OPTIMAL BIT ALLOCATION FOR JOINT CONTOUR-BASED SHAPE CODING AND SHAPE ADAPTIVE 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
email: {skb3, lkondi}@eng.buffalo.edu

ABSTRACT

In this paper, an optimal framework is proposed for the joint encoding of the shape and texture information in object based video. The solution is optimal in the operational rate distortion sense, i.e, given the coding setup, the solution will guarantee the smallest possible distortion for a given rate. The shape is approximated using polygons or higher order curves. We also consider biasing the cost function to favor horizontal and vertical edges, for the case of polygon approximation (biased polygon approximation). The texture is encoded using Shape Adaptive Discrete Cosine Transform (SA-DCT) or Shape Adaptive Discrete Wavelet Transform (SA-DWT) of the MPEG-4 video codec. A comparison is drawn between the two techniques. Both a fixed-width and a variable-width tolerance band for shape coding are considered. The variable width of the tolerance band is a function of the texture profile, i.e, the width is inversely proportional to the magnitude of the image gradient. Experimental results are presented and conclusions are drawn.

1. INTRODUCTION

The second generation video coding techniques represent an image by the shape, motion and texture information of its constituent objects. However, in order to build an efficient video encoder, 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 other words a joint coding scheme needs to be developed.

In [1], [2] 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 tolerance band. However, no scheme for the optimal bit allocation between texture and shape is provided. In [3], a joint shape and texture rate control algorithm for MPEG-4 encoders is proposed. However, no rate-distortion optimal solution is provided. In [4], bit allocation between shape and texture for MPEG-4 codec is provided. However, only bitmap-based shape coding is considered and no shape adaptive texture coding techniques are used.

In this paper, we propose an operational rate distortion optimal bit allocation scheme for shape and texture components. The algorithm is based on the use of polygons or B-splines to encode the shape and SA-DCT or SA-DWT of the MPEG-4 codec to encode the texture. Contour based shape coding technique is known to perform better than the bitmap-based shape encoding method adopted by MPEG-4 [5]. Hence, our algorithm becomes a natural choice. 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. In Section 2 we discuss the contour based shape coding and in Section 3 we discuss the SA-DCT and SA-DWT based texture coding techniques used. In Section 4 the problem formulation and the optimal solution are presented. Section 5 provides our experimental results. In section 6 we draw conclusions.

2. SHAPE CODING

The goal of shape coding is to encode the shape information of a video object to enable applications requiring content-based video access. We have used a contour-based shape coder for our purpose. The shape is approximated using a polygon or B-splines for lossy 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 tolerance band were considered. The reader interested in the details of contour based shape coding is referred to [5].

2.1. Notations

Let $B = b_0, \ldots, b_{N_B-1}$ 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. Let $P = p_0, ..., p_{N_{p-1}}$ denote the set of control points of the polygon approximation, which is also an ordered set with p_k the kth vertex of P and N_p the total number of vertices in P. The kth edge starts at p_{k-1} and ends at p_k. In case of a closed boundary, $b_0 = b_{N_B-1}$. A polygon edge is defined by two control points, its vertices. B-splines can be used to approximate the contour instead of a polygon. A B-spline curve segment is defined by three control points.

2.2. Tolerance Band

A fixed-width tolerance band has a width D_{max} along the boundary B. The approximating contour must lie within the tolerance band.

A variable-width tolerance band requires a D_{max} for every boundary point. We denote this by $D_{max}[i]$, $i = 0, ..., N_B - 1$ where N_B is the number of boundary points. In order to construct the tolerance band we draw circles from each boundary point b_i

0-7803-9134-9/05/$20.00 ©2005 IEEE

with diameter $D_{max}[i]$. The tolerance band consists of the set of points inside the circles. In order to find $D_{max}[i]$, the gradient is computed for an image $f(x,y)$ as:

$$\nabla f(x,y) = \left[\frac{\partial f}{\partial x} \; \frac{\partial f}{\partial y}\right]^T = [f_x \; f_y]^T. \quad (1)$$

In practice, the Sobel edge detector masks are used to calculate the gradient. The magnitude of the gradient is then computed, that is,

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

Let us denote by $gradmin$ and $gradmax$, respectively, the minimum and maximum of the magnitude of the image gradient for the whole image. Let us also denote the desired minimum and maximum values of $D_{max}[i]$ as T_{min} and T_{max}, respectively. Then, a linear mapping is performed between the gradient value of each boundary point and the width of the tolerance band. If the magnitude of the gradient at the boundary point b_i is $grad[i]$, then the width of the tolerance band at this point is given by:

$$D_{max}[i] = T_{min} + \xi(grad[i] - gradmax), \quad (3)$$

where

$$\xi = \frac{T_{max} - T_{min}}{gradmin - gradmax}. \quad (4)$$

In practice, we need to define a threshold (Th) for the gradient magnitude. The boundary points whose gradient magnitude exceeds the threshold should have the minimum possible $D_{max}[i]$. Therefore,

$$D_{max}[i] = T_{min}, \text{if } grad[i] \geq Th \quad (5)$$

2.3. Shape Coding Problem Formulation

In this paper, we approximate boundaries using B-splines, polygons and the biased polygon approximation. 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 $d(p_{k-1}, p_k)$. The bit rate $R(p_0, \ldots, p_{N_{p-1}})$ for the entire curve is given by,

$$R(p_0, \ldots, p_{N_{p-1}}) = \sum_{k=0}^{N_{p-1}} r(p_{k-1}, p_k). \quad (6)$$

The segment distortion for the polygon case is given as

$$d(p_{k-1}, p_k) = \begin{cases} 0: & \text{all points of } G_k(p_{k-1}, p_k) \\ & \text{are inside the tolerance band} \\ \infty: & \text{any point of } G_k(p_{k-1}, p_k) \\ & \text{is outside the tolerance band} \end{cases} \quad (7)$$

Eq. (7) takes a curve segment G_k defined by two control points p_{k-1} and p_k and checks if the curve segment lies with the tolerance band. The curve distortion can be expressed in terms of the segment distortion as,

$$D(p_0, \ldots, p_{N_p-1}) = \max_{k \in [1, \ldots, N_{p-1}]} d(p_{k-1}, p_k). \quad (8)$$

Thus the optimization problem we are solving for the shape coding is:

$$\min \; R(p_0, \ldots, p_{N_p-1}), \text{ subject to: } D(p_0, \ldots, p_{N_p-1}) = 0. \quad (9)$$

This problem reduces to finding the shortest path in a directed acyclic graph (DAG) [5].

The shape-adaptive texture coding using SA-DCT or SA-DWT is expected to be more efficient if the edges of the object are horizontal or vertical. We allow a $bias < 1$ multiplicative factor for the rates of segments defined by control points p_{k-1} and p_k, which correspond to horizontal or vertical edges [1]. Thus,

$$r'(p_{k-1}, p_k) = bias \times [r(p_{k-1}, p_k)] \quad (10)$$

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

B-splines can also be used for shape coding instead of a polygon. The motivation in using B-splines is better coding efficiency for objects in natural images. Such objects tend to have fewer straight lines and narrow corners. The problem formulation and solution are similar to the polygon case.

3. 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) or Shape-Adaptive Discrete Wavelet Transform (SA-DWT) using an MPEG-4 compliant codec.

SA-DCT [7] provides a way of encoding 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. The SA-DCT always shifts the samples in an arbitrarily shaped block towards a certain edge of the rectangular bounding box before performing 1-D DCT along the row or the column. Hence, some spatial correlation may be lost. Again, intuitively, it is not efficient to perform column DCT on a set of coefficients that are from different frequency bands after the row DCT transforms.

To overcome the above difficulties we consider the shape adaptive discrete wavelet transform (SA-DWT) [8] for coding arbitrarily shaped still texture. The SA-DWT transforms the samples in the arbitrary shaped region into the same number of coefficients in the subband domain while keeping the spatial correlation, locality property and self-similarity across subbands. Encoding and decoding the SA-DWT coefficients are the same as encoding and decoding the regular wavelet coefficients except for keeping track of the locations of the wavelet coefficients according to the shape information. A 2D SA-DWT is applied to the texture object, producing the dc component (low frequency subbands) and a number of ac (high-frequency) subbands. The dc subband is quantized, predictively encoded (using a form of DPCM) and entropy coded using an arithmetic encoder. The ac subbands are quantized and reordered ("scanned"), ZeroTree encoded and entropy coded.

4. PROBLEM FORMULATION

The shape of the object is encoded using polygons, B-splines and the biased polygon approximation. Both the fixed-width and the variable-width of the tolerance band are considered for each of these three boundary approximations giving a total of six different shape coding techniques as follows:

1. Polygons using fixed-width of the tolerance band

2. Polygons using variable-width of the tolerance band
3. B-splines using fixed-width of the tolerance band
4. B-splines using variable-width of the tolerance band
5. Biased polygons using fixed-width of the tolerance band
6. Biased polygons using variable-width of the tolerance band

In case of fixed-width of the tolerance band, the parameter of interest is the band width (D_{max}). Considering variable-width, the band is determined by the threshold (Th), minimum (T_{min}) and maximum (T_{max}) width of the tolerance band. In the biased polygon approximation case, the parameter of interest is the bias for the horizontal and vertical edges. When the texture is encoded using SA-DCT, the quantization parameter (QP) is varied and, in SA-DWT based texture coding, the embedded bit stream is decoded at different rates.

Let S be a set that contains the shape coding parameters. It thus specifies one of the six shape coding methods along with the parameters associated with it. For example, for the biased polygon approximation case with fixed width of the tolerance band, S contains the width of the tolerance band and the $bias$ for the horizontal and vertical edges. Let T denote the texture coding parameters (QP for SA-DCT based texture coding and decoded bitrate for texture coding using SA-DWT).

Hence the optimization problem can be written as follows:

$$S^*, T^* = \arg\min_{S,T} R_{total}(S,T) \tag{11}$$

subject to

$$D^{YUV}_{texture}(S,T) \leq D^{YUV}_{budget}$$

where, $R_{total}(S,T) = R_{shape}(S) + R_{texture}(S,T)$.

$D^{YUV}_{texture}(S,T)$ is the texture distortion of the image in the region of the reconstructed shape. The D^{YUV}_{budget} is the maximum allowable texture distortion. The distortion in all of our simulation results is calculated based on the Mean Square Error (MSE) between the original and the reconstructed image in the region of the reconstructed shape. The distortion can be mathematically written as:

$$D^{YUV}_{texture} = \frac{1}{N} \sum_{(x,y)\in C} \left\{\delta_y^2(x,y) + \delta_u^2(x,y) + \delta_v^2(x,y)\right\} \tag{12}$$

where $\delta_y(x,y)$ is the Y component differential intensity value at pixel position (x,y), $\delta_u(x,y)$ is the U component differential intensity value at pixel position (x,y), $\delta_v(x,y)$ is the V component differential intensity value at pixel position (x,y). The input video sequence is in YUV 4:2:0 color space. C is the reconstructed shape region. N is the total number of pels in the region of the reconstructed shape for Y, U and V components. The peak signal-to-noise ratio (PSNR) can be calculated using:

$$PSNR^{YUV}_{texture} = 10\log_{10}\left(\frac{255^2}{D^{YUV}_{texture}}\right). \tag{13}$$

The constrained minimization problem stated in Eq. (11) is converted to an unconstrained problem by using the Lagrangian multiplier method as

$$J_\lambda(S,T) = R_{total}(S,T) + \lambda \cdot D^{YUV}_{texture}(S,T) \tag{14}$$

where λ is the Lagrangian multiplier. Now, according to Eq. (14) if there is a λ^* for which $S^*, T^* = \arg\min_{S,T} J_\lambda(S,T)$ and which satisfies $D^{YUV}_{texture}(S^*,T^*) = D^{YUV}_{budget}$ then S^*, T^* is the optimal solution for Eq. (11). In order to find the optimal solution λ traces the convex hull of the operational rate distortion function. So, λ^* can be found using the bisection algorithm.

5. EXPERIMENTAL RESULTS

A number of experiments were conducted using different frames from the "Bream" sequence and "Children" sequence, some of which are reported below. In one experiment, polygons, B-splines and biased polygon approximation are used for shape coding while SA-DCT of MPEG-4 codec is used for texture coding. The fixed width (D_{max}) of the tolerance band is varied from 0.8 to 3.0 in steps of 0.1. In case of variable width of the tolerance band, the threshold (Th) is varied from 100 to 600 is steps of 50, $T_{min} = 0.8$ and $T_{max} = 3.0$. In case of the biased polygon approximation the $bias$ is varied from 0.1 to 0.9 in steps of 0.1. QP is varied from 2 (fine quantization) to 31 (coarse quantization) in steps of 1. The biased polygon approximation is chosen as the optimal solution for the reported results as seen in Table 1. The third column in the Tables show the threshold (Th) or the tolerance band width (D_{max}) for the variable and the fixed width tolerance band cases respectively.

R_{total} (bits)	Case	Th/ D_{max}	Bias	$PSNR^{YUV}_{texture}$ (dB)
32604	6	600	0.5	44.34
25588	6	600	0.5	40.73
21356	6	600	0.5	38.07
16012	6	600	0.5	34.40
14378	5	2.5	0.5	33.11
10914	5	2.5	0.5	29.93
4362	5	2.5	0.5	22.31
4304	5	3.0	0.5	22.16

Table 1: Optimal results of experiment one using frame 0 of the "Children" sequence.

In another experiment, the six shape coding techniques are used along with SA-DWT based texture coding of MPEG-4. The encoded bit stream is decoded at different rates. The reported results (Table 2) choose the biased polygon approximation for shape coding.

R_{total} (bits)	Case	Th/ D_{max}	Bias	$PSNR^{YUV}_{texture}$ (dB)
27824	6	600	0.5	48.96
21848	6	600	0.5	43.93
14904	5	3.0	0.5	35.49
9920	5	3.0	0.5	29.79
4952	5	3.0	0.7	23.14
4896	6	600	0.5	23.05

Table 2: Optimal results of experiment two using frame 0 of the "Children" sequence.

Fig. 1 shows a rate distortion comparison between SA-DCT and SA-DWT for frame 0 of the "Children" sequence. In Fig.

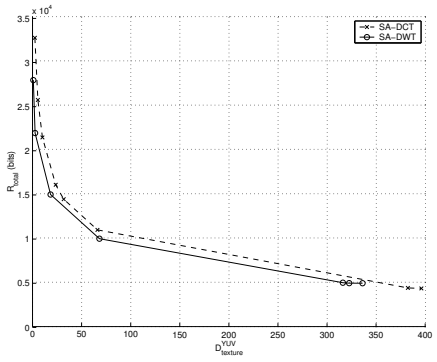

Fig. 1: Comparison of joint shape and texture coding with SA-DCT or SA-DWT based texture coding of the MPEG-4 codec. The results are shown for frame 0 of the "Children" sequence.

2, a subjective comparison for the same frame is shown. Fig. 3 shows a rate distortion comparison for frame 80 of the "Children" sequence. It is clearly observed that the joint shape and texture coding using SA-DWT based texture coding outperforms that using SA-DCT. The experimental results in Tables 1 and 2 choose the biased polygon approximation as the optimal solution among the six different shape coding techniques for a specific texture coding technique. The B-splines have a natural appearance and hence are more efficient in shape coding. The boundary encoding using the biased polygon approximation will favor the horizontal and vertical edges. Hence, it is the least efficient amongst the considered cases for shape coding. However, the spatial correlation between neighboring pels is better maintained while using the biased polygon approximation. As expected, shape coding using biased polygon approximation leads to better performance in texture encoding than both polygons and B-splines. Hence, the optimal solution chooses shape coding using biased polygon approximation to be the most efficient among the considered cases for joint encoding.

6. CONCLUSIONS

We presented an operational rate-distortion optimal bit allocation scheme between texture and shape for the encoding of the object based video. Experimental results show that the SA-DWT based texture coding performs better than SA-DCT based texture coding in the joint encoding of shape and texture. The biased polygon approximation is the least efficient amongst the considered cases for shape coding. However, we used it for efficient texture coding. As expected the optimal solution selected the techniques that uses biased polygon approximation for shape coding. The solution is determined using the Lagrangian multiplier method.

7. REFERENCES

[1] L. P. Kondi, G. Melnikov, A. K. Katsaggelos, "Joint optimal object shape estimation and encoding", *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 14, No. 4, Apr. 2004, pp. 528-533.

[2] 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, Jan. 1998.

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

[4] 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, Sept. 2003.

[5] 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, Jun. 1998.

[6] 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, Jan. 1998.

[7] T. Sikora and B. Makai, "Shape-adaptive DCT for generic coding of video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, pp. 59-62, Feb. 1995.

[8] S. Li, W. Li, "Shape adaptive discrete wavelet transforms for arbitrarily shaped visual object coding", *IEEE Transaction on Circuits and System for Video Technology*, Aug. 2000

(a) (b)

Fig. 2: Frame no. 80 ("Children" sequence),(a) SA-DCT based texture coding, $Th = 600$, $Bias = 0.5$, $QP = 4$, $R_{total} = 21356$ bits, $D^{YUV}_{texture} = 10.14$, (b) SA-DWT based texture coding, $Th = 600$, $Bias = 0.5$, Decoding rate = 21848 bits, $D^{YUV}_{texture} = 2.63$.

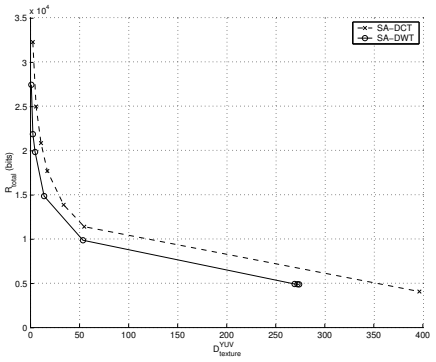

Fig. 3: Comparison of joint shape and texture coding with SA-DCT or SA-DWT based texture coding of the MPEG-4 codec. The results are shown for frame 80 of the "Children" sequence.