JOINT RATE DISTORTION OPTIMAL SHAPE AND TEXTURE CODING

Saurav K. Bandyopadhyay, Lisimachos P. Kondi
Multimedia Communications Laboratory, Dept. of Electrical Engineering,
University at Buffalo, The State University of New York
Buffalo, NY – 14260, USA.
e-mail: {skb3, lkondi}@eng.buffalo.edu

Abstract

In this paper, the problem of joint shape and texture coding in video compression is considered. The shape is approximated using polygons or higher order curves. The texture is encoded using Shape Adaptive Discrete Cosine Transform (SA-DCT). The solution is optimal in the operational rate distortion sense i.e. given a coding setup, the solution will guarantee the smallest possible total rate for a given distortion for both shape and texture coding or the smallest possible distortion for a given rate. Both a fixed-width and a variable width distortion band for shape coding were considered. The variable width of the distortion band is a function of the texture profile i.e. the width is inversely proportional to the magnitude of the image gradient. We present experimental results using the “Kids” and “Bream” sequences.

Keywords: *Shape Coding, Texture Coding, Video Coding, Rate Distortion theory, SA-DCT.*

1. INTRODUCTION

The second generation video coding techniques represent an image by its set of constituent objects or features. Thus, the smallest entity in an image is an object with its associated shape, texture and motion information. However, in order to build an efficient video encoder, the optimal bit allocation between shape and texture is necessary. Again, the coding of the shape is not independent of the coding of the texture of the object. In [1], a vertex based shape coding method is proposed that takes into consideration the texture information. It utilizes the texture information to create a variable-width distortion band. The width of the distortion band is inversely proportional to the magnitude of the image gradient. In [2], a joint shape and texture rate control algorithm for MPEG-4 encoders is proposed. In [3], the operational rate distortion optimal bit allocation between shape and texture for MPEG-4 video coding is proposed.

In this paper, we propose an operational rate distortion optimal bit allocation scheme for shape and texture for object based video. The algorithm is based on the use of polygons or B-splines to encode the shape information and SA-DCT to encode the texture. The solution is optimal in the operational rate distortion sense. For the polygon approximation we also considered biasing the cost function to favor horizontal and vertical edges (biased polygon approximation).

The rest of the paper is organized as follows. The next section briefly describes the methodologies used for shape and texture coding. In section 3 the problem formulation is presented. Section 4 demonstrates the optimal solution. Section 5 provides our experimental results. In section 6 we draw conclusions.

2. OVERVIEW OF ENCODING

In this section the shape and texture coding techniques used are reviewed. Further details can be found in [4, 6, 8].

2.1 Shape Coding

The goal of the shape coding is to encode the shape information of a video object to enable applications requiring content-based video access. Several shape coding techniques can be used for the purpose viz. bitmap-based, contour-based and implicit shape coders. We have used a contour-based shape coder for our purpose. The shape is approximated using a polygon or B-splines for lossless shape coding. In all cases the problem reduces to finding the shortest path in a directed acyclic graph (DAG). Both a fixed-width and a variable-width distortion band were considered. The reader interested in the details of contour based shape coding is referred to [1, 4, 6].

2.1.1 Notations. Let $B = \{b_0,.....,b_{N_B-1}\}$ denote a connected boundary which is an ordered set, where b_j is the j -th point of B and N_B is the total number of points in B . In the case of a closed boundary, $b_0 = b_{N_B-1}$. Let $P = \{p_0,....., p_{N_P+1}\}$ denote the set of control points of the B-spline curve, which is also an ordered set with N_P the total number of curve segments. In case of polygon approximation let the set of control points be denoted by $P = \{p_0,....., p_{N_P}\}$

2.1.2 B-splines. B-splines represent a family of parametric curves [8] which is very useful for boundary encoding. In this paper, we approximate boundaries with quadratic B-splines with a method that is based on the shortest path algorithm for a weighted directed acyclic graph (DAG) [10]. The motivation in

CCECE 2004- CCGEI 2004, Niagara Falls, May/mai 2004
0-7803-8253-6/04/$17.00 ©2004 IEEE

using B-splines is better coding efficiency for objects in natural images. Such objects tend to have fewer straight lines and narrow corners. A B-spline curve segment is defined by three control points. The corresponding bit rate and segment distortion are given by $r(p_{k-1}, p_k, p_{k+1})$ and $d(p_{k-1}, p_k, p_{k+1})$.

2.1.3 Polygons. A polygon edge is defined by two control points, its vertices. The control point rates and the segment distortion depend on two points and are given by $r(p_{k-1}, p_k)$ and $d(p_{k-1}, p_k)$. For details the reader is referred to [4], [5].

2.1.4 Biased Polygon Approximation. The SA-DCT is expected to be more efficient if the edges of the object are horizontal and vertical. However, it will be inefficient for shape coding compared to polygons, B-splines. We allow a *bias < 1* multiplicative factor for rates of points p_{k-1} and p_k which corresponds to horizontal and vertical edges. Thus,

$$\omega(p_{k-1}, p_k) = bias.[r(p_{k-1}, p_k)]$$

if p_{k-1} and p_k define a horizontal and vertical edge. Thus the boundary algorithm will favor horizontal and vertical edges at the expense of increased bit rate for shape coding.

2.1.5 Distortion Band. A fixed width distortion band has a width $2 * \delta_{\max}$ along the boundary B. The approximating contour must lie within the distortion band. However, a variable width distortion band requires a $\delta_{\max}$ for every boundary point. We denote this by $\delta_{\max}[i],\quad i = 0,......, N_B - 1$. In order to construct the distortion band we draw circles from each boundary point b_i with width $\delta_{\max}[i]$. The distortion band consists of the set of points inside the circles.

Considering B-spline curve the segment distortion measure is given by:

$$d(p_{k-1}, p_k, p_{k+1}) = \begin{cases} 0: & \text{all points of } Q(p_{k-1}, p_k, p_{k+1}) \\ & \text{are inside the distortion band} \\ \infty: & \text{any point of } Q(p_{k-1}, p_k, p_{k+1}) \\ & \text{is outside the distortion band} \end{cases}$$

In order to find $\delta_{\max}[i]$ the gradient is computed for an image $f(x, y)$ as:

$$\nabla f(x, y) = \left[\partial f/\partial x \quad \partial f/\partial y \right]^T = \left[f_x \quad f_y \right]^T$$

The Sobel edge detector mask is used to compute the gradient given by:

$$|\nabla f(x, y)| = \sqrt{f_x^2(x, y) + f_y^2(x, y)}$$

Let the minimum and maximum magnitude of the gradient magnitude for the whole image be *gradmin* and *gradmax*. Again, considering the desired minimum and maximum values of $\delta_{\max}[i]$ as $T_{\min}$ and $T_{\max}$ respectively, the width of the distortion band at a point is given by:

$$\delta_{\max}[i] = T_{\min} + \lambda \cdot (grad[i] - gradmax)$$

where,

$$\lambda = \frac{T_{\max} - T_{\min}}{gradmin - gradmax}$$

A threshold is defined for the gradient magnitude. The boundary points whose gradient magnitude exceeds the threshold should have the minimum possible $\delta_{\max}[i]$. In that case, *gradmax* is equal to the threshold.

2.2 Texture Coding

The texture content of each block depends on the reconstructed shape information. It is encoded using Shape-Adaptive Discrete Cosine Transform (SA-DCT). To achieve higher coding efficiency we consider the tightest rectangular boundary of multiple of 8 along both axes to encompass the arbitrarily shaped Video Object. Usually, video frames are encoded by taking the Discrete Cosine Transform of 8x8 image blocks. For the texture of objects, however some 8x8 blocks that are close to the boundary will be partially occupied by the object. Using an 8x8 DCT, we would need to transmit 64 coefficients for these blocks, although the actual number of pels would be smaller. SA-DCT provides for a way of encoding such blocks using a number of coefficients that is equal to the number of object pels in the block. This is accomplished by shifting the object pels towards the origin of the block and then taking one dimensional DCTs row-wise and then column-wise. The length of these one-dimensional DCTs can be less than eight.

Compression is accomplished by quantization of the DCT coefficients followed by entropy coding. It is expected that the block contains an edge or an area with high gradient magnitude, the entropy of the quantized DCT coefficients of this block will be high and a large number of bits will be required for its encoding. If SA-DCT is used and the segmentation is based on gradient, it is beneficial for the shape information to be accurate in areas with high gradient. This way, the number of areas of higher gradient within the object will be minimized and the encoding efficiency of the SA-DCT will be higher.

3. PROBLEM FORMULATION

The goal is to optimally allocate bits between shape and texture to transmit the intra-coded frame at a given acceptable level of quality. Hence the optimization problem can be written as follows:

$$\text{Minimize } R \text{, subject to } D \le D_{\max} \quad (1)$$

where R is the total bit rate for shape and texture given by:

$$R = R_{shape} + R_{texture} \quad (2)$$

R_{shape}, $R_{texture}$ are the rates for shape and texture coding respectively. D is the distortion for texture, and $D_{\max}$ is the maximum allowable distortion. The shape approximation is restricted to lie within the distortion band.

In the same line as above we can solve the dual problem, that is,

$$\text{Minimize } D \text{, subject to } R \le R_{budget} \quad (3)$$

where R_{budget} is the total bit budget for shape and texture coding and D is the distortion for texture.

We solve the optimization problem as in (1) for six discrete cases. The shape of the object is encoded using polygons, B-splines and biased polygon approximation. The fixed width and the variable width of the distortion band are considered for each of the shape coding techniques. Obviously the best set of coding parameters will yield the optimal R-D curves. In case of fixed width of the distortion band the parameter of interest is the band width. Considering variable width the band is determined by the threshold, minimum $(T_{\min})$ and maximum $(T_{\max})$ width of the distortion band. Again, the biased polygon approximation is determined by a specific bias for the horizontal and vertical edges.

3.1 Modeling the Distortion

The distortion is calculated based on the mean square error (MSE) of the reconstructed and the original image in the region of the encoded shape. The distortion can thus be written as

$$D = \frac{1}{N} \sum_{(x,y)\in C} \delta(x,y)^2 \quad (4)$$

where $\delta(x,y)$ is the differential intensity value at pixel position (x,y), C is the intersection of the set of pixels of the original and the reconstructed shapes. N is the number of pixels in C.
The peak signal-to-noise ratio (PSNR) is calculated using

$$PSNR = 10\log_{10} \frac{255^2}{D} \quad (5)$$

4. OPTIMAL SOLUTION

Let U denote the possible solutions for joint coding in one of the six discrete cases mentioned above.
Thus (1) can be written as

$$\text{Minimize } R(U) \text{, subject to } D(U) \le D_{\max} \quad (6)$$

where $R(U) = R_{shape}(U) + R_{texture}(U)$

The constrained minimization problem stated in (6) is converted to an unconstrained problem by using the Lagrangian multiplier method.

$$\begin{aligned} J_\lambda(U) &= R + \lambda \cdot D \\ &= R(U) + \lambda \cdot D(U) \end{aligned} \quad (7)$$

where λ is the Lagrangian multiplier. Now according to [7] if there is a λ^* for which $U^* = \arg\left[\min_u J_\lambda(U)\right]$ and which satisfies $D = D_{\max}$ then U^* is an optimal solution of (6). In order to find the optimal solution λ traces the convex hull of the operational rate distortion function. So λ^* can be found using bisection or fast convex search algorithm. Therefore, if we can find the optimal solution to the unconstrained problem

$$\min\left[R(U) + \lambda \cdot D(U)\right] \quad (8)$$

we can find the optimal λ^* and the convex hull approximation to (6).

To implement the bisection algorithm to solve (8) we formulate a cost function $C(U)$ which represents the minimum total rate and distortion for the frame. Clearly,

$$J_\lambda(U) = \min\ C(U)$$

5. EXPERIMENTAL RESULTS

We coded both the shape and the texture of frame 0 of the “Kids” and “Bream” sequences. Number of experiments were conducted, some of which are reported below.

In one experiment we considered fixed width of the distortion band. The fixed width of the distortion band is varied from 0.8 pels to 3.0 pels. In another experiment we considered variable width of the distortion band. The threshold is varied from 100 to 600, $T_{\min} = 0.8$ and $T_{\max} = 3.0$. In both experiments B-splines, polygons and biased polygon approximation were used for the boundary approximation. When considering biased polygon approximation the degree of favoring the horizontal and vertical edges is varied from 0.1 to 0.9. The texture of the object is coded using Shape-Adaptive DCT with the Quantization Parameter (QP) varied from 2 (fine quantization) to 32 (coarse quantization).
Comparison of the results from the bream sequence for fixed width of the distortion band are shown below:

Contour coding used	Shape Rate (bits)	Texture Rate (bits)	Total Rate (bits)	Distortion	PSNR (dB)
B splines (Width=2.4)	256	30205	30461	6.31	40.13
Polygon (Width = 1.6)	273	30188	30461	6.34	40.11
Biased polygon approximation (Width = 3.0, Bias = 0.2)	433	29568	30001	6.23	40.18

Table 1. Results of the first experiment for QP = 2

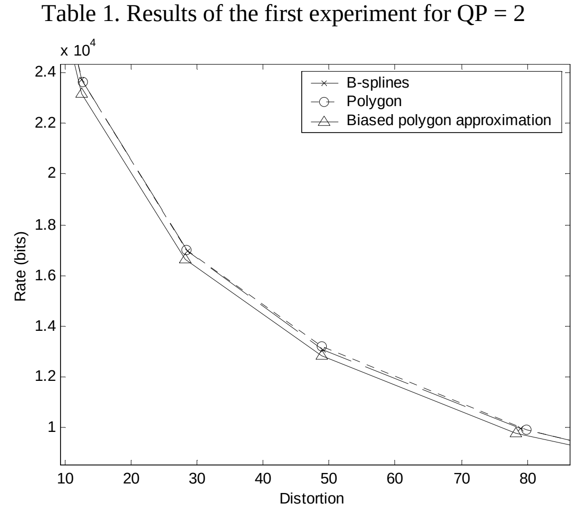

Fig 1. Results of the first experiment

Again comparison of the results from the bream sequence for variable width of the distortion band are shown below:

Contour coding used	Shape Rate (bits)	Texture Rate (bits)	Total Rate (bits)	Distortion	PSNR (dB)
B splines (Threshold=400)	258	30462	30720	6.3	40.14
Polygon (Threshold=400)	255	29884	30139	6.45	40.03
Biased polygon approximation (Threshold=600, Bias = 0.3)	342	29551	29893	6.29	40.14

Table 2. Results of the second experiment for QP = 2.

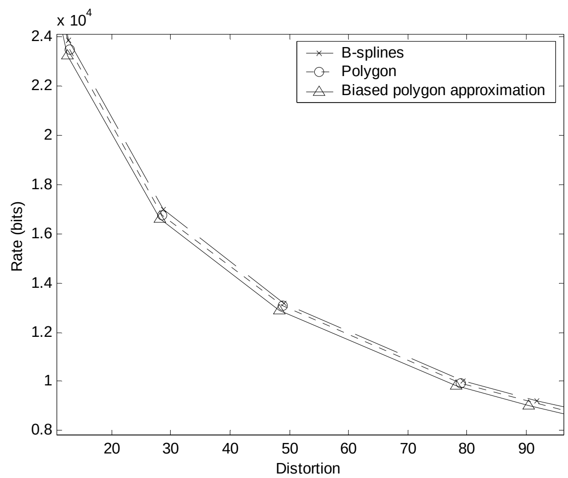

Fig 2. Results of the second experiment

For a Quantization Parameter (QP) = 2, the cases requiring the highest and lowest rate are presented below:

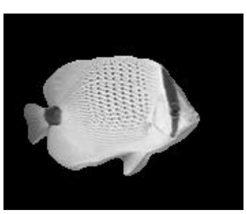

Fig 3. Variable width
with B-splines
Threshold =400

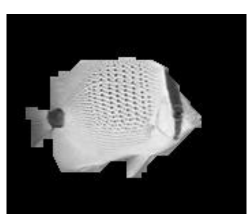

Fig 4. Variable width
with biased polygon approx.
Threshold = 600, Bias = 0.3

In both Table 1 and Table 2 the number of bits required for shape and texture coding is shown with the corresponding distortion and PSNR. It is seen that using a variable width of the distortion band we obtain fewer bits. The number of bits for joint coding of shape and texture is significantly reduced when using the variable width distortion band with bias.

6. CONCLUSIONS

We presented an operational rate-distortion optimal bit allocation scheme between texture and shape for the encoding of the object based video. The contour-based shape coding technique performs better using B-splines than using polygons. The main reason is that B-splines have a more natural appearance than polygons. It is obvious that the biased polygon approximation is the least efficient amongst the considered cases for shape coding. However, we used it for efficient texture coding using SA-DCT. As expected, shape coding using biased polygon approximation leads to better performance in texture encoding using SA-DCT than both polygons and B-splines. The optimal solution is determined using the Lagrangian multiplier method. The optimal approach has higher efficiency than an exhaustive search algorithm.

References

[1] L.P. Kondi, G. Melnikov, and A.K. Katsaggelos, "Joint optimal coding of texture and shape", in *Proc. IEEE International Conference on Image Processing,* Volume III, pp. 94-97, Thessaloniki, Greece, October 2001.

[2] A. Vetro, H. Sun, and Y. Wang, "Joint shape and texture rate control for MPEG-4 encoders", in *Proc. IEEE International Conference on Circuits and Systems,* pp. 285-288, Montrey, USA, June 1998.

[3] H Wang, G. M. Schuster, and A.K. Katsaggelos, "Object-based video compression scheme with optimal bit allocation among shape, motion and texture", in *Proc. IEEE International Conference on Image Processing,* Volume III, pp. 785-788, Barcelona, Spain, September 2003.

[4] A.K. Katsaggelos, L.P. Kondi, F.W. Meier, J. Ostermann, G.M. Schuster, "MPEG-4 and rate-distortion based shape-coding techniques", in *Proc. IEEE,* Vol. 86, Issue. 6, pp. 1126-1154, June 1998.

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

[6] L.P. Kondi, F.W. Meier, G.M. Schuster, and A.K. Katsaggelos, "Joint optimal object shape estimation and encoding", in *Proc. Conference on Visual and Image Processing,* pp. 14-25, San Jose, California, January 1998.

[7] T. Cormen, C. Leiserson, and R. Rivest, *Introduction to algorithms,* McGraw-Hill Book Company, 1991.

[8] L.P. Kondi, G. Melnikov, A.K. Katsaggelos, "Joint optimal object shape estimation and encoding", *IEEE Trans. on Circuits and Systems for Video Technology, April 2004, to appear*.

[9] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers", *IEEE Trans. Acoust., Speech, Signal Processing,* Vol. 36, pp. 1445-1453, September 1988.

[10] G.M. Schuster and A.K. Katsaggelos, "An optimal polygonal boundary encoding scheme in the rate distortion sense", *IEEE Trans. Image Processing,* pp. 13-26, January 1998.