AN OPTIMAL SHAPE ENCODING SCHEME USING SKELETON DECOMPOSITION

Haohong Wang, Guido M. Schuster[*]*, Aggelos K. Katsaggelos, and Thrasyvoulos N. Pappas*

Image and Video Processing Lab (IVPL)
Department of Electrical and Computer Engineering
Northwestern University Evanston, IL 60208, USA
Email: {haohong, aggk, pappas}@ece.northwestern.edu

[*] Abteilung Elektrotechnik
Hochschule für Technik, Rapperswil
CH-8040 Rapperswil, Switzerland
Email: guido.schuster@hsr.ch

***Abstract*—This paper presents an operational rate-distortion (ORD) optimal approach for skeleton-based boundary encoding. The boundary information is first decomposed into skeleton and distance signals, by which a more efficient representation of the original boundary results. Curves of arbitrary order are utilized for approximating the skeleton and distance signals. For a given bit budget for a video frame, we solve the problem of choosing the number and location of the control points for all skeleton and distance signals and for all boundaries within a frame, so that the overall distortion is minimized. The problem is solved with the use of Lagrangian relaxation and a shortest path algorithm in a 4D directed acyclic graph (DAG) we propose. By defining a path selection pattern, we reduce the computational complexity of the 4D DAG shortest path algorithm from $O(N^5)$ to $O(N^4)$, where N is the number of admissible control points for a skeleton. A suboptimal solution is also presented for further reducing the computational complexity of the algorithm to $O(N^2)$. The proposed algorithm outperforms experimentally other competing algorithms.**

Keywords— shape coding, skeleton decomposition, optimal.

1. INTRODUCTION

MPEG-4 [1] is the first international standard capable of encoding video objects with arbitrary shape. Within the MPEG-4 standardization effort [2], several contour-based shape coding methods have been developed and compared. In [3,4], lossy vertex-based polygonal approximations are considered. The placement of vertices allows for a direct control of the local variations of the shape approximation error. Those encoders are not optimal because they do not provide a rigorous tradeoff between the encoding cost and the resulting distortion. In [5,6], a framework for the operationally rate-distortion (ORD) optimal encoding of shape information in the intra and inter modes is proposed. Polygonal/spline approximation techniques are adopted to represent the boundary; the control points of these curves are encoded to achieve the ORD optimal result.

Morphological skeleton decomposition [7] is another approach for shape representation. Techniques utilized for the encoding of morphological skeletons are inefficient (especially with skeletons with many extra branches), since such skeletons are sparsely distributed. In [8], we proposed a new skeleton decomposition, which allows for a more flexible tradeoff between approximation error and bit budget. 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. The skeleton points are connected in the vertical (y-axis) direction. This represents a distinct advantage over morphological skeleton decomposition, especially when progressive transmission of the shape is considered.

In this paper, we propose an overall optimal skeleton-based encoding scheme in the rate-distortion sense. We apply polygonal approximation on both the skeleton and the distance signals. By converting the coding problem into a graph theory problem, a four-dimensional (4D) directed acyclic graph (DAG) shortest path algorithm is applied for obtaining the optimal solution.

This paper is organized as follows. Section 2 provides a description of the skeleton-based shape representation. Section 3 formulates the problem. Section 4 describes the Lagrangian multiplier method and the 4D DAG shortest algorithm applied to solve the proposed constrained problem. Section 5 discusses solutions for more general cases. Section 6 reports experimental results, and section 7 concludes the paper.

II. SKELETON-BASED SHAPE REPRESENTATION

We use the boundary form to represent object shape, by

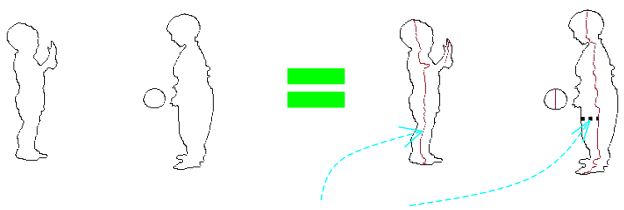
$$B=\{b_1(x,y), b_2(x,y), \ldots\ldots, b_K(x,y)\}, \quad (1)$$

where $b_i(x,y)$ is the ith boundary pixel with x as the horizontal and y as the vertical axes, and K the total number of pixels on the boundary. For any integer i $(1 \le i \le K-1)$, $b_{i+1}(x,y)$ is an 8-connected neighboring pixel of $b_i(x,y)$. We represent the extracted skeletons S as the set of points (x,y) at the "center" of the object and the associated horizontal distance from the boundary (see Fig. 1), i.e.,

$$S=\{(x,y,d)|\ (x+d,y) \in B \text{ and } (x-d,y) \in B\}, \quad (2)$$

where d has half-pixel accuracy.

The decomposition achieved by skeletonization results in two signals (skeleton and distance), which are not correlated with each other. It is therefore expected that the encoding of these two signals will be more efficient than the encoding of the original boundary information. In addition, one of the two signals can be quite inexpensive to encode. This is demonstrated by the two special cases shown in Fig. 2. In Fig. 2(a), the skeleton signal conveys all the information of the boundary (the distance signal is constant), while in Fig. 2(b), the opposite occurs. In both cases, the 2D-shape information is represented by a 1D signal, which results in compression efficiency.



OBJECT = SKELETON + DISTANCE

Figure 1 Example of skeletonization

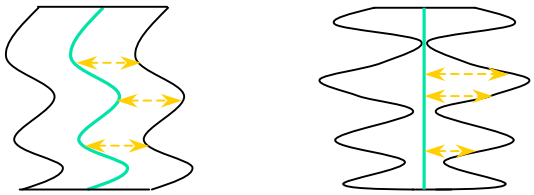


Figure 2 Examples of shapes for which one of the two signals resulting from skeletonization is constant

The situation represented by the synthetic signals in Fig. 2 is also encountered if one considers intervals of an arbitrary shape. For example, the skeleton and distance data of the kid on the left of Fig. 1 is shown in Fig. 3. It can be observed that there are subintervals for which the value of the skeleton or distance data is either constant or can be approximated to be constant without introducing a major error.

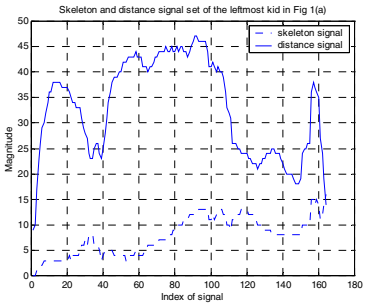


Figure 3 Decomposition into skeleton and distance data

III. PROBLEM FORMULATION

To simplify the problem description, we first assume that there is only one object that contains only one skeleton, and defer the solution for general case of encoding multiple objects with multiple skeletons to section *V*. We seek to define the skeleton signal set $S=\{S_1, S_2, \ldots S_N\}$ and the corresponding distance signal set $T=\{T_1, T_2, \ldots T_N\}$.

A. Distortion Metric

The distortion metric adopted by MPEG-4, which is also utilized in this work, is given by

$$D_{MPEG-4} = \frac{Number\ of\ pixels\ in\ error}{Number\ of\ Interior\ pixels} \qquad (3)$$

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.

Let us denote the distortion of the skeleton by $D(S)=\{D_{S1}, D_{S2}, \ldots, D_{SN}\}$, where D_{Si} is the distortion incurred by the ith skeleton pixel. Correspondingly, the distortion of the distance signal is denoted by $D(T)=\{D_{T1}, D_{T2}, \ldots, D_{TN}\}$. Clearly, all distortion elements are non-negative.

B. Bit Rate

Let us denote the total available bit rate for the encoding of the object shape in a frame by R_{tot}. Then $R_{tot}=R_0+R(S)+R(T)$, where R_0 represents the bits required for the encoding of the starting points of the skeleton, $R(S)$ the bits allocated for the encoding of the skeleton signal, and $R(T)$ the bits allocated for the encoding of the distance signal. The skeleton and distance data will be approximated by a curve of a certain order. For example, if straight lines are used for the approximation, two control points are needed to define a line segment; if on the other hand, second order curves are used, such as splines, three control points are needed to define a curve segment. The location of the control points or vertices is encoded and utilized for the reconstruction of the signal. Assuming that the skeleton has M vertices $\{V_{S1}, V_{S2}, \ldots V_{SM}\}$, $R(S)=\sum_{i=1}^{M} r(V_{Si},\ldots,V_{S(i-o)})$, where $r(V_{Si,\ldots,}V_{S(i-o)})$ is the rate required for the encoding of V_{Si}. Similarly R_T, the rate for encoding the corresponding distance signal, is defined as $R(T)=\sum_{i=1}^{Q} r(V_{Ti},\ldots,V_{T(i-o)})$, where $r(V_{Ti,\ldots,}V_{T(i-o)})$ represents the rate for encoding the V_{Ti}. Therefore,

$$R_{tot} = R_0 + \sum_{i=1}^{M} r(V_{Si},\ldots,V_{S(i-o)}) + \sum_{i=1}^{Q} r(V_{Ti},\ldots,V_{T(i-o)})\} \qquad (4)$$

C. Problem Description

The problem at hand is the operational rate distortion optimal encoding of the shape in a video frame (intra-shape encoding). That is, given a bit budget for the frame, we want to find the encoding of the shapes, which result in the smallest distortion. More specifically, we are solving the following constrained optimization problem with unknown the number and location of the control points,

$$\min D_{tot}, \text{ subject to } R_{tot} \le R_{max}, \qquad (5)$$

where D_{tot} is the total distortion (the sum of the distortion per object boundary) and R_{max} the total given bit budget. It is not hard to prove $D_{tot}(S,T) \le 2 \cdot \sum_{i=1}^{N} \max(D_{Si}, D_{Ti})$. By utilizing Eq. (4), problem (5) can be rewritten as:

$$\min \sum_{i=1}^{N} \max(D_{Si}, D_{Ti}), \text{ subject to} \qquad (6)$$

$$\sum_{i=1}^{M} r(V_{Si},\ldots,V_{S(i-o)}) + \sum_{i=1}^{Q} r(V_{Ti},\ldots,V_{T(i-o)}) \le R_{\max} - R_0$$

IV. OPTIMAL SOLUTION

We define a Lagrangian cost function

$$J_\lambda(V_S, V_T) = \sum_{i=1}^{N} \max(D_{Si}, D_{Ti}) + \qquad (7)$$

$$\lambda\{\sum_{i=1}^{M} r(V_{Si},\ldots,V_{S(i-o)}) + \sum_{i=1}^{Q} r(V_{Ti},\ldots,V_{T(i-o)})\}$$

where λ is called the Lagrange multiplier. Using the Lagrange multiplier method, the constrained problem (6) is relaxed to an unconstrained problem, that is:

$$min\ J_\lambda(V_S, V_T). \tag{8}$$

We solve the problem using a 4D DAG shortest path algorithm. Given a polygonal approximation of both the skeleton and distance signals, we define a node space with elements the 4-tuples *(i,j,p,q)*, representing all combinations of the last two control points in the skeleton approximation *(i)* and *(p)* ($i \leqslant p$), and the last two control points in the distance signal approximation *(j)* and *(q)* ($j \leqslant q$), and links among these elements. Clearly, there is one node space for each possible approximation. There are only three kinds of links starting at node *(i,j,p,q)*. Let *s* denote the next vertex after *p* in the skeleton approximation and *t* the next vertex after *q* in the distance approximation. Then the three links describe the transition *(i,j,p,q)->(p,j,s,q)*, *(i,j,p,q)->(i,q,p,t)*, and *(i,j,p,q)->(p,q.s,t)*.

To simplify the computation, and also to make a dynamic programming technique applicable for seeking the optimal solution of problem (8), we define a state space, which is a subset of the union of all node spaces, with elements (so called states) *(i,j,p,q)* satisfying $i \leqslant q$ and $j \leqslant p$, and edges among elements. This will exclude from consideration those nodes *(i,j,p,q)* with segment *[i,p]* not overlapping with segment *[j,q]*. The motivation for this is twofold: 1) By removing the non-overlapping segments, we can later express the distortion as a sum of link distortions between states. This will make a dynamic programming solution possible. 2) The fewer the states the faster the algorithm, given we do not remove from consideration any feasible polygonal approximations with this pruning. There are only two kinds of edges starting at state *(i,j,p,q)*, which are corresponding to the first two kinds of links in node space. In other words, the two edges describe the transition *(i,j,p,q)->(i,q,p,t)* and *(i,j,p,q)->(p,j,s,q)*. It is important to note that excluding the third possibility does not exclude any optimal path.

To implement the algorithm to solve the optimization problem (8), we create a cost function $C(p_k)$ (assuming p_k is representing state *(i,j,p,q)*), which represents the minimum total rate and "distortion" up to and including state *(i,j,p,q)* in the state space. That is,

$$C(p_k) = \min\left\{\sum_{i=1}^{\min(p,q)} \max(D_{Si}, D_{Ti}) + \lambda\left(\sum_{i=1}^{y_s^{-1}(p)} r(V_{Si},\ldots,V_{S(i-o)}) + \sum_{i=1}^{y_d^{-1}(q)} r(V_{Ti},\ldots,V_{T(i-o)})\right)\right\}$$

where the function $y_s^{-1}(x)=t$ iff $y(V_{St})=x$, and $y_d^{-1}(x)=t$ iff $y(V_{Tt})=x$, where $y(V)$ is the index of the vertex V in the original signal set. The key observation for deriving an efficient algorithm is the fact that given a certain state of a path (p_{k-1}) and the cost function up to and including this state ($C(p_{k-1})$), the selection of the next state p_k is independent of the selection of the previous states $p_0, p_1, \ldots p_{k-2}$. This is true since the cost function can be expressed recursively as a function of the segment rates $\zeta(p_{k-1},p_k)$ and the segment distortion $d(p_{k-1},p_k)$. That is:

$$C(p_k) = min\,(C(p_{k-1}) + w(p_{k-1},p_k)) \tag{9}$$

where

$$w(p_{k-1},p_k) = d(p_{k-1},p_k) + \lambda\zeta(p_{k-1},p_k) \tag{10}$$

and

$$d(p_{k-1},p_k) = \begin{cases} \sum_{t=i+1}^{q} \max(D_{St}, D_{Tt}) & Transition\ occurs\ in\ skeleton\ data \\ \sum_{t=j+1}^{p} \max(D_{St}, D_{Tt}) & Transition\ occurs\ in\ distance\ data \\ \infty & otherwise \end{cases}$$

$$\zeta(p_{k-1},p_k) = \begin{cases} r(y_s^{-1}(p), y_s^{-1}(i)) & Transition\ occurs\ in\ skeleton\ data \\ r(y_d^{-1}(q), y_d^{-1}(j)) & Transition\ occurs\ in\ distance\ data \end{cases}$$

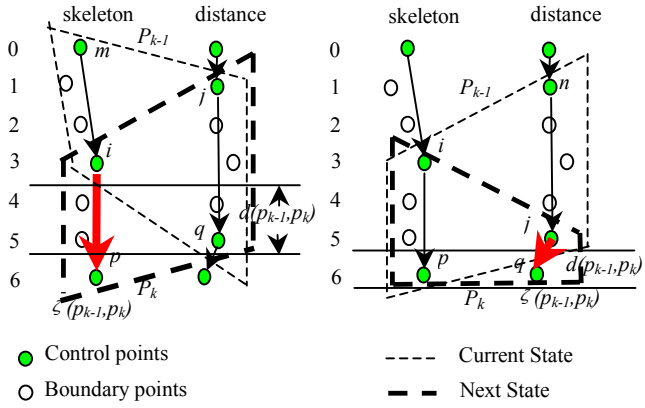


Figure 4 Examples of segment distortion and segment rate

Figure 4 shows an example of the segment distortion and segment rate. The figure on the right shows the next step relative to the figure on the left. It is easy to see how the edge distortions add up to the total distortion. In other words, we are showing that summing the above segment distortions up will result in the total distortion and, that these segment distortions are only dependent on state p_{k-1} and state p_k. Using (9), the problem stated in (8) can be formulated as a shortest path problem as in [5,6,8].

In summary, the state definition and the recursive representation of the cost function in (9) makes the future of the optimization process independent from its past, which is the foundation of the dynamic programming technique. The computational complexity of our 4D DAG shortest path algorithm is $O(N^5)$.

To speed up the performance, a relaxed distortion sub-optimal solution is obtained by assuming that $D_{tot} \approx 2 \cdot \sum_{i=1}^{N} (D_{Si} + D_{Ti})$ [8]. By doing that, problem (8) will be written as:

$$\min\left\{\sum_{i=1}^{N}(D_{Si}+D_{Di}) + \lambda\left(\sum_{i=1}^{M} r(V_{Si},\ldots,V_{S(i-o)}) + \sum_{i=1}^{Q} r(V_{Ti},\ldots,V_{T(i-o)})\right)\right\}$$

Since skeleton and distance data are independent, the original problem can be split into two functions:

$$\min\left\{\sum_{i=1}^{N} D_{Ti} + \lambda\sum_{i=1}^{Q} r(V_{Ti},\ldots,V_{T(i-o)})\right\},\ \min\left\{\sum_{i=1}^{N} D_{Si} + \lambda\sum_{i=1}^{M} r(V_{Si},\ldots,V_{S(i-o)})\right\}$$

which are quite straightforward to solve. The computational complexity is now $O(N^2)$.

V. GENERAL CASES

For those objects with multiple skeletons, assume the L skeletons $\{S^1,S^2,...,S^L\}$ are encoded with the associated distance signals $\{T^1,T^2,...,T^L\}$, then the optimization problem can be stated as

$$\min D_{tot}(S^1,S^2,...,S^L,T^1,T^2,...,T^L),$$
$$\text{subject to } \sum_{m=1}^{L}\{R(S^m)+R(T^m)\}\le R_{budget} \quad (11)$$

where the total distortion of the object depends on all L skeletons and distances. This problem can be solved by

$$\min D_{tot}(S^1,S^2,...,S^L,T^1,T^2,...,T^L)+\lambda\cdot\sum_{m=1}^{L}\{R(S^m)+R(T^m)\}$$

It can be easily shown that

$$D_{tot}(S^1,S^2,...,S^L,T^1,T^2,...,T^L)\le\sum_{i=1}^{L}D_{tot}(S^i,T^i) \quad (12)$$

In most cases, equality holds, and then the optimization problem depends into L problems: min $\{D_{tot}(S^m,T^m)+\lambda[R(S^m)+R(T^m)]\}$ $(1\le m\le L)$, which are identical to problem (8) solved in the previous section. In some rare cases, when the distortion pixel sets corresponding to two skeletons overlap, the total distortion is less than the sum of distortion in Eq. (12). In such cases, solving the decoupling problem yields a suboptimal overall solution. Since these cases are rare and the solution of the overall problem becomes extremely complicated, we will not consider it here.

For multiple objects boundary encoding, since the distortion calculation is defined on an object-by-object basis, the problem remains decoupled even if the distortion pixel sets of the two objects overlap. Then, the results of section *IV* apply.

VI. EXPERIMENTAL RESULTS

Some of the experimental results are shown in Fig. 5. In this experiment, both the skeleton and distance signals were encoded using the polygonal approximation. The results obtained by the application of the 4D DAG shortest path algorithm are indicated by "*". In addition, in Fig. 5, the results obtained by using the algorithm in [6] are shown by "o". Finally, the results obtained by applying the CAE (Context-based Arithmetic Encoding) method are shown. The results are obtained with the SIF sequence "kids". The distortion axis represents the average of the D_{MPEG-4}'s distortion defined in Eq. (3) for one frame, over 100 frames. As it can be inferred from Fig. 5, the decomposition of the boundary data into two signal data sets (skeleton and distance), with different characteristics, allows for their efficient exploitation resulting in better compression results.

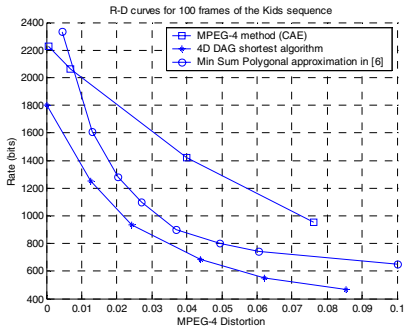


Fig. 5 rate-distortion curve

VII. CONCLUSIONS

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

REFERENCES

[1] R. Koenen, "MPEG-4 multimedia for our time", *IEEE Spectrum,* vol. 36, pp. 26-33, Feb. 1999.

[2] 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.

[3] P. Gerken, "Object-based analysis-synthesis coding of image sequences at very low bit rates", *IEEE Trans. Circuits Syst.*, vol.4, pp. 228-235, June 1994.

[4] K. J. O'Connell, "Object-adaptive vertex-based shape coding method", IEEE Trans. Circuits Syst. Video Technol., vol. 7, pp. 251-255, Feb. 1997.

[5] 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.

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

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

[8] H. Wang, A. K. Katsaggelos and T. N. Pappas, "Rate-Distortion Optimal Skeleton-based Shape coding", in Proc. Int. Conf. Image Processing, Thessaloniki, Greece, pp. 1001-1004, Oct. 2001.