SIMULTANEOUS OPTIMAL BOUNDARY ENCODING AND VARIABLE-LENGTH CODE SELECTION

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 describes efficient and optimal encoding and representation of object contours. Contours are approximated by connected second-order spline segments, each defined by three consecutive control points. The placement of the control points is done optimally in the rate-distortion (RD) sense and jointly with their entropy encoding. We utilize a differential scheme for the rate and an additive area-based metric for the distortion to formulate the problem as Lagrangian minimization. We investigate the sensitivity of the resulting operational RD curve on the variable length codes used and propose an iterative procedure arriving at the entropy representation of the original boundary for any given rate-distortion tradeoff.

1 Introduction

The MPEG-4 standardization effort has revived interest in object-oriented video compression and, therefore, boundary encoding techniques [3]. Research in this area is motivated by such important applications as content-based storage and retrieval, film authoring and mobile communications. In this framework, an image sequence is treated as a collection of disjoint video object planes, each of which is transmitted as texture, motion, and shape information. This bit-stream structure has necessitated development of better segmentation and boundary encoding tools. Within the confines of this structure, it is highly desirable to allocate available bits optimally among bitstream components and within each component.

Currently, in MPEG-4, boundary encoding is treated independently from the boundary estimation step (preliminary efforts have been recently made to couple these two steps [4]). Two types of binary shape coders are considered: bitmap-based and contour-based [3]. The context-based coder [1] and the Modified Modified Read coder [11] belong to the first type. The baseline-based shape coder [5] and vertex-based polynomial coder [2, 7] belong to the second type.

Measuring the distortion between an original and an approximating boundary is a challenging problem. Various distortion metrics are considered in [3, 10]. In MPEG-4 the following distortion metric has been used in evaluating the performance of each algorithm

$$D = \frac{\text{number of pixels in error}}{\text{number of interior pixels}} \qquad (1)$$

However, none of the algorithms considered by MPEG-4 uses the metric of Eq. (1) in the development of the algorithm. Furthermore, none of the algorithms which appeared in the literature can claim optimality in the rate-distortion sense, using the metric of Eq. (1) or any other metric. The first objective of this work is to obtain an optimal in the RD sense approximation of a given boundary using second order splines and the distortion measure of Eq. (1). We have previously proposed optimal approximations of a given boundary using distortion metrics other than the one in Eq. (1) and curves of different orders [3]. In [6] we proposed an optimal approximation using splines and a distortion metric resembling the one in Eq. (1), where in the numerator the area between the original boundary and its continuous approximation was used.

The operationally optimal shape encoding strategies we have considered thus far claim optimality only with respect to the chosen representation of the control points of the curve and their associated variable-length codes (VLC). Therefore, the second objective of this paper is the investigation of the sensitivity of the optimal Operational Rate-Distortion (ORD) curves on the underlying probability model used to derive the VLCs. Furthermore, we propose an iterative procedure for finding the probability model which is locally optimal with respect to the ORD efficiency.

This paper is organized as follows. The algorithm structure is presented in Sec. 2. The additive distortion metric used and the control point encoding issues are described in Sec. 3. The problem is formulated as a graph shortest path problem in Sec. 4. VLC optimization is discussed in Sec. 5. Finally, the results are presented and discussed in Sections 6 and 7.

2 Algorithm

In this paper, we formulate the solution to the contour approximation problem to be optimal in the rate-distortion sense, where the distortion metric used is given by Eq. (1). That is, the distortion metric expresses the number of incorrectly assigned pixels relative to the number of pixels belonging to the object in the original shape. A pixel is judged to be assigned incorrectly if it is either in the interior of the original boundary, but not in the interior of the approximating boundary, or vice-versa.

The original boundary is approximated by 2^{nd}-order B-spline segments. A spline segment is completely defined by 3 consecutive control points (p_{u-1}, p_u, p_{u+1}). It is a parametric curve with parameter t taking values from 0 to 1, which starts at the midpoint between p_{u-1} and p_u and ends at the midpoint between p_u and p_{u+1}. Mathematically, the 2^{nd}-order spline segment used is defined in [6].

Besides solving the interpolation problem at the midpoints, the definition of the spline used makes it continuously differentiable everywhere, including the junction points.

In our implementation, we quantize continuous spline segments to fit the support grid of the original boundary in order to count the number of incorrectly assigned pixels. The decoder performs the same operation when it receives control points. Thus, a set of control points unambiguously defines an approximation to the original boundary.

In principle, control points defining constituent spline segments can be located anywhere in the image, as long as they are ordered. However, because distortion of more than a few pixels can not be tolerated in most applications, and due to computational complexity considerations, we define a region in space to which control points must belong. As shown in Fig. 1, this region is a band centered around the original boundary, with each control band pixel labeled by the index of the boundary pixel closest to it. Boundary pixels themselves are, therefore, ordered and labeled.

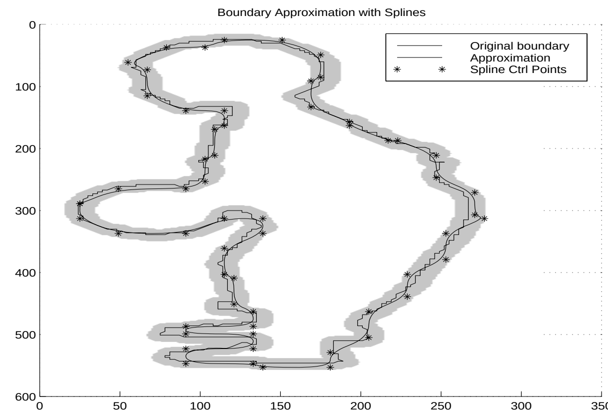

Figure 1: Admissible control point band.

3 Distortion and Rate

In order to define the total boundary distortion the segment distortion needs to be defined first. A segment of the approximating curve is first associated with a segment of the original boundary, as shown in Fig. 2. In it 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'. When more than one boundary pixel is a candidate, we select the one with the larger index. This assures us that starting boundary pixel of the next segment coincides with the last boundary pixel of the current segment. 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 segment distortion, as shown in Fig. 2. We exclude from the

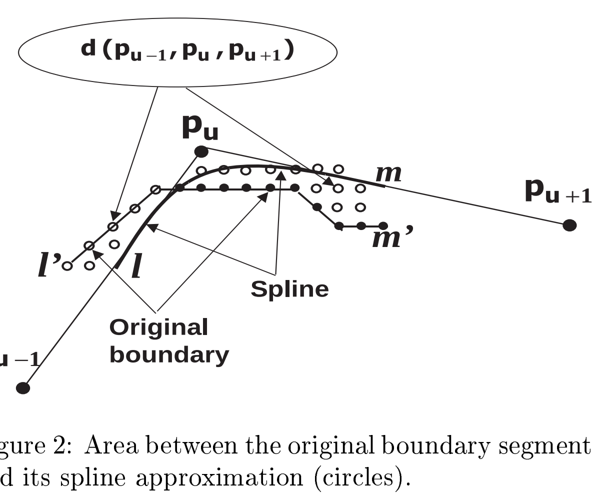

Figure 2: Area between the original boundary segment and its spline approximation (circles).

calculation of $d(p_{u-1}, p_u, p_{u+1})$ any pixels belonging to the line segment (m, m'), so that no pixel in error is counted more than once. Based on the segment distortions, the total boundary distortion is therefore defined by

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

where $p_{-1} = p_{N_p+1} = p_{N_p} = p_0$.
consecutive control point locations are de-correlated using a second-order prediction model [3]. In this model, each control point is described in terms of the relative angle α it forms with respect to the line connecting two previously encoded control points, and by the run length β (in pixels), as shown in Fig. (3A).

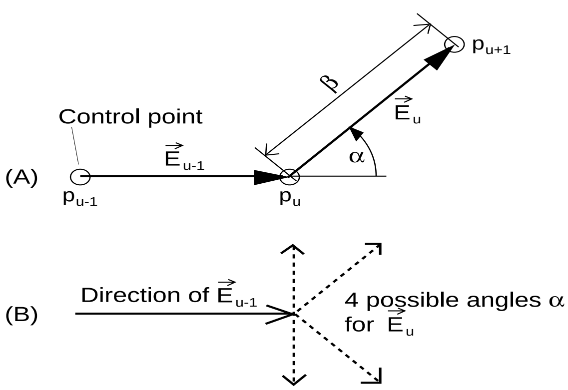

Figure 3: Encoding of a spline control point

Specifically, the angle α takes on values from the set $\{-90^o, -45^o, 45^o, 90^o\}$, thus requiring only 2 bits (Fig. 3B). Angle 0^o was excluded because that orientation is approximately realizable by appropriately placing the preceding control point. The only exeption to this rule is the encoding of the first and the second control points, for which the predictive context does not exist. We also force the control point band of the last boundary point, in the case of a closed boundary, to collapse to just the boundary pixel itself to ensure that the approximation yields a closed contour.

It should be noted, however, that the chosen code structure $(run, angle)$ is somewhat arbitrary. Other DPCM techniques or a different set of permissible angles could be used without loss of generality.

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

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

The issue of selecting efficient variable-length codes for the run β will be discussed in the Sec. 5.

4 Directed Acyclic Graph (DAG) solution

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

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

where both the location of the control points p_i and their overall number N_P have to be determined. The solution to this problem implies that no other selection of control points would result in a lower distortion with a rate lower or equal to R_{max}. It should be understood that 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 control band width and the ordering scheme for boundary points.

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

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

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

$$\{p_0^*, \ldots, p_{N_P-1}^*\} = arg \min_{p_0, \ldots, p_{N_P-1}} J_\lambda^*(p_0, \ldots, p_{N_P-1}) \qquad (6)$$

results in $R(p_0, \ldots, p_{N_P-1}) = R_{max}$. In this case $(p_0^*, \ldots, p_{N_P-1}^*)$ is the optimal solution. We employ a Bezier curve search [9] in order to arrive at λ^* in very few iterations.

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

$$w(p_{u-1}, p_u, p_{u+1}) = d(p_{u-1}, p_u, p_{u+1}) + \lambda \cdot r(p_{u-1}, p_u, p_{u+1}). \qquad (7)$$

Then the overall Lagrangian cost function can be written as,

$$J_\lambda(p_0, \ldots, p_{N_P-1}) = \sum_{u=1}^{N_P-1} w(p_{u-1}, p_u, p_{u+1}) + w(p_{N_P-1}, p_{N_P}, p_{N_P+1}) = J_\lambda(p_0, \ldots, p_{N_P-2}) + w(p_{N_P-1}, p_{N_P}, p_{N_P+1}). \qquad (8)$$

This structure of the problem makes dynamic programming a natural choice of the minimization method. Specifically, this problem is cast as the shortest path problem in a graph with each consecutive pair of control points playing the role of a vertex and incremental costs $w()$ serving as the corresponding weights [3]. This resulting Directed Acyclic Graph (DAG) is searched to find the optimal set of control points [3, 9].

5 VLC Model Optimization

Optimality of the previously proposed vertex-based algorithms ([3, 6]) could be guaranteed only within the confines of the chosen code structures and the chosen VLCs. That is, the following optimization problem was solved:

$$\{p_0^*, \ldots, p_{N_P-1}^*\} = arg[\min_{p_0,\ldots,p_{N_P-1}} J_\lambda^*(p_0,\ldots,p_{N_P-1})|VLC_{fixed}] \quad (9)$$

In this paper, we take this approach a step further by removing the conditioning of the optimal solution on some ad-hoc VLC_{fixed}. As shown in Fig. 4, we employ an iterative algorithm [8] to jointly find the locally optimal parameter distribution model f and the boundary approximation. Hence we solve the following problem:

$$\{p_0^*, \ldots, p_{N_P-1}^*\} = arg \min_{p_0,\ldots,p_{N_P-1}; f\in F} J_\lambda^*(p_0,\ldots,p_{N_P-1}) \quad (10)$$

The iteration process begins with the proposed encoder compressing the input binary images based on a given rate-distortion tradeoff λ and the jointly uniform conditional distribution model for the (*run, angle*) symbol of vertex i, conditioned on the previously encoded *run*. Mathematically, the distribution $f^k(run_i, angle_i|run_{i-1})$ is uniform at iteration $k = 1$.

It should be noted that symbol probabilities rather than codeword lengths are used by the encoder in this iterative process. Hence, symbol entropies $-p\log p$ take place of the bit-rate $r()$ and $R()$ in Eqs. (5,7).

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 $J_\lambda^k()$ in Eq. (5) is a non-increasing function of the iteration k. The iterations stop when $|J_\lambda^k() - J_\lambda^{k-1}()| \le \epsilon$.

This procedure is guaranteed to converge to a local minimum with respect to small perturbation of the joint probability mass function f^M, where M is the number of the last iteration. The resulting symbols (*run,angle*) are then arithmetically encoded and sent to the decoder together with the model f^M.

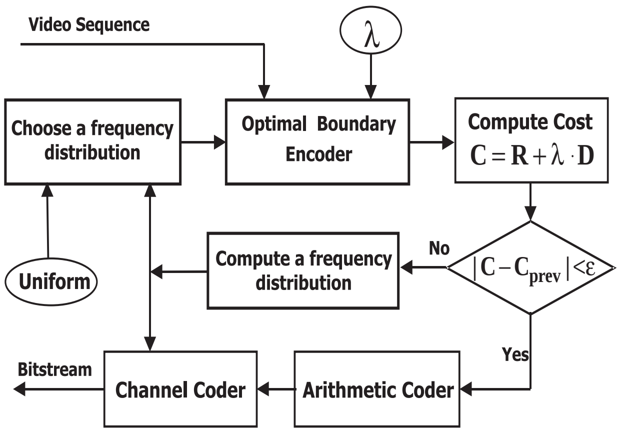

Figure 4: The entropy encoder structure.

6 Results

Figure 5 shows the ORD curve of the proposed algorithm for the SIF sequence "kids" using several different VLCs for the run parameter β. The distortion axis represents the average of the D's defined in Eq.(1) for one frame, over 100 frames. Our results compare favorably with the results of both contour-based and pixel-based algorithms reported in [3]. The proposed approach also outperforms the ORD curve in [6], shown with squares on the plot. 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 evaluated by MPEG-4 [3]. Based on comparing the best ORD curve of Fig. 5 with that produced by the baseline method, the proposed approach performs better or the same as the baseline method for average frame bit rate of 1400 or less. In the very low distortion region ($D \le 0.01$) of operation, however, the proposed algorithm requires more bits than baseline, primarily because of the code structure (angle plus length) and sub-optimality of our code table under these operating conditions. Visually, the accuracy of the optimal solution is demonstrated in Fig. 6, where the original shape is shown in grey and the error is shown in white. This approximation is optimal for frame rate $R = 570$ bits. The optimal way to encode some objects, it turns out, is not to encode them at all, as demonstrated by the space between the legs of the kid on the left, which is shown in white as an erroneous area. In the course of running the experiments we tried several different VLCs for the run length parameter β, as shown in Fig. 5. Clearly, the resulting distortion (D) is extremely sensitive to the VLC and can vary by as much as 50% for the same bit-rate. Figure 5 also shows the

ORD curve based on the locally optimal entropy encoding of the boundaries. It should be noted that, for this curve, entropy was used in place of the bit-rate R, and transmission of the model f^M was not accounted for. The first qualification, however, is not very significant, since the subsequent arithmetic encoding of output symbols is expected to compress the bit-rate close to its entropy.

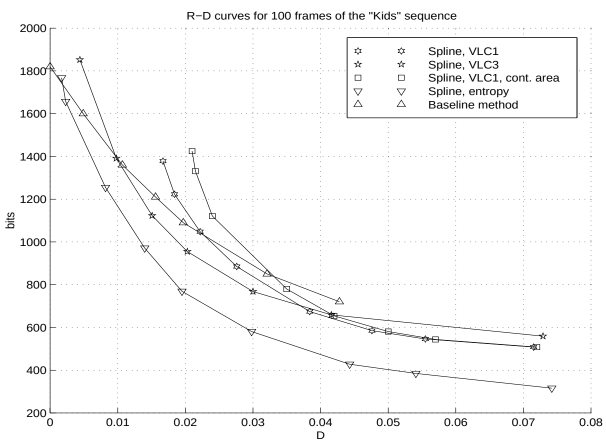

Figure 5: Rate-Distortion curves.

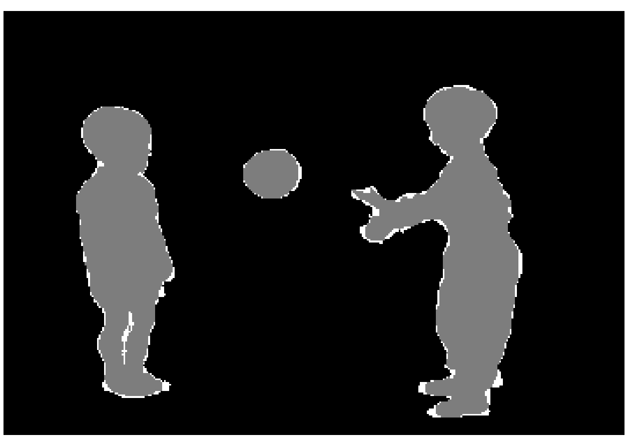

Figure 6: Shape approximation errors (white pixels).

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 pixel-based additive distortion measure was made in the optimization process. The algorithm resulted in a rate reduction (in some cases as high as 20%) for comparable distortions when compared with the methods considered by MPEG-4. It also outperformed the optimal method in which a continuous area approximation was used. This paper also demonstrated the sensitivity of the resulting ORD curve to the VLC table used to encode the control points. We proposed an iterative procedure, which, based on the proposed optimal algorithm, finds both the parameter distribution model and the boundary approximation jointly and optimally.

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". *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. Intl. Symposium on Circiuts and Systems*, June 1998.

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

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

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

[10] 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.

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