MINMAX Optimal Shape Coding Using Skeleton Decomposition

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

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

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

ABSTRACT

In this paper, we consider the rate-distortion optimal encoding of shape information using a skeleton decomposition and the minmum maximum (MINMAX) distortion criterion. For bit budget constrained video communication applications, whose goal is to achieve as low as possible but almost constant distortion, the MINMAX criterion is the natural choice. We propose a 4D DAG (directed acyclic graph) shortest path algorithm implemented by dynamic programming to solve the minimum rate problem, and also provide a solution for the dual minimum distortion problem. Experimental results indicate that our algorithm has an outstanding performance compared with existing methods.

1. INTRODUCTION

Shape coding has attracted a lot of attention recently, as one of the most important parts of object-based video encoding, as supported, for example, by the MPEG-4 standard. A number of existing shape encoding algorithms are described and compared in [1]. Among the rate-distortion optimal ones are vertex-based encoders [2] and skeleton-based encoders [3,4]. Vertex-based approaches use polygons, B-splines, or even higher order curves to approximate the boundary contour of the object. The optimal number and locations of the control points can be found by minimizing the normalized mean-squared error between the boundary and the approximations. By decomposing the object shape into the skeleton (defined as the midpoints between two boundary points) and the distance from the boundary points to the skeleton in the horizontal direction, the skeleton-based approach utilizes curves of arbitrary order for approximating the skeleton and distance signals, and chooses the number and locations of the control points for all skeleton and distance signals and for all boundaries within a frame, to minimize the overall distortion using the MPEG-4 distortion metric. Both methods are using Lagrangian relaxation to obtain the operational rate-distortion optimal solution and dynamic programming to improve the efficiency of the algorithms.

Most shape encoders utilize the minimum average (MINAVE) distortion criterion in measuring distortion. The resulting solutions usually lead to unequal distortion across frames, which can cause "flickering problems" due to abrupt quality changes. An alternative approach to formalize the relationship between rate and distortion is the minimum maximum (MINMAX) distortion approach [5,6]. The philosophy behind this approach is that by minimizing the maximum source distortion, no single source distortion will be extremely high, and hence, the overall quality will be quite constant. In [6], the MINMAX approaches are reviewed and compared with the MINAVE approaches.

In this paper, we are proposing a rate-distortion optimal shape-coding scheme with skeleton decomposition using the MINMAX criterion. A 4D DAG shortest path algorithm is proposed to solve the optimization problem while dynamic programming is adopted to ensure the efficiency of the algorithm.

The rest of the paper is organized as follows. In section 2, the problem formation is presented. Section 3 demonstrates the optimal solution and section 4 provides our experimental results. We draw conclusion in the last section.

2. PROBLEM FORMULATION

The problem at hand is to minimize the rate for encoding a shape while guaranteeing that none of the pixels of the resulting approximated object are located farther than D_{max} (Euclidean distance) away from the original object shape. We also consider the dual problem that of minimizing the D_{max} subject to a bit budget.

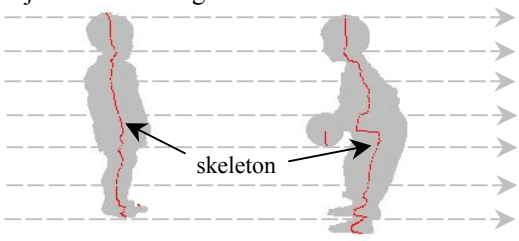


Figure 1 Example of object with skeleton

Figure 1 shows an example of the skeleton representation of the video object, where the object shape is decomposed into the skeleton (midpoints between the two boundary points), and the distance of the boundary points from the skeleton in the horizontal direction. To simplify the problem, we first assume that there is only one object that contains only one skeleton, and defer the solution of the general case of encoding multiple objects with multiple skeletons to section 3.3. We seek to define the skeleton signal set $S=\{S_1, S_2, \ldots S_N\}$ and the corresponding distance signal set $T=\{T_1, T_2, \ldots T_N\}$.

2.1. Distortion

As shown in Fig.2, we define a distortion band (shaded region), around the original boundary in implementing our solution. Clearly D_{max} is a threshold for approximating boundary distortion. In order to build up the relation between D_{max} and the distortion caused by the skeleton and distance, a new concept called maximum horizontal distortion (D_{Hmax}) is introduced, where $D_{Hmax}=\{\{D_{LL1}, D_{LR1}, D_{RL1}, D_{RR1}\}, \ldots, \{D_{LLN}, D_{LRN}, D_{RLN}, D_{RRN}\}\}$, and each vector $\{D_{LLi}, D_{LRi}, D_{RLi}, D_{RRi}\}$ represents the maximum allowable boundary distortion in the horizontal direction at the ith point of the skeleton (see Fig. 3 area B). Each D_{max} uniquely corresponds to a set of D_{Hmax}. Figure 3 shows an example of the concept. In areas A and C, the skeleton is extended in both directions at the length of D_{max} to cover those areas, which also expend the set of D_{Hmax} to include $\{D_{L(-Dmax)}, D_{R(-Dmax)}\}, \ldots, \{D_{L(-1)}, D_{R(-1)}\}, \{D_{L(N+1)}, D_{R(N+1)}\}, \ldots, \{D_{L(N+Dmax)}, D_{R(N+Dmax)}\}$. To simplify our problem, we will focus on area B in the following discussion.

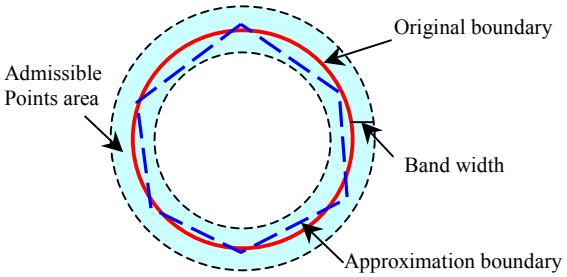


Figure 2. Concept of distortion band

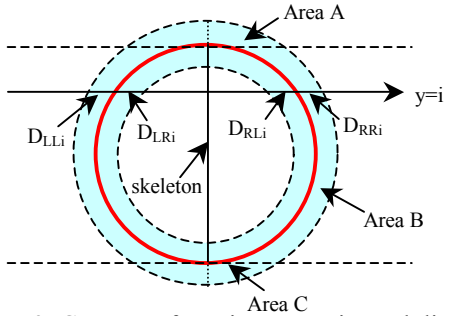


Figure 3. Concept of maximum Horizontal distortion

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 $D(T)=\{D_{T1}, D_{T2}, \ldots, D_{TN}\}$. The distortion elements could be positive or negative, depending on whether the approximated signal is larger or smaller than the original signal. So, the restriction of D_{max} is converted into a new set of restrictions as

$$D_{LLi} \le D_{Si} - D_{Ti} \le D_{LRi} \;\&\; D_{RLi} \le D_{Si} + D_{Ti} \le D_{RRi}\,. \quad (1)$$

The MPEG-4 distortion metric is used to evaluate the quality of reconstructed frames, which is given by

$$D_{MPEG-4} = \frac{Number\ of\ pixels\ in\ error}{Number\ of\ Interior\ pixels}, \quad (2)$$

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.

2.2. 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; while on the other hand, if second order curves are used, such as splines, then three control points will be 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}, which depends on o previous points, with o the order of the curve. 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,

$$R_{tot} = R_0 + \sum_{i=1}^{M} r(V_{Si},\ldots,V_{S(i-o)}) + \sum_{i=1}^{Q} r(V_{Ti},\ldots,V_{T(i-o)})\} \quad (3)$$

2.3. Problem Description

The problem at hand is the operational rate distortion optimal encoding of the shape in a video frame (intra-shape encoding). It can be either a minimum rate problem, or a minimum distortion problem.

The minimum rate problem is to find the minimum bit rate to encode the shape, given a set of maximum horizontal distortion D_{Hmax}. More specifically, we are solving the following constrained optimization problem with unknown the number and location of the control points for both skeleton and distance,

min R_{tot}, subject to $D_{LLi} \le D_{Si} - D_{Ti} \le D_{LRi}$, and

$$D_{RLi} \le D_{Si} + D_{Ti} \le D_{RRi} \text{ for all } i \ (1 \leqslant i \leqslant N). \quad (4)$$

The minimum distortion problem is to find the encoding of the shapes, which results in the set of smallest maximum horizontal distortion, given a bit budget for the frame. More specifically, we are solving the following constrained optimization problem with unknown the number and location of the control points,

$$\min \{D_{Hmax}\}, \text{ subject to } R_{tot} \leq R_{max}, \quad (5)$$

where R_{max} is the total given bit budget.

3. OPTIMAL SOLUTION

3.1 Minimum rate problem

This problem can be easily tackled by first converting it into a graph theory problem and then solving it by a 4D DAG shortest path algorithm. In the following, we are solving the simple case ($o=1$) as an example, since the more complicated case only increases the computational complexity, while in principle, there are no fundamental obstacles.

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)*.

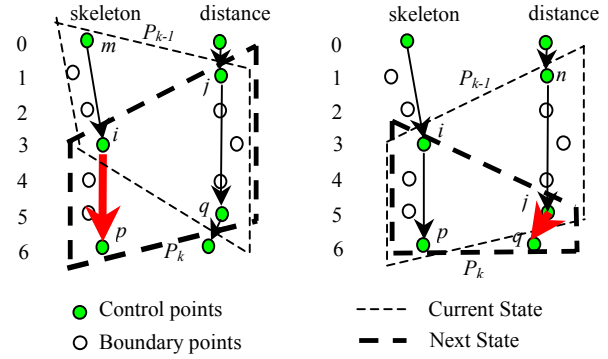


Figure 4 Example of states and edges

To simplify the computation, and also to make a dynamic programming technique applicable for seeking the optimal solution of problem (4), 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 of those nodes *(i,j,p,q)* with segment *[i,p]* not overlapping with segment *[j,q]*. Hence, 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 (4), we create a cost function $C(p_k)$ (assuming p_k is representing state *(i,j,p,q)*), which represents the minimum total rate up to and including state *(i,j,p,q)* in the state space. That is,

$$C(p_k) = \min \left\{ \sum_{i=1}^{y_s^{-1}(p)} r(V_{Si},...,V_{S(i-o)}) + \sum_{i=1}^{y_d^{-1}(q)} r(V_{Ti},...,V_{T(i-o)}) \right\} \quad (6)$$

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 $p_0,p_1,...p_{k-2}$. This is true since the cost function can be expressed recursively as a function of the segment rates $w(p_{k-1},p_k)$. That is:

$$C(p_k) = \min (C(p_{k-1}) + w(p_{k-1},p_k)) \quad (7)$$

where

$$w(p_{k-1}, p_k) = \begin{cases} r(y_s^{-1}(p), y_s^{-1}(i)) & \text{Transition occurs in skeleton data \& for } \forall i \in [i, \min(p,q)], (1) \text{ satisfied} \\ r(y_d^{-1}(q), y_d^{-1}(j)) & \text{Transition occurs in distance data \& for } \forall i \in [j, \min(p,q)], (1) \text{ satisfied} \\ \infty & \text{Otherwise} \end{cases}$$

In words, the rate for a source with a distortion, which is larger than the maximum permissible distortion, is set to infinity. This results in that given that a feasible solution exists, the approximation sequence, which minimizes the total rate, as defined in (4), will not give any source distortion greater than D_{max}.

Figure 4 shows an example of the states and edges. The figure on the right shows the next step relative to the figure on the left. We are showing that summing the above segment rate up will result in the total rates and, that these segment rates are only dependent on state p_{k-1} and state p_k. Using (7), the problem stated in (4) can be formulated as a shortest path problem as in [4].

In summary, the state definition and the recursive representation of the cost function in (7) makes the future of the optimization process independent of its past, which is the foundation of the dynamic programming technique. The computational complexity of our 4D DAG shortest path algorithm is $O(N^5)$.

3.2 Minimum distortion problem

The proposed optimal bit allocation algorithm for the minimum distortion problem is based on the fact that we can optimally solve the minimum rate problem. In other words, for every given D_{max}, we can find the approximation sequence which results in $R^*(D_{max})$, the minimum rate for encoding the combined sources, where

each source distortion has to be restricted by the maximum distortion D_{max}. In [6], it has been proved that $R^*(D_{max})$ is a nonincreasing function of D_{max}. So, we can use bisection to find the optimal D^*_{max} such that $R^*(D^*_{max})=R_{max}$, which solves the minimum distortion problem of (5).

It is not hard to prove $D_{tot}(S,T) \le 2 \cdot \sum_{i=1}^{N} \max(|D_{Si}|,|D_{Ti}|)$. So the corresponding MPEG-4 distortion D_{MPEG-4} of the obtained D_{Hmax} can be found by solving the following problem,

$$\min \sum_{i=1}^{N} \max(|D_{Si}|,|D_{Ti}|), \text{ subject to} \quad (8)$$

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

$$D_{LLi} \le D_{Si} - D_{Ti} \le D_{LRi}, \text{ and}$$

$$D_{RLi} \le D_{Si} + D_{Ti} \le D_{RRi}, \text{ for all } i \ (1 \leqslant i \leqslant N).$$

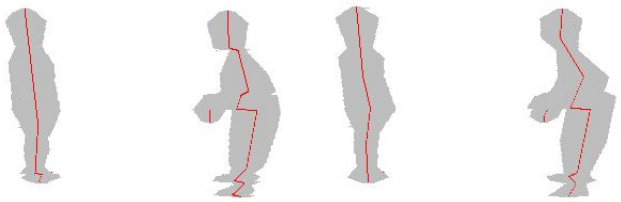
The problem can be solved using Lagrangian relaxation and dynamic programming as the way shown in [4].

3.3. General cases

For multiple object boundary encoding and the case of object with multiple skeletons, since the bit rate for vertices is always additive, and the distortion calculation is defined on separate boundary parts of the object corresponding to various skeletons, the problem can be decoupled into simple problems discussed previously even if the distortion pixel sets of the two objects overlap. Then, the results of this section apply.

4. EXPERIMENTAL RESULTS

A number of experiments have been conducted, some of which are reported here. Figure 5 shows a comparison of the results obtained using the proposed MINMAX approach with $D_{max}=3$ (see Fig. 5(a)) and using the MINAVE approach [4] (see Fig. 5(b)) to compress the frame shown in Fig. 1. The MPEG-4 distortion for both reconstructed images is *8.2%* (percentage of pixels in error). Comparing the feet of the left child in the frame, it is not hard to tell that the MINMAX result keeps the basic shape of the original object, while the MINAVE result may not, because the goal of MINAVE is to reduce the possible average distortion, no matter how distortion may vary from one part of the object to the other.



(a) MINMAX Image (b)MINAVE Image

Figure 5 Comparison of reconstructed images

We conducted experiments on the 100 frames of the "Kids" SIF sequence in the intra mode. Our results are compared with the results obtained by applying the CAE (Context-based Arithmetic Encoding) method [7], which is adopted in MPEG-4 standard, and the results obtained by using the vertex-based MINMAX polygonal algorithm in [2]. As shown in Fig. 5, our algorithm has an overall better performance than other methods, although Vertex-based MINMAX algorithm performs slight better for distortions in the range 0.045-0.073.

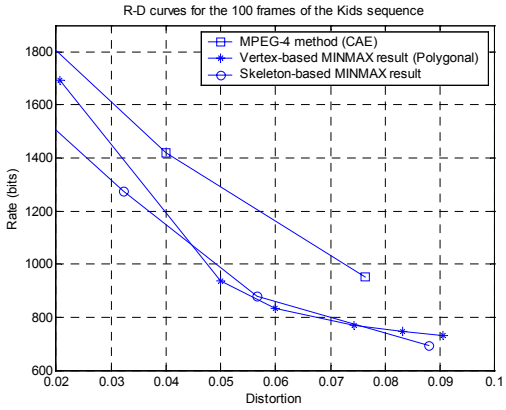


Figure 6 Rate-Distortion curves

5. CONCLUSION

In this paper, we present an optimal skeleton-based shape-coding algorithm using the minimum maximum distortion criterion. The concept of horizontal maximum distortion is introduced to enable the joint processing of skeleton and distance signals. A 4D DAG shortest path algorithm with an efficient dynamic programming implementation is proposed to solve the minimum-rate problem. The minimum-distortion problem is also solved using the fact that we can find the optimal solution to the minimum rate problem, which results in a no increasing operational rate distortion function. Experimental results demonstrate the improved performance of the proposed algorithm.

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, G. Melnikov, and A. K. Katsaggelos, "Operationally optimal vertext-based shape coding", *IEEE Signal Processing Magazine*, pp. 91-108, Nov. 1998.

[3] H. Wang, A. K. Katsaggelos, and T. N. Pappas, "Rate-distortion optimal skeleton-base shape coding", in *Proc. Int. Conf. Image Processing*, Thessaloniki, Greece, pp. 1001-1004, Oct. 2001.

[4] H. Wang, Guido M. Schuster, A. K. Katsaggelos, and T. N. Pappas, "An optimal shape encoding scheme using skeleton decomposition", in *Proc. Int. Workshop Multimedia & Expo*, St. Thomas, USA, Dec. 2002.

[5] Yan Yang, S. S. Hemami, "Minmax Frame Rate Control Using a Rate-Distortion Optimized Wavelet Coder," *Proc. IEEE Int. Conf. on Image Processing,* Kobe, Japan, October 1999.

[6] G. M. Schuster, G. Melnikov, and A. K. Katsaggelos, "A review of the minimum maximum criterion for optimal bit allocation among dependent quantizers", *IEEE Trans. Multimedia*, vol. 1, pp. 3-17, March 1999.

[7] MPEG-4 video VM 18.0, ISO/IEC JTC1/SC29/WG11 N3908, Pisa, Jan. 2001.