JOINT OPTIMAL 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 texture and shape information in object based video. The solution is optimal in the operational rate distortion sense i.e. given one coding setup the solution will guarantee the smallest possible distortion for a given rate. The shape is approximated using polygons or higher order curves. The texture is encoded using Shape Adaptive Discrete Cosine Transform (SA-DCT) and Shape Adaptive Set Partitioning In Hierarchical Trees (SA-SPIHT) and a comparison drawn between the two techniques. Both a fixed-width and a variable width tolerance band for shape coding were 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.

Keywords: *Shape coding, SA-SPIHT, SA-DCT*

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, 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], [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. The width of the tolerance band is inversely proportional to the magnitude of the image gradient. In [3], a joint shape and texture rate control algorithm for MPEG-4 encoders is proposed. In [4], 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 or SA-SPIHT 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. In Section 2 we discuss the contour based shape coding and SA-DCT and SA-SPIHT based texture coding techniques. 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. DESIGNING THE ENCODER

In this section the shape and texture coding techniques used are reviewed. Further details can be found in [5], [6].

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. 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].

2.1.1. Tolerance Band

A fixed width tolerance band has a width $2 \times D_{max}$ 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 set of boundary points. In order to construct the tolerance band we draw circles from each boundary point b_i with width $D_{max}[i]$. The tolerance band consists of the set of points inside the circles. Considering B-spline curve the segment distortion measure is given by:

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

where Q_k is defined by three control points p_{k-1}, p_k, p_{k+1} in case of a second order B-spline curve segment.

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:

$$D_{max}[i] = T_{min} + \lambda(grad[i] - gradmax), \qquad (1)$$

where

$$\lambda = \frac{T_{max} - T_{min}}{gradmin - gradmax}. \qquad (2)$$

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

2.1.2. Contours for Shape Coding

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 polygon used to approximate B, which is also an ordered set with p_k the kth vertex of P, N_p the total number of vertices in P and the kth edge starts at p_{k-1} and ends at p_k. A polygon edge is defined by two control points, its vertices.

In this paper, we also approximate boundaries with B-splines with a method that is based on the shortest path algorithm for a weighted directed acyclic graph (DAG). 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. A B-spline curve segment is defined by three control points.

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 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.

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) and Shape Adaptive Set Partitioning in Hierarchical Trees (SA-SPIHT).

While using SA-DCT [7] 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. 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.

The SA-DWT [8] transforms the samples in the arbitrarily shaped region into the same number of coefficients in the sub-band domain while keeping the spatial correlation, locality and self-similarity across sub-bands. We used the lifting in a 9-7 filter tap Cohen Daubechies filter for better PSNR of the joint encoding. The resultant coefficients are entropy coded using Shape Adaptive Set Partitioning in Hierarchical Trees (SA-SPIHT) [9]. Shape adaptivity is rendered to the SPIHT coding algorithm with a modification to handle partial wavelet trees that have wavelet coefficients corresponding to the pixels out of the shape boundary. The generated bit stream is fully embedded, allowing the best reduction of distortion with every additional bit sent. It can be truncated at any point to achieve the best reconstruction for the actual number of bits sent.

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:

$$\min R_{total} \text{ subject to } D^Y_{texture} \leq D^Y_{budget} \qquad (3)$$

where $D^Y_{texture}$ is the texture distortion of the image in the region of the reconstructed shape. The D^Y_{budget} is the maximum allowable texture distortion. The shape is encoded using curve segments as discussed and the reconstructed shape is made to lie within the tolerance band. The total rate for joint shape and texture coding is given by

$$R_{total} = R_{shape} + R_{texture} \quad (4)$$

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

$$\min D^{Y}_{texture} \text{ subject to } R_{total} \leq R_{budget} \quad (5)$$

where R_{budget} is the total rate for the shape and texture and $D^{Y}_{texture}$ is the distortion for the texture. The R_{shape}, $R_{texture}$ and $D^{Y}_{texture}$ are computed from the codec.

We solve the optimization problem for six cases of shape coding for a specific texture coding technique. The shape of the object is encoded using polygons, B-splines and biased polygon approximation. Both the fixed width and the variable width of the tolerance band are considered for each of these three results giving a total of six different shape coding techniques. Obviously, the best set of coding parameters will yield the optimal R-D curves. In case of fixed width of the tolerance 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 tolerance band. While using biased polygon approximation, 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-SPIHT based texture coding the embedded bit stream is decoded at different rates.

3.1. Modeling the distortion

The distortion in all of our simulation results is calculated based on the mean square error (MSE) of the original and the reconstructed image in the region of the reconstructed shape. The distortion can be mathematically written as:

$$D^{Y}_{texture} = \frac{1}{N} \sum_{(x,y) \in C} \delta(x,y)^2 \quad (6)$$

where $\delta(x,y)$ is the differential intensity value at pixel position (x,y), N is the number of pixels in the region of the reconstructed shape. The peak signal-to-noise ratio (PSNR) can be calculated using:

$$PSNR^{Y}_{texture} = 10 \log_{10}\left(\frac{255^2}{D^{Y}_{texture}}\right) \quad (7)$$

4. OPTIMAL SOLUTION

Thus Eq.(3) can be written as

$$V^{*} = \arg\min \quad R_{total}(V), \quad (8)$$

subject to

$$D^{Y}_{texture}(V) \leq D^{Y}_{budget}$$

where, V specifies one of the six possible shape coding techniques along with the parameters associated with it and $R_{total}(V) = R_{shape}(V) + R_{texture}(V)$. The constrained minimization problem stated in Eq.(4) is converted to an unconstrained problem by using the Lagrangian multiplier method as $J_{\lambda}(V) = R_{total}(V) + \lambda \cdot D^{Y}_{texture}(V)$ where λ is the Lagrangian multiplier. The unconstrained problem now becomes the minimization of the cost function $J_{\lambda}(V)$ and can be solved using the bisection algorithm.

5. EXPERIMENTAL RESULTS

In one experiment the polygons, B-splines and the biased polygon approximation are used for shape coding while SA-DCT is used for texture coding. The fixed width 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 $threshold$ is varied from 100 to 600 is steps of 50, while $T_{min} = 0.8$ and $T_{max} = 3.0$. In the case of the biased polygon approximation the $bias$ is varied from 0.1 to 0.9 in steps of 0.1. The QP is varied from 2 (Fine quantization) to 31 (Coarse quantization) in steps of 1. The biased polygon approximation with the fixed width of the tolerance band is chosen as the optimal solution for all the texture distortions as seen in Table 1.

Width	Bias	QP	R_{total} (bits)	$D^{Y}_{texture}$	$PSNR^{Y}_{texture}$ (dB)
3.0	0.1	2	30162	6.19	40.22
2.5	0.5	3	23244	12.65	37.11
2.5	0.5	4	19209	20.26	35.06
2.5	0.5	7	12971	48.13	31.13

Table 1: Results of experiment one

In another experiment, the six shape coding techniques are used along with SA-SPIHT based texture coding. The encoded bit stream is decoded at different rates. The reported results (Table 2) choose the biased polygon approximation with the fixed width of the tolerance band. From

Width	Bias	R_{total} (bits)	$D^{Y}_{texture}$	$PSNR^{Y}_{texture}$ (dB)
3.0	0.9	30232	1.63	46.00
3.0	0.9	25232	3.58	42.59
3.0	0.9	20232	8.13	39.03
3.0	0.7	15234	18.61	35.43

Table 2: Result of experiment two

Fig. 1, it is clearly observed that the joint shape and texture coding using SA-SPIHT 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 main reason for using B-splines is their natural appearance over polygons. The boundary encoding using 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 neighbouring pixels 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 most efficient among the considered cases for joint encoding of the shape and texture.

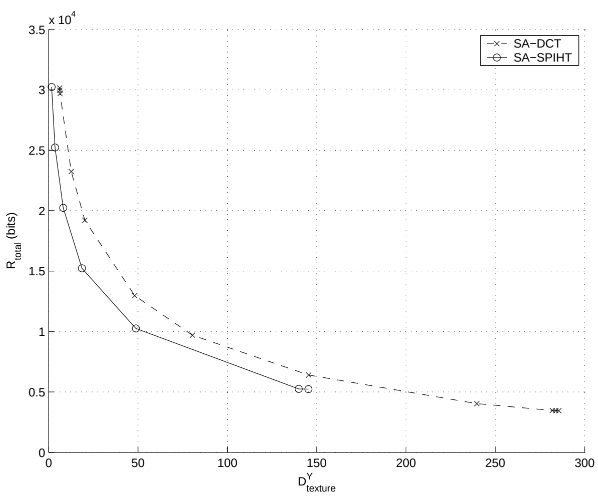

Fig. 1: Comparison of the joint shape and texture coding with SA-DCT or SA-SPIHT based texture coding using Frame 0 of the "Bream Sequence".

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 SA-SPIHT based texture coding performs better than SA-DCT based texture coding in the joint encoding of shape and texture. 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. 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. The optimal approach has higher efficiency than an exhaustive search algorithm.

7. 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, Oct. 2001.

[2] 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.

[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] 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.

[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

[9] G. Minami, Z. Xiong, A. Wang, and S. Mehrota, "3-D wavelet coding of video with arbitrary regions of support," *IEEE Trans. Cir. Sys. Video Tech.*, vol. 11, no. 9, pp. 1063-1068, Sept. 2001.

[10] G. Schuster and A. Katsaggelos, *Rate-Distortion Based Video Compression.* Boston: Kluwer Academic Publishers, 1997.