New Dynamic Enhancements to the Vertex-Based Rate-Distortion Optimal Shape Coding Framework

Journal Item

New Dynamic Enhancements to the Vertex-Based Rate-Distortion Optimal Shape Coding Framework

Ferdous Ahmed Sohel, *Student Member, IEEE*, Laurence S. Dooley, *Senior Member, IEEE*, and Gour C. Karmakar, *Member, IEEE*

***Abstract*—Existing vertex-based operational rate-distortion (ORD) optimal shape coding algorithms use a vertex band around the shape boundary as the source of candidate control points (CP) usually in combination with a tolerance band (TB) and sliding window (SW) arrangement, as their distortion measuring technique. These algorithms however, employ a fixed vertex-band width irrespective of the shape and admissible distortion (AD), so the full bit-rate reduction potential is not fulfilled. Moreover, despite the causal impact of the SW-length upon both the bit-rate and computational-speed, there is no formal mechanism for determining the most suitable SW-length. This paper introduces the concept of a variable width admissible CP band and new adaptive SW-length selection strategy to address these issues. The presented quantitative and qualitative results analysis endorses the superior performance achieved by integrating these enhancements into the existing vertex-based ORD optimal algorithms.**

***Index Terms*—Vertex-based shape coding, video coding.**

I. Introduction

SHAPE CODING has become a popular topic in multimedia technology research as evidenced by the corpus of literature [1]–[9]. Shape coders are generally classified as being either bit-map or contour based [1], with examples of the former including the context based arithmetic encoder (CAE) [5], and new digital straight line segments based context coding (DSLSC) [6], while the baseline [7] and vertex-based polynomial approaches [1] are two popular examples of contour-based shape coding.

A concise treatise of the shape coding algorithms has been presented in [1] with both polygon and B-spline based operational rate-distortion (ORD) frameworks being developed which form the basis for a suite of algorithms [1]–[4] that all have as their aim, for some prescribed distortion, to optimally encode a shape contour in terms of the number of bits, by selecting the set of control points (CPs) that incurs the lowest bit rate and vice versa. The CP are selected from a set of vertices contained in the admissible control point band (ACB) around the shape contour. Algorithms [1], [2] employ a single admissible distortion (AD) for all boundary points, while those in [3], [4] embrace variable AD for each boundary point, though all use a fixed-width admissible CP band (FCB) with the width being arbitrarily selected. This does not guarantee an optimal source of admissible CP unless the width is sufficiently large, and large FCB widths imply increased computational complexity. This paper addresses this issue by proposing a variable-width admissible control point band (VCB), where the width of each boundary point is individually determined from the admissible peak distortion and shape curvature information.

The sliding widow (SW) is also a core element of the ORD algorithms, as its length directly impacts upon the bit-rate. It was originally introduced [1] to force the encoder to follow the shape boundary so avoiding trivial solutions, as well as improving computational speed. In fact, the SW confines the search space for the next CP to only those points within the SW-length, so compromising global optimality in a bit-rate sense. For computationally efficient encoding, a small SW-length is desirable though decreasing the length increases the overall bit-rate overhead. While the SW is a deep-rooted idea, no formal mechanism has been developed for determining the most appropriate SW-length, so this paper also presents a shape adaptive sliding window (ASW) strategy based on the curvature of boundary points.

These enhancements can all be seamlessly embedded into the vertex-based ORD optimal shape coding framework, with the performance of these ORD algorithms integrating the enhancements being analyzed, and consistently shown to provide superior RD results in terms of both lower bit-rate and faster encoding speed compared with [1]–[4].

The rest of this paper is organized as follows. Section II provides a brief overview of existing vertex-based ORD optimal shape coding algorithms and identifies some limitations, while Section III details the underlying theoretical principles of the proposed enhancements. Section IV presents an analysis of the empirical results and performance comparison for the original ORD algorithms with the new enhancements embedded, while some concluding remarks are provided in Section V.

II. Existing Vertex-Based Rate-Distortion Optimal Shape Coding Framework

The algorithms [1]–[4] seek to determine and encode a set of CP to represent a particular shape within prescribed RD constraints. Assume boundary $B = \{b_0, b_1, \ldots, b_{N_B-1}\}$ is an ordered set of shape points, where N_B is the total number of points and $b_0 = b_{N_B-1}$ for a closed boundary. $P = \{p_0, p_1, \ldots, p_{N_P-1}\}$ is an ordered set of CP used to approximate B, where N_P is the total number of CP and $P \subseteq C$, where C is the ordered set of vertices in ACB. In the core of these algorithms a weighted directed acyclic graph (DAG) is formed using the vertices in and the shortest path from the first vertex to the last vertex is searched. These algorithms however, possess some inherent limitations.

Manuscript received March 05, 2006; revised March 10, 2007. Portions of this paper were presented at the IEEE International Conference on Image Processing (ICIP'06), Atlanta, GA, Oct. 8–11, 2006. This paper was recommended by Associate Editor S. Li.

The authors are with the Gippsland School of Information Technology (GSIT), Monash University, Churchill, Vic. 3842, Australia (e-mail: lsdaussie@ieee.org; Gour.Karmakar@infotech.monash.edu.au; Ferdous.Sohel @infotech.monash.edu.au).

Digital Object Identifier 10.1109/TCSVT.2007.903788

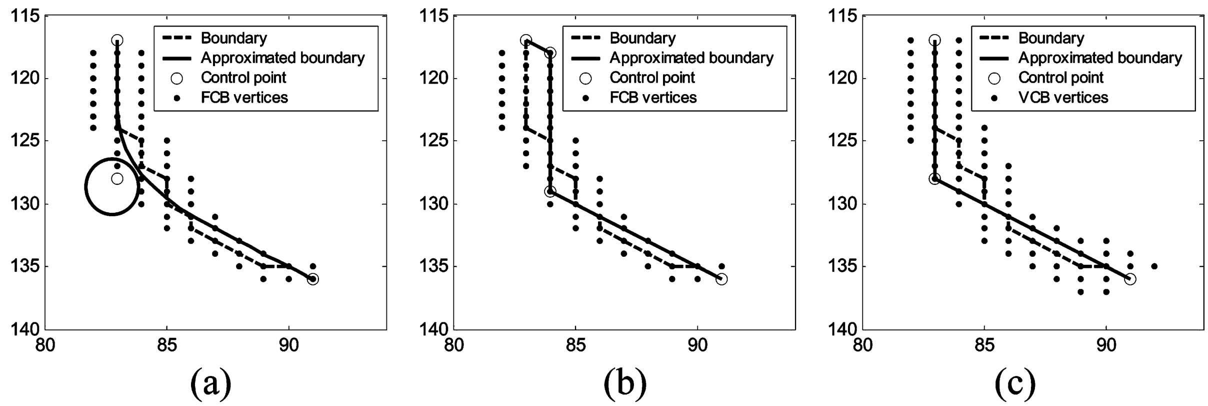

Fig. 1. B-spline based approximation: (a) shape maintains the AD though one CP (encircled) lies outside the FCB. Polygon approximations: (b) FCB requires 4 CP and (c) VCB requires 3 CP.

A. *Fixed Width Admissible CP Band*

In [1], CP are selected from a set of vertices around the boundary namely the ACB. In [1], [2], a single admissible peak distortion $(D_{\max})$ is considered the FCB width $(W_{\max})$ for the entire shape, which is optimal for polygonal approximations, but may not be so for B-spline based approximations as a CP lying outside this band can still produce a shape-approximating curve that upholds the peak AD. The *Neck* segment of the 31st frame of the *Miss America* video sequence in Fig. 1(a) illustrates this scenario, where one CP (encircled) is located outside the band, yet the resulting approximation curve still maintains the peak AD of $D_{\max} = 1$ pel. For efficient coding, the ORD algorithms [3], [4] employ two admissible peak distortion bounds ($T_{\max}$ and $T_{\min}$) to obtain the AD $T[i]$ at boundary point b_i, by permitting a smaller distortion at boundary points having a lower image intensity gradient or shape curvature [4], and vice versa. A FCB is again employed however, with $W_{\max} = 1$ pel for example in [3], being invariant of the AD value so the philosophy of variable AD is not fully exploited in bit-rate minimization, with FCB accordingly no longer being optimal for either polygonal or curve-based approximations. The FCB example in Fig. 1(b) corroborates this with the encoder mandating 29 bits (4 CP) for a polygonal shape approximation, while the corresponding VCB solution needs only 24 bits (3 CP) in Fig. 1(c) for $T_{\max} = 2$ and $T_{\min} = 1$ pel.

B. *Impact of a Sliding Window*

The overall computational complexity of polygon-based ORD algorithms is $O\left(N_B^4\right)$, since the DAG mandates $O\left(N_B^2\right)$ and the distortion measurement, which is an integral DAG loop calculation, also incurs $O\left(N_B^2\right)$ time. If an SW of length L is applied, where $L < N$, then the complexity drops to $O\left(N_B L^3\right)$, so the overall computational overhead is always lower when a SW is used, because it limits the CP search space. This conversely increases the bit-rate requirement, with the nexus between bit-rate and computational speed being illustrated in Fig. 2, where typical rates and CPU-times are plotted for various SW-lengths on the same shape and prescribed AD.

A smaller SW clearly incurs a lower time penalty but a higher bit-rate and vice versa, so the significance in selecting the most appropriate SW-length is a key design parameter for achieving better admissible bit-rate utilization and reduced computational costs.

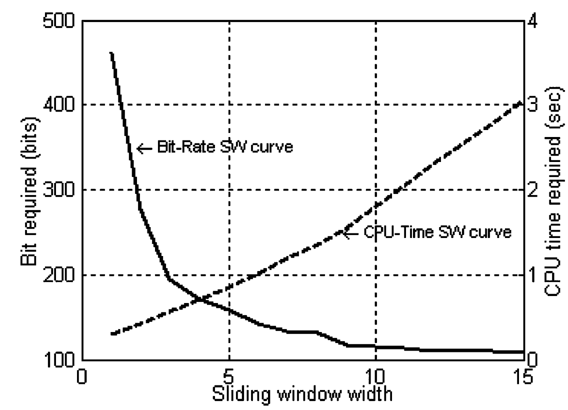

Fig. 2. Effect of SW-length in pel on bit-rate and CPU-time in sec.

III. Proposed Enhancements to ORD Optimal Shape Coding Framework

This section firstly elucidates the theoretical basis of the VCB, before formally developing bounds for the width of this band for every boundary point. A brief discussion upon the proposed enhancement to the distortion measurement technique is then provided, followed by SW-length selection strategy.

A. *Variable Width Admissible CP Band*

VCB is an ordered set of vertices formed around the shape, with every boundary point having some associated vertices in the band. As the AD for each boundary point differs, they will have different VCB widths and so the number of VCB points associated with each point will vary. Let $W[j]$ be the width of VCB for the boundary point b_j.

Bounds for the width of the admissible control point band: Lemmas 1 and 2 focus upon the ACB width for a polygonal and quadratic B-spline based coding, respectively.

Lemma 1: For a polygonal approximation, the width of the ACB for any boundary point is bounded by the peak AD for that point, i.e., $W[j] \leq T[j]$ for b_j.

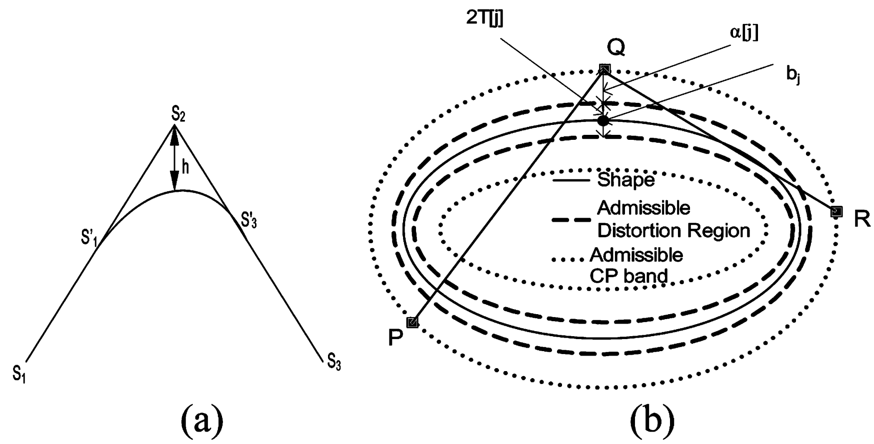

Fig. 3. (a) Distance between a quadratic B-spline curve and its CP and (b) Maximal width of the admissible CP band calculation.

Proof (by Contradiction): Let there be such a vertex u associated with boundary point b_j where $u \in C$ and the distance of u from b_j is greater than $T[j]$, i.e., $W[j] > T[j]$. Assume u is now selected as a CP. Since it is a polygonal approximation, the approximated shape will pass through this vertex (in fact, it will be an end point of one edge), so the distortion at this vertex will exceed the peak AD for b_j, and for b_j such a vertex u can never be selected as a CP and lie within the VCB vertices associated with b_j.

Lemma 2: For a quadratic B-spline based RD constrained shape approximation,

$$\alpha[j] \leq \min\left\{(3\delta + 4T_{\max} + 2T[j]/6), \left(\rho\sqrt{2}/4\right)\right\},$$

where δ and ρ are respectively the longest chord length of the boundary and the largest run-length possible for the code employed, $\alpha[j]$ is the difference between the corresponding AD and width of the admissible CP band, i.e., $W[j] = \alpha[j] + T[j]$.

Proof: Fig. 3(a) shows a uniform quadratic B-spline curve produced by CP S_1, S_2, and S_3. This is actually a Bezier curve (BC) generated by S_1', S_2 & S_3', where $S_1' = (1/2)(S_1 + S_2)$ and $S_3' = (1/2)(S_2 + S_3)$, with h being the minimum distance of the middle CP S_2 from the BC. It thus follows from [10] that $2h \leq \max\{|S_1'S_2|, |S_2S_3'|\}$, where $|S_2S_3'|$ is the length of edge S_2S_3', so $4h \leq \max\{|S_1S_2|, |S_2S_3|\}$.

In Fig. 3(b) example, three CP P, Q, and R are employed to encode a shape segment that includes the boundary point b_j which has an AD $T[j]$. Assuming $PQ \geq QR$, the distance of the B-spline curve from Q is always $\leq (1/4)|PQ|$. The maximum length of PQ is $\delta + T_{\max} + T_{\max} + \alpha_{\max} + \alpha_{\max} = \delta + 2T_{\max} + 2\alpha_{\max}$ where $\alpha_{\max}$ is the maximum value of α. So $\delta + 2T_{\max} + 2\alpha_{\max} \geq 4\alpha_{\max}$. Hence, $\alpha_{\max} \leq (\delta/2) + T_{\max}$. Now the corresponding $\alpha[j]$ for boundary point b_j is given by

$$4\alpha[j] \leq \delta + T_{\max} + \alpha_{\max} + T[j] + \alpha[j] \tag{1}$$

$$\text{so, } \alpha[j] \leq \tfrac{1}{6}\left(3\delta + 4T_{\max} + 2T[j]\right). \tag{2}$$

The encoding strategy adopted can limit the length of an edge since if ρ is the length of codebook, it is able to encode a maximum length of $\rho\sqrt{2}$ pel (through the diagonal) so

$$\alpha[j] \leq \frac{\rho\sqrt{2}}{4}. \tag{3}$$

From (2) and (3), $\alpha[j] \leq \min\{(3\delta + 4T_{\max}+2T[j]/6), \left(\rho\sqrt{2}/4\right)\}$.

B. Sliding Window Length Selection Strategies

As discussed earlier, a SW improves computational complexity while concomitantly compromising global optimality in a bit-rate sense. Thus, rather than using an arbitrary SW-length, there are cogent arguments for applying an appropriate SW-length within the ORD coding framework. This section presents an ASW-length strategy that adjusts the SW-length automatically based on a shape's cornerity. As the cornerity of points on a shape do not change abruptly, but rather gradually follow a trend to and from the local maximum value [11], its value can be used to determine the most appropriate SW-length. By monitoring the relative cornerity of shape points, an adaptive algorithm has been developed so boundary points with a higher cornerity induce a smaller SW and vice versa. This ensures more CP for the boundary-segment with sharp changes and corners and fewer CP for segments where the rate of change in shape is more gradual (longer window lengths). Let $K[j]$ be the cornerity of the jth boundary point, and $K_{\max}$ and $K_{\min}$ respectively be the maximum and minimum cornerity values of the vertices on a boundary. The SW-length for the jth boundary point is then given by(4), shown at the bottom of the page.

The ASW strategy is effective for both cursive and non-cursive shapes, while for the special case of a straight line, when $K_{\max} = K_{\min}$ the SW-length is always $L_{\max}$, since both $K_{\max}$ and $K_{\min}$ will be zero. The encoder may be constrained by either bit-rate or AD, with in both cases, for logarithmic

$$L[j] = \begin{cases} L_{\max}, & \text{if } K_{\max} = K_{\min} \\ L_{\min} + \frac{L_{\max} - L_{\min}}{e^{K_{\max}} - e^{K_{\min}}} \cdot \left(e^{K_{\max}} - e^{K[j]}\right), & \text{else.} \end{cases} \tag{4}$$

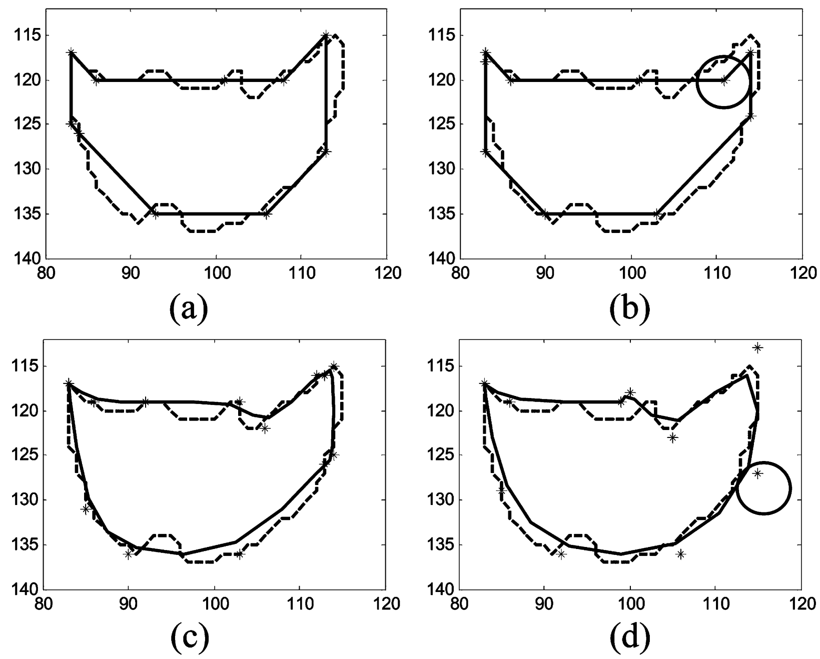

Fig. 4. Results for the *Neck* region with $T_{\max} = 3$ and $T_{\min} = 1$ pel (Legends—solid line: approximated boundary; dashed line: original boundary; asterisk: CP). (a) Polygon-FCB, (b) polygon-VCB, (c) B-spline-FCB, (d) B-spline-VCB.

coding (LC) [2], $L_{\max}$ being 15 pel. If the bit-rate is constrained, $L_{\min} = 1$ while for an encoder constrained by AD, $L_{\min} = \lceil T_{\max} \rceil$ as otherwise, the AD is not fully exploited.

The ASW technique is computationally very efficient requiring order of complexity $O(N_B)$ to both calculate the cornerity using [11] and determine the SW-length for all boundary points.

IV. Results and Analysis

To both qualitatively and quantitatively analyze the performance of the proposed enhancements compared with the existing ORD optimal shape coding algorithms, experiments were performed upon various test video sequences: *Kids*.sdtv (spatial resolution 720×486 pixels, 300 frames), *Stefan*.sdtv (720×480, 300), *Kids*.sif (352×240, 100), *Weather*.qcif (176×144, 100), *Foreman*.cif (352×240, 207) and *MissAmerica*.qcif (176×144, 100). For presentational clarity, this section will concentrate on analyzing the results for the *Neck* region of the 31st frame of the *MissAmerica*.qcif sequence, with the results for the other sequences being summarised in tabular form. To also clarify the nomenclature adopted in this section, the following four parameter notation is used:

$$\text{Approx.type} - \text{ACB type} - [\text{SW}]$$

Approx. type refers to either polygonal or quadratic B-spline based approximation; *ACB type* indicates either VCB or FCB is deployed; optional part SW indicates the SW types used with its default length being 15 pel, Polygon-VCB means the algorithm is based upon a polygonal approximation, with VCB as the admissible CP band and a SW-length of 15 pel. ASW indicates that the new adaptive sliding window strategy, rather than a fixed-length SW has been applied together with the Beus-Tiu cornerity calculation method [11]. In addition, the accurate distortion measurement technique in [9] has been applied in lieu of the shortest absolute distance (SAD), distortion band (DB) or tolerance band (TB) metrics, since it has been proven in [9] that in certain circumstances, both SAD and DB have limitations in their measurements and as the TB is in fact a generalization of the DB [3], it too inherits the same restrictions as the DB.

TABLE I
Maximum Distortion $T_{\max}$ (pel) Required for the *Neck* Region for Different Algorithms for Various Admissible Bit-Rates With $T_{\min} = 1$ pel

Algorithms	**75 *bits***	**90 *bits***	**105 *bits***
Polygon–FCB	3.61	2.45	2.24
Polygon–VCB	3.17	2.24	2.0
B-spline–FCB	3.17	2.24	2.0
B-spline–VCB	2.0	2.0	1.42

The first series of experiments performed upon the *Neck* region focused upon the peak distortion for a prescribed set of admissible bit-rate values with the respective numerical results produced by different algorithms being summarized in Table I. These confirm VCB-based algorithms produce lower distortions than their FCB counterparts, so for example with an admissible bit-rate of 75 bits, B-spline-FCB and B-spline-VCB respectively produced $T_{\max}$ values of 3.17 and 2.0 pel as a direct consequence of VCB providing a larger and more dynamic set of potential admissible CP.

The next series of experiments concentrated upon the bit-rate and distortion for a prescribed set of AD settings. The respective results produced by the various ORD algorithmic combinations are shown in Fig. 4(a)–(d) for a peak distortion bound of $T_{\max} = 3$ pel, $T_{\min} = 1$ pel with the corresponding numerical results given in Table II. These reveal that the *Polygon-FCB* and *Polygon-VCB* approximations with $T_{\max} = 2, T_{\min} = 1$ pel required 109 and 99 bits, respectively, with a similar trend being

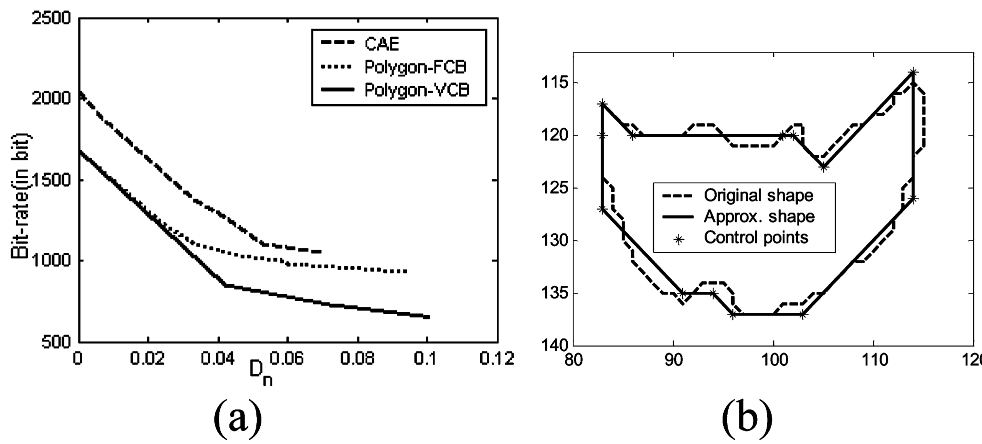

Fig. 5. (a) Comparative RD performances for various polygon-based algorithms using the D_n metric upon the *Kids*.sif sequence. (b) Shape coding for the *Neck* region using *Polygon-VCB-ASW* for $T_{\max} = 2, T_{\min} = 1$ pel.

TABLE II
BIT REQUIREMENTS (IN BIT) FOR ADMISSIBLE $T_{\max}$ AND $T_{\min}$ FOR THE VARIOUS ORD OPTIMAL SHAPE CODING ALGORITHMS UPON THE NECK REGION

AD→	$T_{\max}=2$, $T_{\min}=1 pel$	$T_{\max}=3$, $T_{\min}=1 pel$	$T_{\max}=3$, $T_{\min}=2 pel$
Algorithms ↓	**Bit-rate**	**Bit-rate**	**Bit-rate**
Polygon–FCB	109	83	82
Polygon–VCB	99	80	76
B-spline–FCB	88	78	68
B-spline–VCB	65	61	62

TABLE III
LOSSLESS COMPRESSION RESULTS (AVERAGE NUMBER OF BITS PER FRAME) GENERATED BY CONTEMPORARY SHAPE CODING ALGORITHMS APPLIED TO VARIOUS TEST VIDEO SEQUENCES

Test Sequence	*Polygon-VCB*	DSLSC	CAE (Intra)	CAE (Inter)
Kids.sif	1704	1792	2040	1736
Weather.qcif	472	480	520	320
Foreman.cif	1040	904	1472	1344
Kids.sdtv	5400	5392	7248	5848
Stefan.sdtv	1200	1152	1808	1680

observed in respect of the bit requirements for all VCB based algorithms, namely a lower value than for the FCB-based algorithms due to the richer and more dynamic set of potential admissible CP. The various encircled CP in Fig. 4(b) and (d), lie outside the $W_{\max} = 1$ pel FCB yet approximate a significant portion of the shape, so lowering the bit-rate in the VCB case.

To substantiate the improved performance of the VCB-based strategies, additional experiments were conducted using the MPEG-4 shape distortion metric D_n which is defined as the ratio of the number of erroneously represented pels of an approximating shape to the total number of pels in the original shape. Again the various enhancements presented in this paper provided better RD performance compared to MPEG-4 CAE [5] technique, when D_n is calculated for the first 100 frames of the *Kids*.sif sequence which contains multiple objects in a frame. The corresponding RD results for various polygon-based algorithms are plotted in Fig. 5(a) and clearly reveal the superior performance achieved by the VCB enhancements over both the FCB based framework and the CAE, especially at higher distortion values.

In addition to analyzing the lossy vertex-based ORD optimal shape coding framework, comparative results for a number of standard and contemporary lossless shape coders are presented in Table III. Both *Polygon-VCB* and DSLSC [6] in intra mode provided comparable bit-rate results for lossless compression and are evidently much more efficient than CAE encoders in terms of intra- and motion-compensated inter-modes for most shapes. It needs to be emphasized that while the framework in [1]–[4] is based on a tradeoff mechanism which is optimal in the ORD sense, the enhancements presented in this paper are specifically designed to secure greater efficiency for lossy compression, while DSLSC and CAE [5] are both lossless entropy coders. Obviously, the VCB-based paradigm also affords lossless compression whenever $T_{\max} = T_{\min} = 0$.

Fig. 5(b) demonstrates the performance of the ASW-length strategy, by displaying the decoded shape using *Polygon-VCB* for $T_{\max} = 2, T_{\min} = 1$ pel, where it is palpable the approximated shape has retained all its sharp corners by using more CP in these regions, while concomitantly using fewer CP in the flatter regions so vindicating using shape cornerity at the boundary points in the adaptation process.

V. CONCLUSION

This paper has presented a series of innovative enhancements to existing vertex-based ORD optimal shape coding algorithms. It has proposed a dynamic and flexible variable-width admissible control point band (VCB) instead of the conventional fixed-width band as the source of potential control points to fully utilize the variable admissible distortion (AD) in minimizing the bit-rate. For both polygonal and B-spline based approximations, theoretical bounds on the VCB-width have been established. The paper has also investigated automatically determining a shape adaptive length of a sliding widow suitable for RD constraints, with empirical results confirming the improvements in RD performance and distortion measurement accuracy over the existing ORD optimal shape coding algorithms.

REFERENCES

[1] A. K. Katsaggelos, L. P. Kondi, F. W. Meier, J. Ostermann, and G. M. Schuster, "MPEG-4 and rate-distortion-based shape-coding techniques," *Proc. IEEE*, vol. 86, no. 6, pp. 1126–1154, Jun. 1998.

[2] G. M. Schuster and A. K. Katsaggelos, *Rate-distortion Based Video Compression: Optimal Video Frame Compression and Object Boundary Encoding*. Norwell, MA: Kluwer, 1997.

[3] L. P. Kondi, F. W. Meier, G. M. Schuster, and A. K. Katsaggelos, "Joint optimal object shape estimation and encoding," in *Proc. SPIE Conf. Vis. Comm. Image Processing*, 1998, pp. 14–25.

[4] L. P. Kondi, G. Melnikov, and A. K. Katsaggelos, "Jointly optimal coding of texture and shape," presented at the Int. Conf. Image Processing, Sophia-Antipolis, France, Sep. 3–5, 2001.

[5] N. Brady, F. Bossen, and N. Murphy, "Context-based arithmetic encoding of 2D shape sequences," in *Proc. Int. Conf. Image Processing*, 1997, vol. 2, pp. 29–32.

[6] S. M. Aghito and S. Forchhammer, "Context based coding of bi-level images enhanced by digital straight line analysis," *IEEE Trans. Image Processing*, vol. 15, no. 8, pp. 2120–2130, 2006.

[7] S. Lee, D. Cho, Y. Cho, S. Son, E. Jang, J. Shin, and Y. Seo, "Binary shape coding using baseline-based method," *IEEE Trans. Circuits Syst. Video Tech.*, vol. 9, no. 1, pp. 44–58, 1999.

[8] F. A. Sohel, L. S. Dooley, and G. C. Karmakar, "Variable width admissible control point band for vertex based operational-rate-distortion optimal shape coding algorithms," presented at the Int. Conf. Image Processing, Atlanta, GA, Oct. 8–11, 2006.

[9] F. A. Sohel, L. S. Dooley, and G. C. Karmakar, "Accurate distortion measurement for generic shape coding," *Pattern Recog. Lett.*, vol. 27, no. 2, pp. 133–142, 2006.

[10] D. Nairn, J. Peters, and D. Lutterkort, "Sharp, quantitative bounds on the distance between a polynomial piece and its Bezier control polygon," *Comput. Aided Geometric Design*, vol. 16, no. 7, pp. 613–631, 1999.

[11] H. L. Beus and S. H. Tiu, "An improved corner detection algorithm based on chain coded plane curves," *Pattern Recog.*, vol. 20, no. 3, pp. 291–296, 1987.