RATE-DISTORTION OPTIMAL SKELETON-BASED SHAPE CODING

Haohong Wang, Aggelos K. Katsaggelos, and Thrasyvoulos N. Pappas

Department of Electrical and Computer Engineering
Northwestern University, Evanston, IL 60208, USA
Email: {haohong, aggk, pappas}@ece.northwestern.edu

ABSTRACT

We present a new shape-coding approach, which decouples the shape information into two independent data sets, the skeleton and the distance of the boundary from the skeleton. The major benefit of this approach is that it allows a more flexible trade-off between accuracy of the approximation and bit-allocation cost, and thus, provides the possibility of better performance in the operational rate-distortion (ORD) optimal sense than other reported techniques. The characteristics of these data sets are studied and various approximation approaches are applied on each of them to reach an ORD optimal result. We apply, for example, polygonal approximation on both the skeleton and distance data. We demonstrate that the resulting approach outperforms existing ORD optimal approaches.

1. INTRODUCTION

In recent years, object oriented video coding has received a lot of attention because it facilitates retrieval, interactive editing, and manipulation of videos. Within the MPEG-4 standardization effort [1], contour-based shape coding methods have been developed and proved to be very efficient. However, any good coding algorithm should provide a rigorous tradeoff between the encoding cost and the resulting distortion. In [2] and [3], a framework for the rate-distortion operationally optimal encoding of shape information in the intra and inter modes is proposed. Polygon/spline approximation techniques are adopted to represent the boundary, and the control points of these curves are encoded to achieve the ORD optimal result.

Recent investigations have considered alternative representations of shape, which could allow a more flexible trade-off between accuracy and bit-allocation cost. In [5], the skeleton decomposition was proposed as an alternative shape description method for biology research. The main idea in the skeleton decomposition is to find a set of points that are equidistant from the object boundary by means of maximal inscribed disks. The description consists of the locus of the center of each inscribed disk and its associated radius [6]. This is the morphological definition of a skeleton, and has been employed for coding of binary images [7] and contour objects [8]. However, this definition can result in many extra branches (bones) in the skeleton, when the boundary has many outward ripples (see example in [6, p. 377]), which results in coding inefficiencies. In compressing the skeleton points, all of these approaches [7,8] do not appear to take advantage of the spatial continuity of the skeleton. Furthermore, if progressive contour transmission of images is considered, the case becomes worse. This is because in the progressive system, the skeleton points at coarser levels are farther apart from each other, resulting in lower compression rates.

In this paper, a variant of the morphological skeleton is defined and an ORD optimal shape coding approach is presented. The object shape is decomposed into the skeleton (defined as the midpoints between the two boundary points) and the distance of the boundary points from the skeleton in the horizontal direction. These two data sets are not correlated. This decoupling allows more flexibility in encoding, that is, it allows the application of different transform and compression methods for each data set, according to their characteristics. Furthermore, the skeleton of an object can be used for the estimation of the object motion in the Inter-mode. As an example of ways to encode the two data sets, we apply a polygonal approximation on the skeleton data and a polygonal or DCT-based approximation on the distance data. Arithmetic coding is used for the approximation encoding of the resulting symbols.

This paper is organized as follows. Section 2 provides a description of the skeleton-based shape representation. Section 3 describes the lossy polygonal approximation of the skeleton data, and the lossy coding of the distance data, and illustrates the ORD optimal process for allocating bits between the skeleton and distance data sets. Section 4 presents experimental results, and conclusions are drawn in the last section.

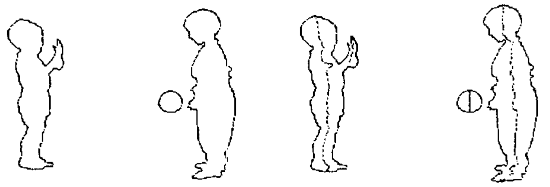
2. SKELETON-BASED SHAPE REPRESENTATION

The basic idea of skeletonization is to represent an object by one or more 2D curves (skeletons with associated distance data). Each pixel on the skeleton is associated with the distance to the closest boundary pixel in some direction. We use the horizontal distance for both simplicity and efficiency. The object can be recovered from the skeleton. Extracted skeletons contain features of the object like object center, object height etc., which could be helpful in video processing. Figure 1 shows examples of skeletonization for a frame of the "kids" sequence.

The binary shape of an object is defined by:

$$M_k=\{m_k(x,y) \mid 0\le x<X,\ 0\le y<Y\},\ m_k(x,y)\in\{0,1\}, \quad (1)$$

where X and Y are the dimensions of the frame. A pixel is part of the visual object if it satisfies $m_k(x,y)=1$, and belongs to the background if $m_k(x,y)=0$.



(a) original objects (b) objects with skeleton

Figure 1 An example of skeletonization

We can represent the extracted skeletons as the set of points (x,y) at the "center" of the object in the horizontal direction and the associated distance from the boundary, i.e.,

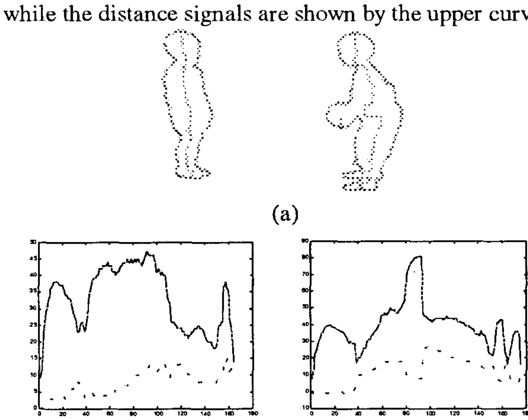
$$S=\{(x,y,d) \mid m_k(x-d-1,y)=0 \text{ and } m_k(x+d+1,y)=0 \text{ and } m_k(x-a,y)=1 \text{ and } m_k(x+a,y)=1 \text{ for } \forall a \text{ with } 0\le a\le d\}, \quad (2)$$

where x (the horizontal axis), a and d have half-pel accuracy, and $x-a$, $x+a$ are integers.

3. LOSSY CODING OF SHAPE DATA

The decoupling of the boundary data into skeleton and distance data allows us to apply different lossy coding techniques on each of them. Since the skeleton and the distance data typically have different characteristics, this method allows us to take advantage of such difference. For instance, if the object has near axial symmetry, the skeleton will be smooth. Similarly, if the two boundaries are nearly parallel, the distance function will be smooth. In both cases there is high correlation between the boundary data, and our representation decorrelates them. Our experimental results verify that this decoupling provides a compression advantage.

Figure 2 shows an example of decomposition into skeleton and distance data. In Fig. 2(a), a frame with 2 objects and 3 major skeletons is shown. Fig. 2(b) shows the skeleton and distance data of the leftmost major skeleton for 2(a), while Fig. 2(c) compares the skeleton data and distance data of the rightmost major skeleton for 2(a). The skeleton signals are shown by the bottom curve, while the distance signals are shown by the upper curve.



(a)

(b) (c)

Figure 2 Decomposition into skeleton and distance data

Notice that in Fig. 2(b), the skeleton is smoother than the distance function. In Fig. 2(c), there is a lot less correlation between the boundaries, and therefore, neither function is very smooth. Since highly correlated boundaries are quite common, our approach results in a substantial advantage.

3.1 Polygonal approximation

We utilized the ORD optimal polygonal approximation of [2,3,4] for encoding the two data sets. The approach is based on Lagrangian relaxation and a shortest path algorithm. There is an inherent tradeoff between the rate and the distortion in the sense that a small distortion requires a high rate, whereas a small rate results in a high distortion. So the task is to solve the problem:

$$\min_{N,p_0,p_1,\dots p_{N-1}} D(P_0,\dots P_{N-1}),\ \textit{subject to}: R(P_0,\dots P_{N-1})\le R_{\max}, \quad 3(a)$$

or its dual

$$\min_{N,p_0,p_1,\dots p_{N-1}} R(P_0,\dots P_{N-1}),\ \textit{subject to}: D(P_0,\dots P_{N-1})\le D_{\max}, \quad 3(b)$$

where P_0, P_1, … P_{N-1} are the locations of the unknown vertices of the approximating polygon, N is their unknown number, and R(.) and D(.) are the rate and distortion functions, respectively.

3.2 Lossless skeleton and lossy distance data coding

If we represent the skeleton in a lossless manner, and only apply polygonal approximation to the distance data, then the problem becomes:

$$\min_{N,p_0,p_1,\dots p_{N-1}} D(P_0,\dots P_{N-1}),\ subject\ to: R_d \le R_{max} - R_s, \qquad (4)$$

where R_s is the rate for representing the skeleton data, which is defined first, and R_d is the rate for representing the distance data. R_s is fixed once the skeleton is extracted.

As a means of evaluating the distortion we use the metric adopted by MPEG-4, that is,

$$D_{MPEG-4} = \frac{Number\ of\ pixels\ in\ error}{Number\ of\ Interior\ pixels}, \qquad (5)$$

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 vice-versa.

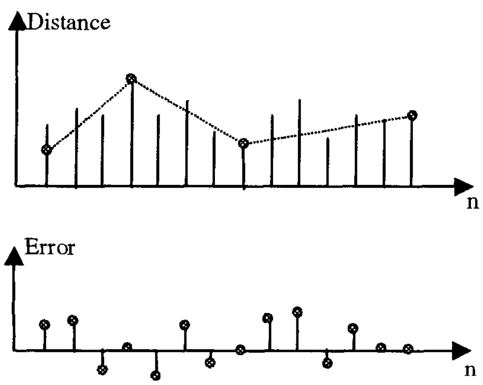
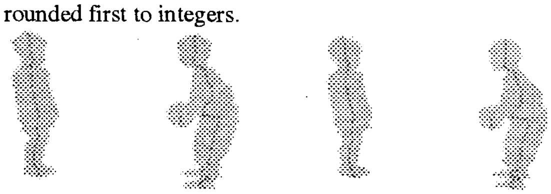


Figure 3 Polygonal approximation on distance data

Figure 3 illustrates a polygonal approximation of the distance data (dashed lines) we obtain by the ORD algorithm. The vertices of the polygon are obviously only needed for the approximation and need to be encoded. As shown in this figure, the distortion is the sum of the absolute value of the errors, also shown in Figure 3 (bottom part). The floating-point values of the error are rounded first to integers.



(Rate=976, Distortion=3.9%) (Rate=1080, Distortion=2.35%)

Figure 4. Examples of distance data approximation

Figure 4 shows two examples with different rate-distortion tradeoffs on the distance data.

3.2 Lossy coding of both skeleton and distance data

In order to get a more flexible tradeoff between coding accuracy and bit allocation cost, we apply polygonal approximation on both the skeleton and distance data. Then, problem (3) becomes:

$$\min_{N,M,p_0,p_1,\dots p_{N-1},q_0,q_1,\dots q_{M-1}} [D_s(p_0,\dots p_{N-1}) + D_d(q_0,q_1,\dots q_{M-1})],$$
$$subject\ to: \quad R_s+R_d \le R_{max}, \qquad (6)$$

where we assume that the distortion from the skeleton approximation (D_s) and the distortion from the distance approximation (D_d) are additive, although, in general, they are not. It is straightforward to show that $D_{MPEG-4} \le D_s+D_d$, which means that in practice, we may end up with a slightly better solution than the one provided by the algorithm. The optimization problem (6) is solved by first relaxing it and then minimizing $D_s+D_d+\lambda(R_s+R_d)$. By considering various values of λ, the operational rate distortion (ORD) curve in Figure 5 is obtained. In this figure, the actual MPEG-4 distortion is also shown, which results in an improved ORD curve. Both curves were obtained by processing the first frames of the "kids" sequence.

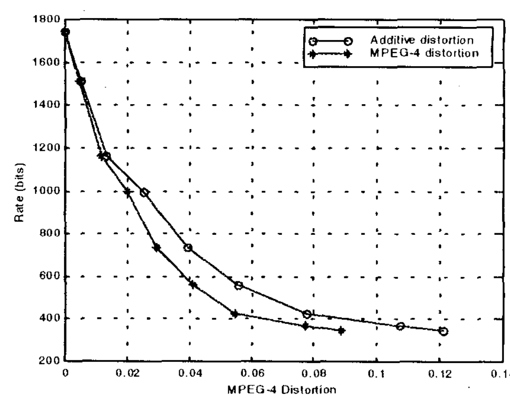
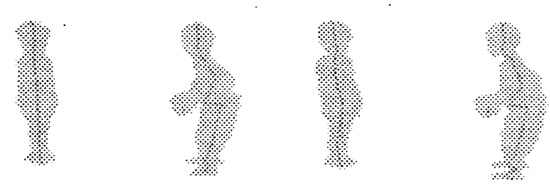


Figure 5. ORD curves for problem (6)



(rate=376, Distortion=8.86%) (rate=592, Distortion=4.1%)

Figure 6 Shape approximation results according to Eq.(6)

Figure 6 shows two results using the optimization of Eq. (6). Comparing this figure with Figure 4, we can see that by introducing distortion on the skeleton, the rate is reduced by about 40% (for distortion around 4%).

4. EXPERIMENTAL RESULTS

A number of experiments have been contacted. Some of these address the suitability of various compression techniques for compressing each of the skeleton and distance data. Two sets of experimental results are reported in Figure 7. In the first experiment, the skeleton was encoded with no error, while polygonal approximation was applied to the distance data. The results are indicated by the "+" in Figure 7. In the second experiment, both the skeleton and the distance were encoded using the polygonal approximation. The results of the optimal approximation are indicated by "*". In addition, in Figure 7, the results obtained in [3] are shown, indicated by curves with "o" for polygonal approximation, and with "x" for spline approximation. The results are obtained with the SIF sequence "kids".

The distortion axis represents the average of the $D_{MPEG\text{-}4}$'s distortion defined in Eq. (5) for one frame, over 100 frames. As it can be inferred from Figure 7, the decomposition of the boundary data into two data sets (skeleton and distance), with different characteristics, allows for their coefficient exploitation resulting in better compression results. In addition, applying polygonal approximation on both skeleton and distance data results in a better performance than when applying polygonal approximation only on the distance data.

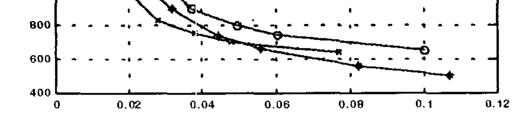


Figure 7 Rate-Distortion curves

The figure also shows that the spline approximation used in reference [3] to encode the boundary of the object gives the best performance for low distortions (0.01-0.05). If a similar spline approximation is used to encode the skeleton data, we expect similar improvements in performance.

5. CONCLUSIONS

In this paper, we presented a skeleton-based shape-coding algorithm. By decoupling the shape object data into the skeleton and distance data, we create a new scheme that reduces their correlation. This approach together with polygonal approximation of the skeleton and the distance data results in a significant improvement in rate-distortion efficiency with respect to other ORD optimal shape encoders.

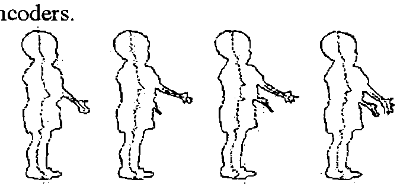


Figure 8 Same object in consecutive frames

Current work concentrates on answering the question of which is the most suitable algorithm for compressing the skeleton and the distance data. In addition, we are investigating the extension of the method to the 3-D case, where the skeleton and distance data can be predicted from the previous frame. As indicated in Fig. 8, in this case, skeleton approximation is expected to offer additional advantages.

6. REFERENCES

[1] A. K. Katsaggelos, L. Kondi, F. W. Meier, J. Ostermann, and G. M. Schuster, "MPEG-4 and Rate Distortion Based Shape Coding Techniques", *Proc. IEEE*, pp.1126-1154, June 1998.

[2] G. M. Schuster and A. K. Katsaggelos, "An optimal boundary encoding scheme in the rate distortion sense", *IEEE Trans. on Image Processing*, vol. 7, pp. 13-26, January 1998.

[3] G. M. Schuster, G. Melnikov and A. K. Katsaggelos, "Operationally Optimal Vertext-based Shape Coding", *IEEE Signal Processing Magazine*, pp. 91-108, November, 1998.

[4] R. Nygaard, G. Melnikov, and A. K. Katsaggelos, "A Rate Distortion Optimal ECG Coding Algorithm", *IEEE Trans. on Biomedical Engineering*, vol. 48, pp. 28-40, January, 2001.

[5] H. Blum, "A transformation for extracting new descriptors of shape", In *Models for the Perception of Speech and Visual Form*, (W. Wathen-Dunn, editor), pp. 362-380, Cambridge, Mass., MIT Press, 1967.

[6] J. Serra, *Image Analysis and Mathematical Morphology*, Academic Press, 1982.

[7] P. A. Maragos and R. W. Schafer, "Morphological skeleton representation and coding of binary images", *IEEE Trans. on Acoustics, Speech and Signal Processing*, 34(5): 1228-1244, October, 1986.

[8] P. Salembier, P. Brigger, J. R. Casas and M. Pardas, "Morphological Operations for Image and Video Compression", *IEEE Trans. on Image Processing*, vol. 5, No. 6, pp. 881-898, June, 1996.