

# RATE-DISTORTION OPTIMAL BOUNDARY ENCODING USING AN AREA DISTORTION MEASURE

*Gerry Melnikov, Passant V. Karunaratne, Guido M. Schuster\* and Aggelos K. Katsaggelos*

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

\*3COM  
Advanced Technologies Research Center  
Mount Prospect, Illinois 60056, USA  
Email: gmschust@ece.nwu.edu

## ABSTRACT

In this paper an optimal boundary encoding algorithm in the rate-distortion sense is proposed. Second-order B-spline curves are used to model object boundaries. An additive area distortion measure between the original boundary and its approximation is employed in the optimization process. The problem is formulated in a Directed Acyclic Graph (DAG) paradigm, and the shortest path solution is used to optimally select control point locations of the B-spline curve approximation based on the desired rate-distortion tradeoff.

## 1. INTRODUCTION

Object oriented video compression has assumed an important role in recent years, partly because aspects of it are considered by the ongoing MPEG-4 standardization effort. Applications such as content-based storage and retrieval of video information, mobile communications and studio film authoring are the driving force behind the research in this area. Within the object-oriented framework, analysis and synthesis of an image sequence is done by treating it as a collection of disjoint video object planes (VOP). Evolution of each VOP in time is described by sending texture, motion, and segmentation information to the decoder. The fixed structure of the bitstream makes it necessary not only to develop more efficient motion estimation, segmentation, and boundary encoding algorithms, but, also, to solve the problem of the optimal allocation of resources among these video object descriptors. Consequently, efficiency in the representation of video objects is paramount for applications requiring a high accuracy in the description of video scenes or very high compression ratios.

Two main classes of binary shape coders have emerged in the MPEG-4 standardization effort: bitmap-based and contour-based coders [9]. The former includes context-based (CAE) [1] and Modified Modified Read (MMR) [8] coders. The latter includes baseline [4] and vertex-based polynomial coders [2, 6]. Although some efforts have been made to couple the process of shape coding with that of VOP segmentation [3], these processes, as well as motion

estimation, are still treated independently by the developing standard.

Most shape encoding methods reported in the literature [9] lack optimality. For the objective evaluation of competing shape coding algorithms, MPEG-4 has chosen a distortion measure defined by the number of incorrectly assigned pixels normalized by the number of interior pixels in the original boundary. None of the algorithms adopted by MPEG-4 are optimized with respect to this area-like distortion measure.

In this paper the problem of optimal boundary approximation is solved using second order B-spline segments. Unlike the approach used in [5], where optimization was carried out with respect to a global maximum distortion between the original and an approximating boundary, we utilize an additive area distortion measure. That is, we minimize the area between the original pixel boundary and its continuous approximation. Our chosen distortion metric is in line with the one employing the number of pixels in error, since for fine grid resolutions there is little difference between the two metrics.

The paper is organized as follows. The optimization problem is formulated in section 2, and the adopted distortion metric is discussed in section 3. Section 4 details the computation of the bit rate, and section 5 describes how the optimal solution is obtained using the Directed Acyclic Graph (DAG) paradigm. Results of the proposed algorithm are discussed in section 6. Conclusions and future work are presented in section 7.

## 2. PROBLEM FORMULATION

The boundary is approximated by connected spline segments. Each segment is represented by a  $2^{nd}$ -order parametric function, parameterized by  $t$ . A spline segment is uniquely specified by 3 ordered control points  $(p_{u-1}, p_u, p_{u+1})$ . As the parameter  $t$  is varied from 0 to 1, a segment is traced from the midpoint between  $p_{u-1}$  and  $p_u$  to the midpoint between  $p_u$  and  $p_{u+1}$ . From the family of parametric curves satisfying this interpolation constraint, we select the one which is continuously differentiable, including the endpoints.

This  $2^{nd}$ -order spline is defined as follows:

$$Q_u(p_{u-1}, p_u, p_{u+1}, t) = \begin{bmatrix} t^2 & t & 1 \end{bmatrix} \cdot \begin{bmatrix} 0.5 & -1.0 & 0.5 \\ -1.0 & 1.0 & 0.0 \\ 0.5 & 0.5 & 0.0 \end{bmatrix} \cdot \begin{bmatrix} p_{u-1,x} & p_{u-1,y} \\ p_{u,x} & p_{u,y} \\ p_{u+1,x} & p_{u+1,y} \end{bmatrix}, \quad (1)$$

where  $p_{u,x}$  and  $p_{u,y}$  are the Cartesian coordinates of the point  $p_u$ . Clearly, specifying a set of control points  $p_0, \dots, p_{N_P-1}$ , where  $N_P$  is the set cardinality, is equivalent to specifying a (lossy) approximation of the given boundary.

It is desired to solve the problem of control point placement optimally (in the rate-distortion sense) and fast. That is, for any chosen bit-rate  $R$ , and its associated additive area-based distortion measure  $D$ , no other selection of control points would result in a rate lower than  $R$ , while having a distortion lower or equal to  $D$ . Mathematically, the following optimization problem is to be solved:

$$\begin{aligned} \min_{p_0, \dots, p_{N_P-1}} & D(p_0, \dots, p_{N_P-1}), \\ \text{subject to: } & R(p_0, \dots, p_{N_P-1}) \leq R_{max}, \end{aligned} \quad (2)$$

where both the location of the control points  $p_i$  and their overall number  $N_P$  has to be determined. Clearly, our claim of optimality is valid only within the chosen code structure ( $2^{nd}$ -order B-splines whose control points are encoded with the given variable-length code). Optimality is also contingent upon the width of the admissible control point band [3, 5] and the ordering scheme for the boundary points [7]. However, the experiments showed no significant gain in performance when the control point width was increased beyond 2 pixels, while almost doubling the running time. It should be noted, also, that the solution depends on the first point, which is fixed and is assumed to be encoded elsewhere in the VOP bitstream.

### 3. DISTORTION

In our implementation, instead of quantizing continuous splines to fit the support grid of the original boundary in order to count the number of incorrectly assigned pixels, we perform numerical integration to estimate the area of the distortion region. The higher the resolution of the original boundary grid, the closer the numerical integral becomes to the measure involving the count of mislabeled pixels. Furthermore, employing a continuous area metric makes it possible for the decoder to reconstruct the boundary on a finer grid while still guaranteeing a given fidelity of approximation.

Let us define by  $d(p_{u-1}, p_u, p_{u+1})$  the segment distortion, as shown in Fig. 1. In determining the segment distortion, the boundary points closest to the midpoints of the line segments  $(p_{u-1}, p_u)$  and  $(p_u, p_{u+1})$ , which are called knots, are used. Based on the segment distortions, the total boundary distortion used in Eq. (2) is therefore defined by

$$D(p_0, \dots, p_{N_P-1}) = \sum_{u=0}^{N_P} d(p_{u-1}, p_u, p_{u+1}), \quad (3)$$

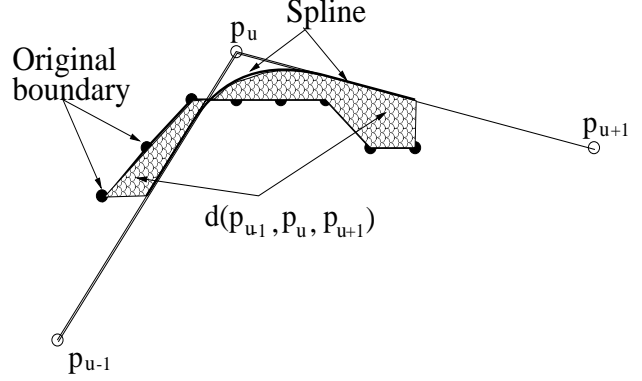


Figure 1: Area between the original boundary segment and its spline approximation

where  $p_{-1} = p_{N_P+1} = p_{N_P} = p_0$ .

To make a comparison between various algorithms the average number of pixels in error is used, defined by

$$D_n = \frac{\text{number of pixels in error}}{\text{number of interior pixels}} \quad (4)$$

for every frame. This metric is also used by MPEG-4 to evaluate performance of several boundary encoding algorithms [9]. Here, a pixel is counted as mislabeled if it belongs to the interior of the one boundary but not the interior of the other.

### 4. RATE

We de-correlate consecutive control point locations in order to encode them by employing a second-order prediction model. As in [5], each control point is described by the relative angle  $\alpha$  it makes with respect to the line connecting two previously encoded control points, and by the length  $\beta$  in pixels as shown in Fig. (2A).

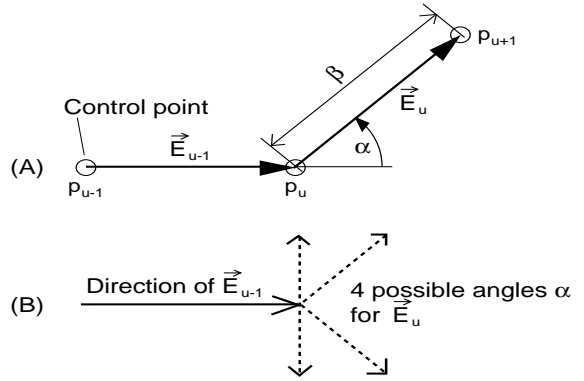


Figure 2: Encoding of a spline control point

The relative angle is allowed to take on values from the set  $\{-90^\circ, -45^\circ, 45^\circ, 90^\circ\}$ , thus requiring only 2 bits

(Fig. 2B). We exclude a  $0^\circ$  angle because a spline with that orientation can always be realized by placing a preceding control point appropriately. Special care must be taken when encoding the first and second control points because differential encoding is not applicable and to ensure that the resulting approximation is a closed boundary.

Let  $r(p_{u-1}, p_u, p_{u+1})$  denote the segment rate for representing  $p_{u+1}$  given control points  $p_{u-1}, p_u$ . Then the total rate in Eq. (2) is given by

$$R(p_0, \dots, p_{N_P-1}) = \sum_{u=0}^{N_P-1} r(p_{u-1}, p_u, p_{u+1}). \quad (5)$$

In order to reduce the algorithmic complexity and to make code words shorter, we limit the length  $\beta$  to 15 pixels. To satisfy the conflicting requirements of being able to use long length spline segments when possible, on one hand, and to do so with the fewest number of bits, we use a code table where only half of the length values have a finite code. Only lengths in the set  $(1, 2, 4, 6, 7)$  were allowed in our implementation, encoded with  $(2, 2, 2, 3, 3)$  bits respectively.

Conceptually, control points are not restricted in terms of their location in the image. However, we want them to be close to the original boundary for several reasons. First of all, while we minimize the global additive area measure, we want to be sensitive to local maximum distortions for subjective reasons. Coding efficiency and algorithmic complexity are the other considerations. For this reason we define a 2-pixel wide band, centered around the original boundary, to which all control points must belong to. Pixels in this band are labeled by the label of the closest boundary pixel [3, 5]. Boundary pixels are themselves ordered. This is done to force our solution forward along the boundary.

## 5. DIRECTED ACYCLIC GRAPH SOLUTION

The constrained minimization problem, stated in Eq. 2, is converted into an unconstrained one by using a Lagrangian multiplier method,

$$J_\lambda(p_0, \dots, p_{N_P-1}) = D(p_0, \dots, p_{N_P-1}) + \lambda \cdot R(p_0, \dots, p_{N_P-1}), \quad (6)$$

where for any choice of the multiplier  $\lambda$ ,  $J_\lambda$  is the cost function to be minimized. If a solution to the constrained problem exists, there must also exist a  $\lambda^*$  such that

$$\{p_0^*, \dots, p_{N_P-1}^*\} = \arg \min_{p_0, \dots, p_{N_P-1}} J_{\lambda^*}(p_0, \dots, p_{N_P-1}) \quad (7)$$

results in  $R(p_0^*, \dots, p_{N_P-1}^*) = R_{max}$ . As a result,  $(p_0^*, \dots, p_{N_P-1}^*)$  constitutes the optimal solution. When a certain specified bit budget is to be met, a Bezier curve search [7] is employed in order to arrive at  $\lambda^*$  in very few iterations. For the purposes of this paper, however, we wish to compare our solution to solutions reported in the literature [9] over a range of bit rates, and, therefore, we let  $\lambda$  vary in the range  $(0.1, \dots, 10.0)$ .

Let us define an incremental cost of encoding one spline segment as,

$$w(p_{u-1}, p_u, p_{u+1}) = d(p_{u-1}, p_u, p_{u+1}) + \lambda \cdot r(p_{u-1}, p_u, p_{u+1}). \quad (8)$$

Then the overall Lagrangian cost function can be written as,

$$\begin{aligned} J_\lambda(p_0, \dots, p_{N_P-1}) &= \\ & \sum_{u=0}^{N_P-1} w(p_{u-1}, p_u, p_{u+1}) + w(p_{N_P-1}, p_{N_P}, p_{N_P+1}) \\ &= J_\lambda(p_0, \dots, p_{N_P-2}) + w(p_{N_P-1}, p_{N_P}, p_{N_P+1}). \end{aligned} \quad (9)$$

This structure of the problem (global optimality implies up-to-a-level optimality) makes dynamic programming a natural choice of the minimization method. Specifically, this problem is formulated as the shortest path problem in a DAG. We map each consecutive pair of control points into a vertex and incremental costs  $w()$  into the corresponding weights [5]. This resulting DAG is searched to find the optimal set of control points [5, 7].

In addition to making our solution optimal for one object, our result is optimal over all objects in all frames. To achieve this, we use a fixed value of  $\lambda$ , the tradeoff between rate and distortion, throughout the test sequence, thus keeping the slope of the operational rate distortion (ORD) curve for every object fixed [7]. The slope condition, together with the convexity property of ORD curves, ensure that decreasing the rate for one object in the sequence will result in an increase in distortion for the object. If the saved bits were to be used for the encoding of another object, the gain in quality would not compensate the loss of quality in the first object.

## 6. RESULTS

Figure 3 shows the operational rate-distortion (ORD) curve of the proposed algorithm for the SIF sequence “kids”. The distortion axis represents the average of the  $D_n$ ’s defined in Eq.( 4) for one frame, over 100 frames. Our results compare favorably with the results of both contour-based and pixel-based algorithms reported in [9]. In the low bit rate region of operation ( $R < 800$  bits per frame) we obtain a 15% – 20% reduction in the rate for the same distortion as compared to the four approaches adopted by MPEG-4 [9]. In the low distortion region of operation, however, the proposed algorithm requires more bits, primarily because of the code structure (angle plus length) and suboptimality of our code table under these operating conditions. For comparison purposes, Figures 4 and 5 show the original bit map and its approximation of frame 5. The optimal way to encode the smallest object shown in the original frame (space between the legs of the kid on the left), in this case, is not to encode it at all, as the algorithm has chosen.

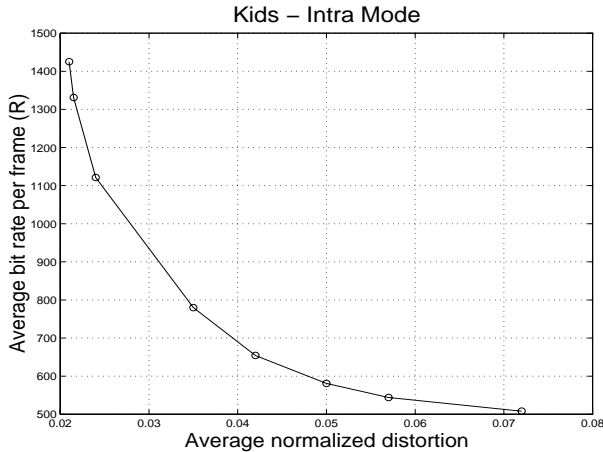


Figure 3: Rate-Distortion curve.

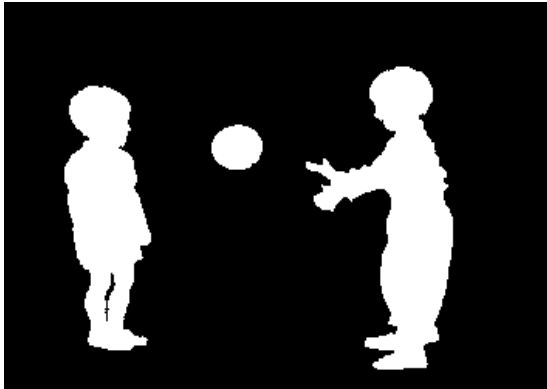


Figure 4: Frame 5 of the Kids sequence (original)

## 7. DISCUSSION

We have proposed an optimal boundary encoding algorithm which uses a second-order B-spline code structure, combined with a DAG shortest path formulation. Explicit utilization of the area-based additive distortion measure was made in the optimization process. The algorithm resulted in a 15%-20% rate reduction for comparable distortions when compared with the methods adopted by MPEG-4. Future research will consider quantized spline segments, and associated distortions, instead of their continuous counterparts. The possibility of using splines of degrees other than 2 and hybrid order splines will also be investigated. Finally, the adaptive switching between several Huffman tables depending on the target bit rate is currently under investigation.

## 8. REFERENCES

[1] N. Brady, F. Bossen, and N. Murphy, "Context-based arithmetic encoding of 2D shape sequences," *Proc. Int. Conf. Image Processing*, (Santa Barbara), pp. 29-32, 1997.

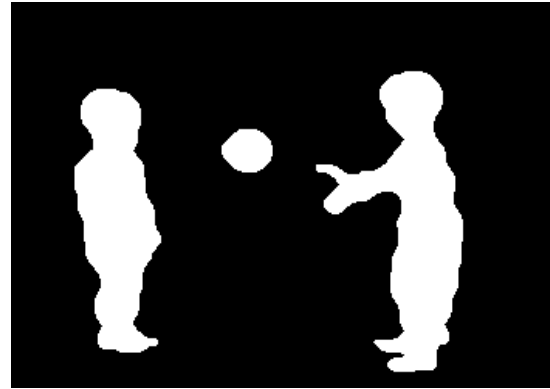


Figure 5: Frame 5 of the Kids sequence (approx.)

- [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] L. Kondi, F. W. Meier, G. M. Schuster, A. K. Katsaggelos, "Joint optimal object shape estimation and encoding", *Proc. SPIE Conf. Visual Communication and Image Processing*, vol. 3309, pp. 14-25, Jan. 1998.
- [4] S. Lee, D. Cho, Y. Cho, S. Son, E. Jang, and J. Shin "Binary shape coding using 1-D distance values from baseline", *Proc. Int. Conf. on Image Processing*, (Santa Barbara), pp. 508-511, 1997.
- [5] F. W. Meier, G. M. Schuster, and A. K. Katsaggelos "An efficient boundary encoding scheme which is optimal in the rate distortion sense", *Proc. Int. Conf. Image Processing*, (Santa Barbara), pp. 9-12, 1997.
- [6] K. J. O'Connell "Object-adaptive vertex-based shape coding method", *IEEE Trans. Circuits and Systems for Video Technology*, vol. 7, pp. 251-255, Feb. 1997.
- [7] G. M. Schuster and A. K. Katsaggelos, *Rate-Distortion Based Video Compression, Optimal Video frame compression and Object boundary encoding*. Kluwer Academic Press, 1997.
- [8] N. Yamaguchi, T. Ida, and T. Watanabe, "A binary shape coding method using modified MMR", *Proc. Int. Conf. Image Processing*, (Santa Barbara), pp. 504-508, 1997.
- [9] A. K. Katsaggelos, L. Kondi, F. W. Meier, J. Ostermann, G.M. Schuster, "MPEG-4 and Rate Distortion Based Shape Coding Techniques". *Proc. IEEE*, to appear, July 1998.
- [10] G.M. Schuster, A. K. Katsaggelos, "An optimal boundary encoding in the rate-distortion sense,". *IEEE Trans. Image Processing*, vol. 7, pp. 13-26, Jan. 1998.