two degrees of freedom and discarding inappropriate representation models and segmentation strategies. This operation is signal-independent and defines a profile that fits the application specifications by filtering out inadequate coding strategies. For instance, if a limited computational complexity is allowed, representation models and segmentation strategies which require a high degree of complexity will be *a priori* discarded during this first operation.

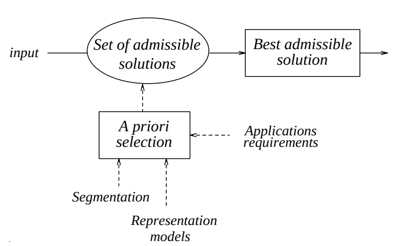

Figure 2: Dynamic coding two-step procedure

This first step of *a priori* selection defines a set of solutions in which the system is allowed to search for an appropriate coding scenario. Each coding strategy belonging to the set of admissible solutions is a very particular way to represent a set of data and results in different characteristics such as rate, distortion, error signal statistics, number of segmented regions, etc. The second major issue is therefore to point out the coding scenario which is optimal with respect to a predefined criterion. This requires to define a criterion together with a procedure for the determination of the optimal solution. This procedure is transparent to the decoder.

Dynamic coding allows to gather contradictory requirements such as openness, flexibility, complexity, efficiency, and genericity in an optimal way.

3 DYNAMIC VIDEO CODING

Dynamic coding is a global approach to visual data representation and many variations on the same theme are possible. Specific profiles differentiate themselves by their particular set of admissible solutions and the criterion with respect to which the optimal admissible solution is defined. This section aims at providing an illustrative example of the dynamic coding approach, a video compression system for video-telephone/conference applications, based on the principles presented in the previous section. The algorithm has been submitted to the MPEG-4 committee as the *EPFL/LTS* proposal [1].

As summarized in Fig. 2, the first step consists in delimiting the set of admissible solutions with respect to the requirements imposed by video-telephone/conference applications. Section. 3.1 describes the set of admissible coding strategies and justifies the choices. Once the set of solutions is delimited, the best coding strategy must be identified. The criterion by which optimality is defined and the procedure for determining the best solution are described in Sec. 3.2.

3.1 Set of Admissible Solutions

As depicted in Fig. 2, the first step consists in defining the set of admissible solutions with respect to the application in question. Video-telephone/conference applications call for low encoding/decoding delay (due to full-duplex nature of communications) and very high compression (owing to the narrow channel bandwidth supporting these applications). Accordingly, the set of coding strategies is confined to solutions which fulfill these two major requirements.

In order to achieve low coding delay, the segmentation is restricted to generate two-dimensional regions. This means that the processing and compression of the video signal is performed on a frame by frame basis. The segmentation is further limited to a quadtree partitioning. The motivation is twofold. First, quadtree segmentation constitutes a good trade-off between efficiency and representation cost. Second, the hierarchical structure of the quadtree segmentation greatly alleviates the succeeding problem of the optimal solution selection, as will be seen in the next section.

In the perspective of maintaining a reasonable encoding/decoding delay, representation models exploiting forward temporal redundancy are *a priori* discarded. Only spatial (intra) and backward temporal redundancy removal are permitted. Five representation models in agreement with these requirements and acknowledged as effective for high compression have been selected. These five representation models are (i) the Discrete Cosine Transform (DCT) followed by an adjustable quantization of the coefficients (ii) a fractal-based compression technique (iii) a graphic-oriented model (iv) a non-moving region model and (v) motion compensation followed by error coding (DCT with adjustable quantization).

The set of admissible solutions includes all possible quadtree frame segmentations together with one representation model associated with each resulting region. The cardinality of the set is very large and grows as $O(K^{4^d})$ as a function of the maximal allowed tree depth d. However, it is possible to build, at reasonable computational cost, the spine of the set of solutions thanks to its hierarchical structure. The spine is constructed as follows. Starting from the entire picture, the image is recursively split into four equally-sized subblocks. At each recursion step, the descriptions of the corresponding blocks are computed with respect to each of the representation models. The resulting rate and distortion quantities are also evaluated with respect to each representation model. The recursion stops when the tree reaches the maximal allowed tree depth. This operation generates a complete tree in which to each node are associated the parameters of the description within the available representation model,s together with the resulting rate and distortion values. This spine allows a fast evaluation of the characteristics of any admissible solution.

3.2 Optimal Admissible Solution

The set of admissible solutions includes all the different strategies that can be used to represent the current frame. As depicted in Fig. 2, the next step is to determine the optimal solution with respect to a predefined criterion. In the context of digital source coding, the traditional framework is given by Shannon's rate-distortion theory, where the overall source distortion is minimized subject to a channel rate constraint. Among admissible strategies, the goal is to determine the coding scenario minimizing the distortion provided that a given bitrate budget, R_{budget}, is not exceeded. Mathematically, the optimization criterion may be formulated as follows:

$$\min_{B \in \mathcal{S}} D(B) \ \textit{subject to} \ R(B) \leq R_{budget}, \qquad (1)$$

where $\mathcal{S}$ is the set of admissible solutions, $D(B)$ is the distortion, and $R(B)$ the bitrate resulting from encoding according to the scenario $B \in \mathcal{S}$.

The optimization technique that has been used relies on a fundamental theorem of Lagrange multipliers developed in the framework of optimal resources allocation theory [2]. The technique consists in converting the constrained minimization of (1) into an equivalent unconstrained problem by merging rate and distortion through the Lagrangian multiplier λ. The unconstrained problem comes down to selecting the coding strategy which gives the minimum of the Lagrangian cost function expressed as :

$$J(\lambda) = D(B) + \lambda R(B). \qquad (2)$$

For a given multiplier λ, the minimization of the cost function $J(\lambda)$ results in a solution $B^*(\lambda)$, an associated rate $R^*(\lambda)$ and a distortion $D^*(\lambda)$. It has been demonstrated [2, 3] that, as we sweep λ over positive values, the pairs ($R^*(\lambda)$,$D^*(\lambda)$) trace out the optimal rate-distortion curve. In particular, if for a given $\lambda = \lambda_c$, $R^*(\lambda)$ happens to coincide with the targeted rate R_{budget}, then $B^*(\lambda)$ is the solution of the constrained problem (1). Readers further interested in this theorem and its implications are referred to [2, 3].

The optimization problem is solved by iterative minimization of $J(\lambda)$ for λ values converging to the appropriate λ_c. The convergence towards the appropriate λ_c is guaranteed thanks to the monotonic nature of optimal solution rate $R(B^*)$ versus λ. Indeed, the Lagrangian multiplier λ plays the role of a quality factor that acts as a trade-off distortion for rate. The appropriate λ leading to $R(B^*) = R_{budget}$ can therefore be retrieved by intelligently searching over λ values by fast convex search such as the bisection technique.

The above-mentioned search for the appropriate λ requires the minimization of $J(\lambda)$ for arbitrary arguments λ. It can be shown that, given a complete image partition, $\{P_i, 0 \leq i < N\}$, the minimization of the cost function $J(\lambda)$ over the entire frame can be performed independently over each segment P_i. The coding strategy minimizing $J(\lambda)$ is identified by pruning the spine of the set of solutions along a bottom-up path. At each node, the option, yielding the minimum of $J(\lambda)$ over the associated block, is selected. The operation stops when the root of the tree is reached. At this point, the partition, together with its associated coding models which minimize $J(\lambda)$, are identified for a given quality factor λ. Readers interested in further details are referred to [4].

Although in the previous example the distortion is minimized given a maximum rate budget, the algorithm is able to find out the strategy that gives the minimum rate for a desired quality. We simply swap the rate and distortion functions in Eq. (1), and the dual Lagrangian cost function becomes $\bar{J}(\lambda) = R + \lambda D$. Since $\bar{J}(\lambda)$ possesses the same properties as $J(\lambda)$, the optimization procedure can be conducted without any modification by simply inverting the rate and distortion functions. The system is therefore capable of switching from a rate control to a distortion control mode.

4 MULTI-CRITERIA SEGMENTATION

The determination of the optimal admissible strategy corresponds to a joint optimization of the frame segmentation together with the representation model associated with each segment. Therefore, the optimization procedure results in a multi-criteria picture segmentation which differentiates from classical object-based compression approaches by the adaptive definition of the notion of *object*. Indeed, in classical approaches the definition of *object* is specified *a priori* and its definition conditions the entire compression procedure. As an example, second-generation coding algorithms define an object to be any set of adjacent pixels having similar grey-level intensities [5]. After the data segmentation, objects are described by polynomial approximation or some other technique appropriate for describing nearly uniform regions. As opposed to classical techniques, the proposed algorithm defines the criterion by which the objects are created from a library of criteria put in competition.

Since the definition of object is adaptive, the resulting regions created by this approach may not correspond to actual objects in the scene, from a semantic point of view. However, if object interactivity features are desired, it is possible to bring them into this scheme by special masking operations as demonstrated in [6].

5 EXPERIMENTAL RESULTS

Experiments have been carried out on two color video sequences, *Hall Monitor* and *News*. The input format is QCIF (144×176), 4:2:0, at 30 Hz. Input sequences have been temporally down-sampled at 5 Hz. Simulations were performed on 10 seconds of the video signals.

The system have been configured so as to lead to a constant quality in the decoded sequences. In other words, for each frame the rate is minimized subject to a given constant distortion. However, for each test sequence, the distortion values have been chosen in order to reach a specific overall bitrate. These distortion values leading to a specific bitrate have been determined experimentally. Bitrates of 10 and 48 kbit/sec. were targeted for the *Hall Monitor* scene whereas rates of 24 and 48 kbit/sec. were aimed at for the *News* sequence.

Experimental results are reported in Fig. 3 and 4. These plots describe the distribution of the bitstream around frames together with the peak signal-to-noise ratio (PSNR) of the luminance (Y component) . As expected, there are significant variations in the bitrate according to the local temporal and spatial scene complexity. Since experiments are performed at constant distortion, the PSNR behavior remains flat.

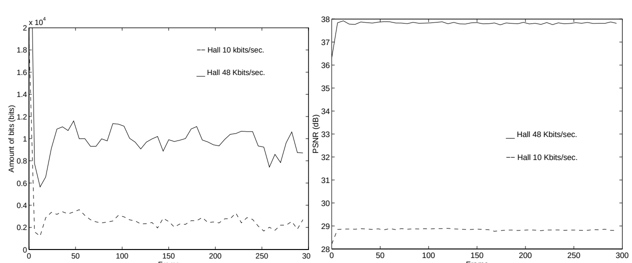

Figure 3: Rate distribution and distortion behavior for *Hall monitor*. (Left) Rate distribution at 10 and 48 kbit/sec. (Right) Distortion behavior at 10 and 48 kbit/sec.

The most noticeable degradation, in reconstructed images, is the loss of details in intricate regions such as

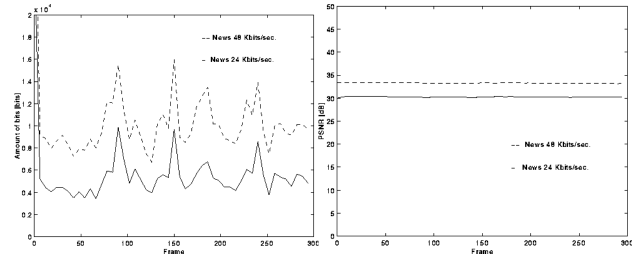

Figure 4: Rate distribution and distortion behavior for *News*. (Left) Rate distribution at 24 and 48 kbit/sec. (Right) Distortion behavior at 24 and 48 kbit/sec.

on face areas. However, edges remain sharp and motion rendition is natural. As at high compression, blocking artifacts may become visible, and a post-processing algorithm in charge of reducing this type of degradation has been applied [7].

6 CONCLUSIONS

This paper has introduced a novel approach, named dynamic coding, to the problem of visual data compaction in the context of multimedia. As illustrative example of the approach, a video compression system based on the principles of dynamic coding has also been presented.

References

[1] T. Ebrahimi *et al.* “Dynamic Coding of Visual Information”. MPEG4 proposal JTC/SC29/ WG11/M0320, Signal Processing Laboratory, Swiss Federal Institute of Technology, October 1995.

[2] H. Everett. “Generalized Lagrange Multiplier Method for Solving Problems of Optimum Allocation of Resources”. *Operation Research*, Vol. 11, pp. 399–417, 1963.

[3] Y. Shoham and A. Gersho. “Efficient Bit Allocation for an Arbitrary Set of Quantizers”. *IEEE transactions on Acoustics, Speech, and Signal Processing*, Vol. 36, No. 9, pp. 1445–1453, September 1988.

[4] 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.

[5] M. Kunt, A. Ikonomopoulos, and M. Kocher. “Second Generation Image Coding Techniques”. *Proceedings of the IEEE*, Vol. 73, No. 4, pp. 549–575, April 1985.

[6] C. Le Buhan, E. Reusens, and T. Ebrahimi. “Object-Scalable Dynamic Coding of Visual Information”. In *Proc. of VIII European Signal Processing Conference*, Trieste, Italy, September 1996.

[7] R. Castagno and G. Ramponi. “ A Rational Filter for the Removal of Blocking Artifacts in Image Sequences Coded at Low Bitrate”. In *Proc. of VIII European Signal Processing Conference*, Trieste, Italy, September 1996.