INTER MODE VERTEX-BASED OPTIMAL SHAPE CODING

Gerry Melnikov, Guido M. Schuster and Aggelos K. Katsaggelos*

Northwestern University
Electrical and Computer Engineering Dept
Evanston, Illinois 60208, USA
Email: {gerrym,aggk}@ece.nwu.edu

*3COM
Advanced Technologies Research Center
Mount Prospect, Illinois 60056, USA
Email: Guido_Schuster@3com.com

Abstract

This paper investigates the problem of optimal lossy encoding of object contours in the Inter mode. Contours are approximated by connected second-order spline segments, each defined by three consecutive control points. Taking into account correlations in the temporal direction, control points are chosen optimally in the rate-distortion (RD) sense. Applying motion to contours in the reference frame followed by the temporal context extraction, we predict the next control point location, given the previously encoded one. Based on the chosen differential encoding scheme and an additive MPEG4-based distortion metric, the problem is formulated as Lagrangian minimization. We utilize an iterative procedure to jointly find the optimal solution and the associated DPCM parameter probability mass functions.

1. INTRODUCTION

One of the important problems in object-oriented video coding is the operationally optimal allocation of available bits among texture, motion, and shape components, as well as, within each component separately.

The tasks of boundary estimation and encoding are separated in MPEG-4, although there have been some efforts to couple the two [4]. In the process of evaluating competing techniques for the standard, several binary coders were considered. Up until now, however, optimality was lacking from their Inter, as well as, Intra, mode of operation. In the context-based (CAE) framework [1], temporal redundancy is capitalized upon by doing motion compensation and extending the context template into the neighboring pixels of the reference frame. This operation is then followed by (ad-hoc) Intra mode techniques to achieve rate control. Similarly, the MMR Inter mode algorithm [13] differs from its ad-hoc Intra mode counterpart in the choice of pixels serving as context. In the Baseline approach (Inter mode) [5] a contour in the current frame is approximated through global and local motion by a contour in the previous frame, with areas exceeding a certain error threshold coded in the Intra mode. In the case of the vertex-based polynomial coders [2, 8], temporal redundancy is exploited by applying global motion compensation and Intra-coding error segments whose error exceeds some predetermined threshold. Clearly none of these approaches are operationally optimal since they fail to take the tradeoff between the rate and the distortion into account, and they do not use the distortion metric used for their evaluation in the encoding process.

We have previously proposed optimal approximations of a given boundary based on curves of different orders and for various distortion metrics [3]. Recently, operationally optimal vertex-based coders were proposed for the Intra mode [12, 6]. In [7] this problem was solved optimally and jointly with the variable-length code selection. In this work we extend this operational rate-distortion (ORD) optimal framework to take into account the temporal redundancies present in typical video sequences and develop an Inter mode ORD optimal coder.

In addition to arriving at the ORD optimal representation for a particular coding framework, characterized by fixed VLC tables, the second objective of this paper is to find the set of parameter VLC tables resulting in the most efficient ORD curve.

This paper is organized as follows. The algorithm structure is presented in Sec. 2. The additive distortion metric is discussed in Sec. 2.1. Temporal context extraction and the control point encoding issues are described in Sec. 2.2. Section 2.3 describes how the problem can be formulated as a shortest path problem and discusses VLC optimization issues. Finally, results are presented and discussed in Section 3.

2. PROPOSED ALGORITHM

In this paper we solve the problem of contour approximation optimally in the ORD sense. Contours are approximated by connected 2^{nd}-order B-spline segments, each defined by 3 consecutive control points, (p_{u-1}, p_u, p_{u+1}). Thus an ordered set of control points constitutes a code for a shape approximation. A 2^{nd}-order spline is a parametric curve that starts at the midpoint between p_{u-1} and p_u and ends at the midpoint between p_u and p_{u+1} as the parameter t sweeps from 0 to 1. These midpoints are also called knots. A precise mathematical definition of this curve is given in [6]. A sequence of 2^{nd}-order B-splines solves the interpolation problem at the knots, while being differentiable everywhere, including the knots. This smoothness property, coupled with the simplicity of definition, makes B-splines a natural choice for the shape coding applications.

Although an ordered set of control points defining approximating splines may contain elements from anywhere in the image, it is unlikely that locations far from the original boundary would lead to an ORD optimal approximation. This leads naturally to the concept of the admissible control point band [11], thus excluding all pixels located farther than the band width away from the original boundary from consideration.

2.1. Distortion

Distortion between an original boundary and its approximation can be quantified based on either a (non-additive) maximum operator or an (additive) summation operator. In MPEG-4 the following additive distortion metric has been used per frame to evaluate performance of competing algorithms:

$$D_{MPEG4} = \frac{\text{number of pixels in error}}{\text{number of interior pixels}}, \quad (1)$$

where a pixel is said to be in error if it belongs to the interior of the original object and the exterior of the approximating object, or vise-versa.

Spline segment distortions need to be defined in order to evaluate the total (additive) boundary distortion. This is done by first associating segments of the approximating curve with segments of the original boundary, as shown in Fig. (1). Here the midpoints of the line segments (p_{u-1}, p_u) and (p_{u+1}, p_u), l and m, respectively, are associated with the points of the boundary closest to them, l' and m'. That is the segment of the original boundary (l', m') is approximated by the spline segment (l, m).

Let us now define by $d(p_{u-1}, p_u, p_{u+1})$ the spline segment distortion, as shown in Fig. (1). Specifically, to calculate

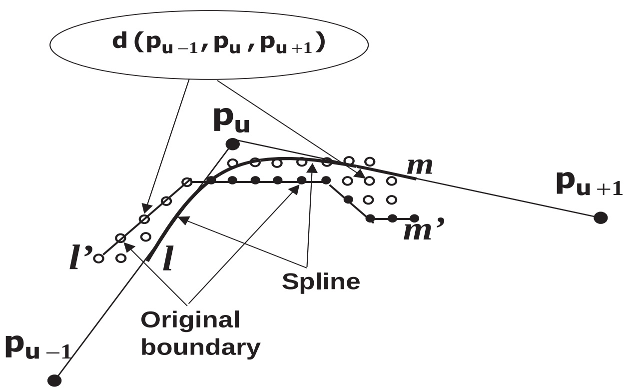

Figure 1: Area between the original boundary segment and its spline approximation (circles).

$d(p_{u-1}, p_u, p_{u+1})$ we count the pixels in error after quantizing the continuous spline to fit the pixel grid of the image. These pixels are shown as hollow circles in the figure. Special care must be taken when solving the correspondence problem to ensure that the starting boundary pixel of the next segment coincides with the last boundary pixel of the current segment and that error pixels on the border line between m and m' are not counted twice when computing the next segment distortion. Based on the segment distortions, the total boundary distortion is therefore defined by

$$D(p_0, \ldots, p_{N_P-1}) = \sum_{u=0}^{N_p} d(p_{u-1}, p_u, p_{u+1}), \quad (2)$$

where N_p is the number of control points and $p_{-1} = p_{N_p+1} = p_{N_p} = p_0$. The last equality ensures that an approximation to a closed contour is also closed and simplifies implementation.

2.2. Temporal Correlation and Rate

Object boundaries and control points representing them exhibit correlation in the temporal direction. To capitalize on this we employ global motion compensation to align a previously encoded object with the current object. The global motion vector is chosen to minimize the number of mismatched pixels. We employ the $(angle,\ run)$ framework, also used in the INTRA mode [12, 7], to encode consecutive control point locations.

Contexts are generated based on the motion compensated object for each pixel in the admissible control point band. These contexts serve a dual purpose by adaptively selecting a probability model describing the location of the next control point and by providing a reference $(angle_{ref},\ run_{ref})$ used by the DPCM scheme to encode that location. Figure 2 depicts a hypothetical context window in the reference frame after motion compensation. Centered on a pixel in the admissible control point band, it is used to extract both the most likely direction and the most likely length of the vector pointing from that pixel to the next potential control point. That is, if an actual control point is located at the current position, this context estimates where the next control point is most likely to be. The directional context is obtained by

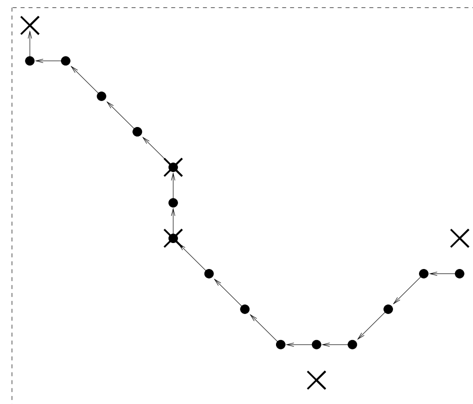

Figure 2: Control points (X marks) and boundary pixels (circles) in a temporal context window. Context: NW direction, run of 4.

selecting the direction in which the maximum number of boundary pixel transitions occur, based on the chosen pixel ordering scheme. This corresponds to the North-West (NW) direction in Fig. (2). Similarly, the run length context is obtained by selecting the most frequently occurring distance between consecutive control points in that window. In this example, a run length of 4, occurring 3 times, is selected.

Once the direction and the run length contexts have been determined, a context-dependent variable-length code is used for coding the location of the next control point of the object in the current frame. Clearly, it is desirable to assign shorter codewords to directions and runs closest to the context. Figure (3A) shows a typical direction probability distribution, given the NW context. Here vector lengths are proportional to the probability of the corresponding direction. Sometimes the context information may not be available due to inhomogeneous motion, occlusion, or lack of the reference frame. In that case, we revert to the INTRA mode and the direction is encoded with respect to the previously encoded direction in the same object. This concept is illustrated in Fig. (3B), where the dashed line represents the previously encoded direction, and the solid-line vectors represent the 4 equally probable moves. Similarly, a hypothetical but likely conditional probability distribution for the run length is shown in Fig. (4A) for the case the context is 4 and in Fig. (4B) for the case the context is unavailable. The issue of selecting efficient variable-length codes for the direction and the run will be discussed in Sec. 2.3. Regardless of the context, the first control point location is en-

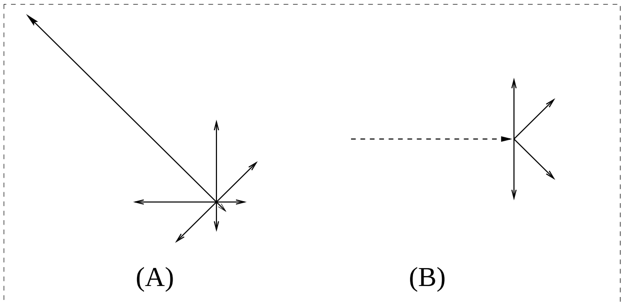

Figure 3: Probability assignments for the direction (proportional to vector lengths), (A) context is NW, (B) no context.

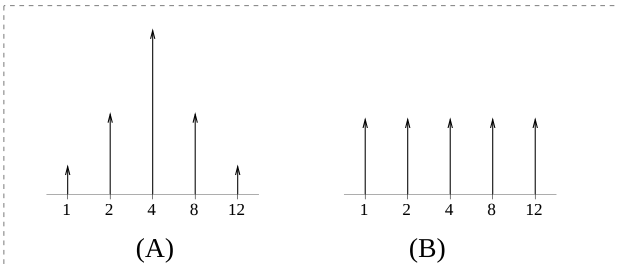

Figure 4: Probability assignments for the run length (proportional to vector lengths), (A) context is 4, (B) no context.

coded absolutely and that cost together with the cost of sending a global motion vector, constitutes an overhead outside the realm of the ORD optimization described in Sec. 2.3.

If $r(p_{u-1}, p_u, p_{u+1})$ denotes the segment rate for representing p_{u+1} given control points p_{u-1}, p_u, then the total rate is given by

$$R(p_0, \ldots, p_{N_P-1}) = \sum_{u=0}^{N_p-1} r(p_{u-1}, p_u, p_{u+1}). \tag{3}$$

The idea of using contexts for INTER mode shape coding is not new. However, unlike the CAE [1] and the MMR [13] coders, which use contexts to predict whether a given pixel belongs to the boundary, we avoid pixel-level noise effects by applying contexts to more noise-resilient features, such as, the localized boundary orientation and smoothness.

2.3. Determining the Optimal Solution

Having defined the total distortion and rate in the previous section, we are solving the following optimization problem:

$$\min_{p_0,\ldots,p_{N_P-1}} D(p_0, \ldots, p_{N_P-1}), \quad subject\ to: \\ R(p_0, \ldots, p_{N_P-1}) \leq R_{max}, \tag{4}$$

where both the location of the control points p_i and their overall number N_P have to be determined. It should be understood, however, that the solution to this problem is optimal only within the chosen code structure. Thus, a different motion compensation scheme, a different set of VLC tables, a wider control point band - all may result in a more efficient ORD performance of the algorithm.

We convert the above constrained minimization problem into an unconstrained one by forming the Lagrangian

$$J_\lambda(p_0, \ldots, p_{N_P-1}) = \\ D(p_0, \ldots, p_{N_P-1}) + \lambda \cdot R(p_0, \ldots, p_{N_P-1}), \tag{5}$$

where for any choice of the multiplier λ, J_λ is the cost function to be minimized. The above cost function is expressed as a sum of incremental spline segment costs defined as,

$$w(p_{u-1}, p_u, p_{u+1}) = \\ d(p_{u-1}, p_u, p_{u+1}) + \lambda \cdot r(p_{u-1}, p_u, p_{u+1}). \tag{6}$$

The optimal set of control points $(p_0^*, \ldots, p_{N_P-1}^*)$ is then found by casting the problem as a shortest path in a Directed Acyclic Graph (DAG) with control points playing the role of vertices and incremental costs $w()$ serving as edge weights [3]. Dynamic Programming (DP) is employed to find the shortest path in the DAG for a fixed rate-distortion tradeoff λ. We employ a Bezier curve search [10] in order to arrive at λ^*, the multiplier resulting in the total rate closest to the target rate of R_{max}, in very few iterations.

The optimal solution to the shape coding problem in the INTRA mode was shown in [7] to be highly sensitive to the VLC table used. An iterative procedure for finding a locally optimal parameter distribution model was proposed. In this work we extend this idea to the INTER mode by removing the conditioning of the operationally optimal solution on an ad-hoc VLC. That is, the conditional optimization of

$$\{p_0^*, \ldots, p_{N_P-1}^*\} = \\ arg[\min_{p_0,\ldots,p_{N_P-1}} [D^*(\cdot) + \lambda \cdot R^*(\cdot)] | VLC_{fixed}] \tag{7}$$

is converted using an iterative procedure [9, 7], depicted in Fig. 5, into the following problem:

$$\{p_0^*, \ldots, p_{N_P-1}^*\} = \\ arg \min_{p_0,\ldots,p_{N_P-1}; f \in F} [D^*(\cdot) + \lambda \cdot R^*(\cdot)], \tag{8}$$

where f is the parameter probability mass function conditioned on the context. Hence the shape approximation and the parameter probability model are found jointly and ORD optimally.

The iteration process begins with the proposed INTER mode encoder compressing the input binary images based on a given rate-distortion tradeoff λ and some initial probability models for (*direction — context*) and (*run — context*) parameters. Conditional symbol probabilities rather than codeword lengths are used by the encoder in this iterative process. Hence, symbol entropies $-p \log p$ take the place of the bit-rate $r()$ and $R()$ in Eqs. (5) and (6).

Having encoded the input sequence at iteration k, based on the probability mass function $f^k()$, we use the frequency of the output symbols to compute $f^{k+1}()$, and so on. It is straightforward to show that the total cost $D^k() + \lambda \cdot R^k()$ in Eq. (5) is a non-increasing function of the iteration k. This procedure is guaranteed to converge and the iterations stop when the cost improvement is less than ϵ. Finally, the symbols are arithmetically encoded and sent to the decoded together with their associated probabilities, which are fixed for the entire video sequence.

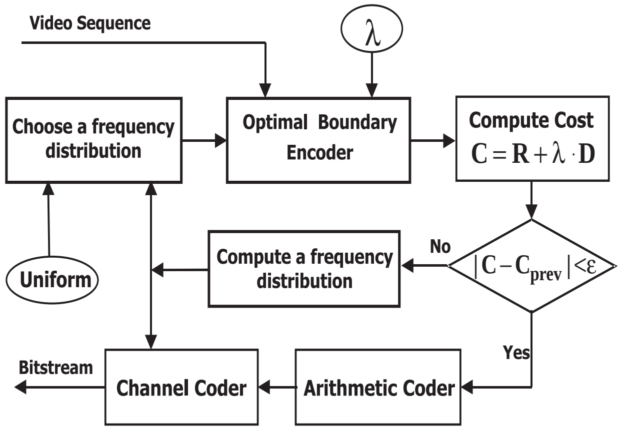

Figure 5: The entropy encoder structure.

3. RESULTS AND CONCLUSIONS

Figure 6 shows the ORD curves of the proposed (INTER mode) algorithm for the SIF sequence "kids". The distortion axis d_n represents the average of the D_{MPEG4}'s defined in Eq. (1) for one frame, over 100 frames. As the figure demonstrates, our result compares favorably with the both the baseline [5] and the vertex-based [8] algorithms in the INTER mode across most of the range of bitrates. Also, as expected, it outperforms the ORD optimal INTRA mode encoding without VLC optimization (shown with $\circ$ symbols). The proposed algorithm produces relatively better results in the low bitrate ($R < 1000$ bits per frame) region. In the very low distortion region ($d_n \leq 0.015$) of operation, however, the proposed algorithm requires more bits than both the baseline and the vertex-based methods. This is due to the fact that for near-lossless boundary encoding the chosen code structure (direction plus run) is inefficient. We trained our VLC model iteratively on the first 10 frames of the test sequence and used that model to encode the rest of the frames. However, in principle, VLC optimization could have been performed iteratively on all 100 frames of the sequence.

In this implementation, 8 directions (separated by 45^o) were allowed in the case the context is present. They are encoded differentially with respect to the context and correspond to the 8 out of 12 conditional symbols for the direction component. The other 4 symbols are used when a context is not present (see Fig. 3B). The run component was represented by 20 symbols, with only 5 symbols (corresponding to runs of 1, 2, 4, 8, and 12) used for any given context. Bit-rate reported in Fig. 6 for the proposed method takes into account bits for the global motion vectors, searched in a 32×32 window.

Despite its ORD optimality, it is important to recognize that the results of the proposed algorithm do not differ significantly from the best INTRA mode results [7]. Further efficiency can be gained with the use of more sophisticated motion model and context utilization. For example, the affine motion model and/or segment motion can lead to more useful contexts.

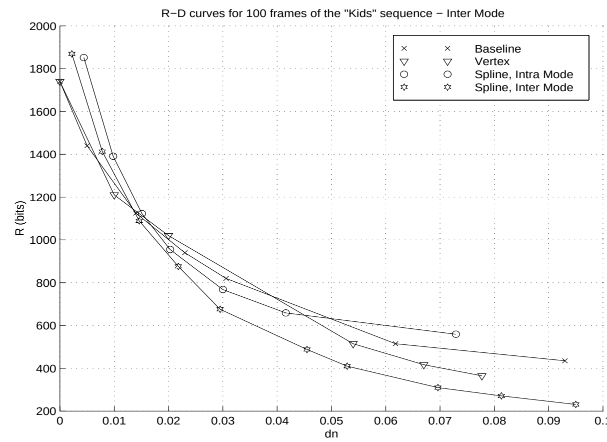

Figure 6: Rate-Distortion curves.

4. REFERENCES

[1] N. Brady, F. Bossen, and N. Murphy, "Context-based arithmetic encoding of 2D shape sequences", *Proc. ICIP97*, pp. I-29-32, 1997.

[2] M. Hötter, "Object-oriented analysis-synthesis coding based on moving two-dimensional objects", *Signal Processing: Image Communications*, vol. 2, pp. 409-428, Dec. 1990.

[3] A. K. Katsaggelos, L. Kondi, F. W. Meier, J. Ostermann, G.M. Schuster, "MPEG-4 and Rate Distortion Based Shape Coding Techniques", *Proc. IEEE*, pp. 1126-1154, June 1998.

[4] L. Kondi, F. W. Meier, G. M. Schuster, A. K. Katsaggelos, "Joint optimal object shape estimation and encoding", *Proc. of SPIE*, vol. 3309, pp. 14-25, Jan. 1998.

[5] S. Lee, *et al*. "Binary shape coding using 1-D distance values from baseline", *Proc. ICIP97*, pp. I-508-511, 1997.

[6] G. Melnikov, P. V. Karunaratne, G. M. Schuster, A. K. Katsaggelos "Rate-Distortion optimal boundary encoding using an area distortion measure", *Proc. ISCAS98*, Jun. 1998.

[7] G. Melnikov, G. M. Schuster, A. K. Katsaggelos "Simultaneous Optimal Boundary Encoding and Variable-Length Code Selection", *Proc. ICIP98*, Oct. 1998.

[8] K. J. O'Connell "Object-adaptive vertex-based shape coding method", *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 7, pp. 251-255, Feb. 1997.

[9] D. Saupe "Optimal piecewise linear image coding", *Proc. SPIE Conf. on Visual Comm. and Image Proc.*, vol. 3309, pp. 747-760, 1997.

[10] G. M. Schuster and A. K. Katsaggelos *Rate-Distortion Based Video Compression, Optimal Video frame compression and Object boundary encoding*. Kluwer Academic Press, 1997.

[11] G. M. Schuster and A. K. Katsaggelos, "An Optimal Polygonal Boundary Encoding Scheme in the Rate-Distortion Sense," *IEEE Trans. Image Processing*, vol. 7, no. 1, pp. 13-26, Jan. 1998.

[12] G. M. Schuster, G. Melnikov, and A. K. Katsaggelos, "Optimal Shape Coding Techniques," *Signal Processing Magazine*, Nov. 1998.

[13] N. Yamaguchi, T. Ida, and T. Watanabe, "A binary shape coding method using modified MMR", *Proc. ICIP97*, pp. I-504-508, 1997.