SCALABLE SHAPE CODING OF VIDEO OBJECTS USING TEMPORAL PREDICTION

Mingyou HU, A. H. Sadka, S. Worrall, A. M. Kondoz

Centre for Communication System Research (CCSR),
University of Surrey, Guildford,
Surrey, United Kingdom

ABSTRACT

Scalable shape encoding is an important requirement of highly scalable object-based video coding. In this paper, a new scalable vertex-based shape coding scheme is proposed that uses temporal prediction. During shape coding of a video object, the object shape in the first frame is scalable intra-coded using the method described in [1]. For scalable shape coding of subsequent frames, temporal prediction is conducted during the coding of coarser layers. Contour matching in the curvature scale space domain is conducted in order to get higher matching accuracy. Experimental results show that the proposed scalable shape coding scheme can achieve better R-D performance than existing predictive shape coding methods [2] and CAE in MPEG-4 [3].

1. INTRODUCTION

Much research has been conducted into determining methods for shape representation and coding. In order to transmit the shape of an object efficiently and robustly, a large number of techniques have been proposed [1-2, 4-10]. These coding algorithms can achieve lossy and/or lossless coding. However, the coding efficiency achieved by intra-shape coding can not satisfy the requirement of low bitrate video coding, even though current state-of-the-art compression ratios are high. Since a contour sequence has very high correlation in the temporal domain, motion estimation and compensation can be used to achieve further compression [2, 4-6]. The contour in the current frame can be predicted from the contour obtained in the previous frame. Only the contour segment which can not be predicted must be encoded, which is achieved by using the intra-shape coding technique. This can reduce the bit rate of shape coding drastically.

In the published contour motion estimation schemes [2] [4], the object contour is assumed to undergo a translational motion. When there is more complex motion such as zoom and/or rotation, the contour cannot be well compensated. In Figure 2 (a), for example, the object contours in two neighbouring frames are overlapped. We can find that different motion patterns exist for different contour segments. It is difficult to use one motion model to describe them.

In [6], an affine global motion compensation scheme is investigated. The parameter vector of affine transformation is estimated according to two available contours and is encoded. The problem with this method is the coding of affine parameters. As the affine parameters are floating point, 10-12 bits are required to represent each parameter. So, for most sequences, the compression ratio is not very high. Furthermore, it can not provide scalable shape coding.

In this paper, a novel scalable predictive coding scheme is proposed, which can achieve higher compression efficiency than state-of-the-art shape coding methods due to the use of temporal information. In order to achieve scalable predictive coding, it is necessary to represent and estimate the contour motion hierarchically. In our proposed scheme, the contour motions in level i are first estimated. They are predicted from the MVs of the previously transmitted levels and/or the encoded MVs of the current level. Contour matching in CSS images is applied to find the correspondence of two contours during contour motion estimation, which can achieve more accurate contour motion estimation. This paper is organised as follows. The proposed scalable shape coding scheme is described in detail in section 2. Section 3 presents some experimental results. Section 4 gives the conclusions.

2. SCHEME DESCRIPTION

Figure 1 shows the diagram for predictive scalable shape coding. The novelties of this method are twofold. First, we propose an efficient contour motion estimation scheme,

0-7803-9332-5/05/$20.00 ©2005 IEEE

which is based on the curvature information of an object contour and is used to predict the motion vectors of vertices in the coarser level. Second, a scalable shape coding scheme is proposed, in which the motion of each contour is estimated hierarchically. A multi-model encoding scheme is included to improve the compression efficiency. Our scheme consists of two steps: contour motion estimation and scalable predictive shape coding, which will be discussed in detail in the following sections.

2.1. Contour motion estimation

Instead of using corner matching for contour motion estimation in [6], in our method, curvature scale space (CSS) images and curvature information are used in contour matching. CSS image is currently used for shape feature selection in shape indexing and retrieving and bas been selected as shape descriptors for the MPEG-7 standard [11]. Compared with the method in [6], the proposed method can achieve more accurate motion estimation. The contour motion estimation algorithm includes CSS image calculation and contour matching in the CSS domain.

The CSS image is computed by first convolving a path-based parametric representation of the contour with a Gaussian function, as the standard deviation of the Gaussian varies from a small to a large value. Next, curvature is computed on each smoothed contour. As a result, curvature zero-crossing points can be recovered and mapped to the CSS image in which the horizontal axis represents the arc length parameter on the original contour, and the vertical axis represents the standard deviation of the Gaussian filter. The CSS image has the properties that it is invariant under rotation, uniform scaling and translation of the contour.

Figure 2 shows the two object contours of Motr_dhtr sequence, and their zero crossing of contour curvature image with the scale . It can be found that the zero crossing of object contour curvature can identify important geometric properties of contour and two object contours have large similarity in CSS images.

As the two contours have the same direction (counter-clockwise in our experiments), the matching of two CSS images just tries to find the optimal horizontal shift of the maxima in one of the CSS images that would yield the best possible overlap with the maxima of the other CSS image. The contour-matching algorithm [12] (in Chapter 2) has been employed to achieve contour matching in CSS image.

2.2. Scalable predictive shape coding

The proposed scalable predictive shape coding scheme used different encoding schemes for different approximation levels. For each level except the finest one, the motion vectors of allocated approximating vertices are estimated and the contour of the current frame is predicted by motion compensation. For the motion failure segments, where the approximation error band is larger than the predefined threshold, new vertices are inserted to make it satisfy the error band. The coordinates of these vertices are intracoded and are transmitted to the decoder. As an adaptive update scheme is proposed and used in the codec, the order of the maintained and rejected vertices are not coded and transmitted to the decoder. This is different from the method in [9].

The predictive scalable shape coding algorithm can be summarized as follows:

- For the vertices of level 0 in the previous frame, which have the maximal curvature, their corresponding point in the CSS of current frame is estimated based on the contour matching algorithm in Section 2.1. For other vertices of level 0, which try to make the contour satisfy the error band, their motion vectors are estimated by trying to minimise the approximation error. After estimating the motion vectors, these motion vectors are encoded by using a variable length coding scheme (VLC). In order to achieve higher coding efficiency, each motion vector is divided into a global motion part and a local motion part. For the motion failure segments, where approximation can not satisfy the predefined error band; new vertices are inserted based on the CSS image and maximal error distance. The coordinates of these vertices are intra-coded and transmitted to the decoder after the motion vectors.
- If the vertices in level 1 and level 2 in the previous frame are located on the salient points with high curvature, their motion vectors are estimated by using the method described in Section 2.1. Otherwise, the motion vectors (MV) are estimated by trying to minimise the approximation error. During shape MV encoding, motion vectors are first predicted from MVs in the coarser levels and/or the current level. The prediction error is encoded by using a VLC scheme. Some video objects have complicated shapes, requiring more vertices in level 1 and 2 to represent them. Therefore, the MV estimation and prediction-coding method is not efficient. In the proposed scheme, the multi-model selection method is used in these two levels. The selected coding models are MV estimation/prediction-coding method and scalable intra-shape coding method. The method, which generates the shorter bitstream, is selected and encoded, together with the shorter bitstream.
- For level 3, as there are more approximation vertices, the size of the list update information is significant.

Furthermore, it is very hard to estimate the correspondence of two contour segments based on the curvature information, which provide a lossless approximation of contour. The performance of the MC-based method is not satisfactory. Therefore, in the proposed method, the scalable intra-encoding scheme described in [1] is used. Therefore, for the vertices of level 3, no motion estimation is conducted as the motion estimation of vertices is less efficient for shape coding.

After predictive coding of each layer, some vertices may change their layer status, for example, one vertex changes from layer 1 to layer 0 or to layer 2 due to shape deformation. This information should be updated and used for the shape coding of next video frame. This update process can be achieved in encoder and decoder simultaneously without information transmission.

3. EXPERIMENTAL RESULTS

The performance of the proposed algorithm has been tested using “Weather” and “Kids” sequences, which are widely used during test in other published papers. The performance is compared with the predictive shape coding scheme (GPSC) in [2] and CAE in MPEG-4 [3]. For the CAE method, the inter-model is used instead of intra-coding model.

From the various ways to measure distortion, we utilize the following additive distortion metric per frame, which has also been used in the MPEG-4 standardization process to evaluate the performance of competing algorithms:

$$D_n = \frac{Number\ of\ Pixels\ in\ Error}{Number\ of\ Interior\ Pixels} \tag{1}$$

where a pixel is said to be in error if it belongs to the interior of the original object and the exterior of the approximating object, or vise-versa.

Figure 3 presents the bit distortion curves of the proposed algorithm for (a) Weather and (b) Kids sequence. It shows that our proposed algorithm can achieve better R-D performance than that of CAE and GPSC techniques.

4. CONCLUSIONS

Experimental results show that the proposed scalable predictive shape coding scheme can achieve better R-D performance than the existing published predictive shape coding method and the CAE method in MPEG-4. The proposed scalable shape coding method can achieve great improvement in compression performance by exploiting the geometrical knowledge of coarser levels, statistical entropy coding, and novel contour motion estimation scheme. Most importantly, the proposed scheme can achieve scalable shape coding, which facilitates error protection and error concealment of shape information. It also allows other functions, such as shape retrieval.

5. REFERENCES

[1] M. Hu, S. T. Worrall, A. H. Sadka, A. M. Kondoz, “A scalable vertex-based shape intra-coding scheme for video objects” *In Proceedings of IEEE ICASSP-2004*, Vol. 3, Montréal, Canada, pp. 273-276, May 2004

[2] J.I. Kim, A.C. Bovik, and B.L. Evans, “Generalized predictive binary shape coding using polygon approximation”, Signal Processing: Image Communication, pp. 643-663, July 2000

[3] ISO/IEC 14496-2: 2001, “Information Technology – Coding of audio-visual objects- Part 2: Visual”, 2001

[4] Cho Sung Ho, Kim R. C., Oh Seung S., and Lee Sang U., “A coding technique for the contours in smoothly perfect eight-connectivity based on two-stage motion compensation”, IEEE Trans. On CSVT, Vol.9, no.1, pp.59-69, Feb. 1999

[5] G. Melnikov, G. M. Schuster, and A. K. Katsaggelos, "Inter mode vertex-based optimal shape coding," in Proc. ICASSP-1999, Vol.6, pp.3125-3128, 1999

[6] Lu Z., and Pearlman W. A., "Motion Compensated Two-link Chain Coding for Binary Shape Sequence," *Visual Communications and Image Processing 2002*, Proceedings of SPIE, Vol. 4671, Jan. 21-23, pp. 356-362, 2002

[7] G. Melnikov, and A. K. Katsaggelos, "A rate-distortion optimal scalable vertex based shape coding algorithm," in Proc. ICASSP-2000, Vol.4, Turkey, June 5-9, pp.1947-1950, 2000

[8] G. Melnikov, and A. K. Katsaggelos, “Shape approximation through recursive scalable layer generation”, In Proc. IEEE ICIP 2000, Vol.2, pp. 915-918, Sept. 2000

[9] O’Connell K. J., “Object-adaptive vertex-based shape coding method”, IEEE Trans. on CSVT, Vol.7, no.1, pp.251-255, Feb. 1997

[10] H. Wang, G. M. Schuster, A. K. Katsaggelos, T. N. Pappas, “An efficient rate-distortion optimal shape coding approach utilising a skeleton-based decom-position”, IEEE Trans. on IP, Vol.12, no.10, pp. 1181-1193, Oct. 2003

[11] ISO/IEC 15938-3:2002, “Information Technology – Multimedia Content description Interface- Part 3: Visual”, 2002

[12] Mokhtarian F and Bober M., *Curvature Scale Space Representation: Theory, Applications & MPEG-7 Standardization*, Kluwer Academic, ISBN: 1-4020-1233-0, 2003

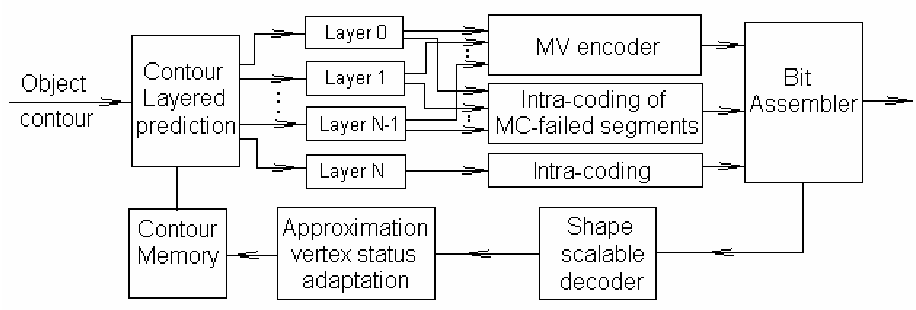
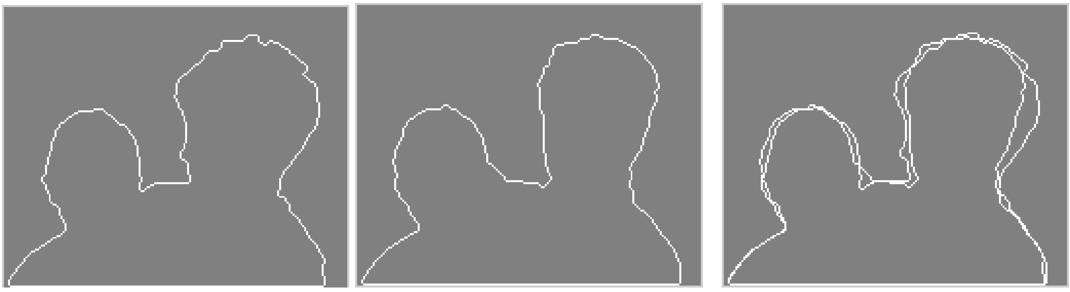
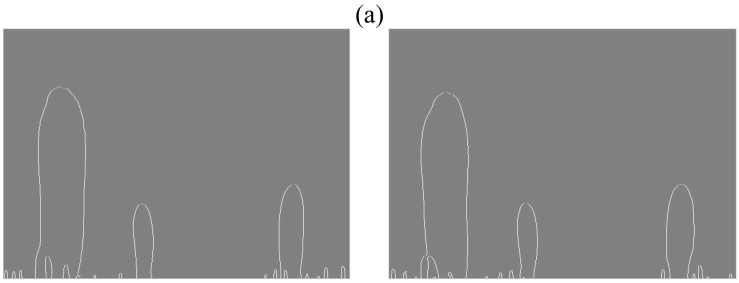


Figure 1 – Diagram for predictive scalable shape coding

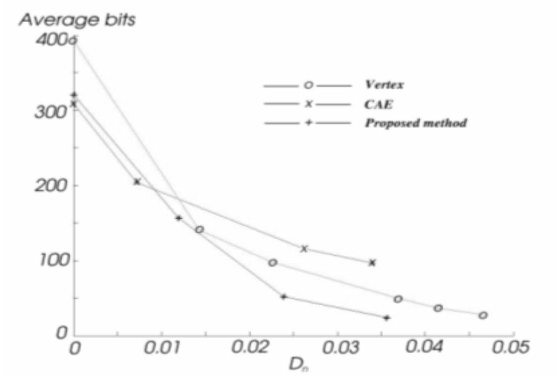


(a)

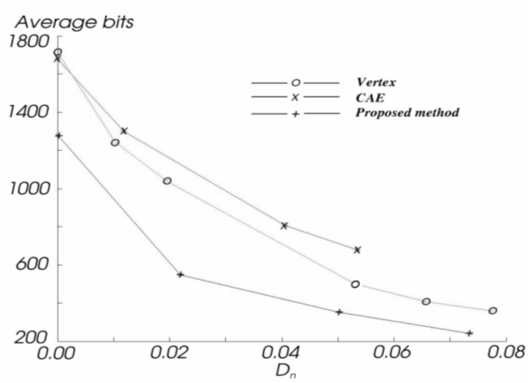


(b)

Figure 2 - Two object contours from Motr_dhtr sequence, and their zero crossing of contour curvature image with the scale : (a) Original object contours and their overlap to show contour motion; (b) the zero crossing of curvature across the scale .



(a)



(b)

Figure 3 Comparison of R-D performance of the proposed scalable inter-shape coding method with those of other coding scheme for (a) Weather and (b) Kids sequence