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 minimum 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 formulation 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.
Let us denote the distortion of the skeleton by $D(S) = \{D_{S_1}, D_{S_2}, ..., D_{S_N}\}$, where $D_{S_i}$ is the distortion incurred by the $i$th skeleton pixel. Correspondingly, the distortion of the distance signal is denoted by $D(T) = \{D_{T_1}, D_{T_2}, ..., D_{T_N}\}$.

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

$$D_{LLi} - D_{Si} - D_{LRi} \leq D_{S_i} \leq D_{LRi} + D_{RT} \leq D_{RRi} \leq D_{RRi}.$$  

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

$$D_{MPEG-4} = \frac{\text{Number of pixels in error}}{\text{Number of Interior pixels}},$$

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 $\{V_{S_1}, V_{S_2}, ..., V_{S_M}\}$, $R(S) = \sum_{i=1}^{M} r(V_{S_i}, ..., V_{S(i-o)})$, where $r(V_{S_i}, ..., V_{S(i-o)})$ is the rate required for the encoding of $V_{S_i}$, 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 $R(T) = \sum_{i=1}^{o} r(V_{T_{i-0}}, ..., V_{T(i-0)})$, where $r(V_{T_{i-0}}, ..., V_{T(i-0)})$ represents the rate for encoding the $V_{Ti}$. Therefore,

$$R_{tot} = R_0 + \sum_{i=o}^{M} r(V_{S_{i-o}}, ..., V_{S(i-o)}) + \sum_{i=o}^{o} r(V_{T_{i-o}}, ..., V_{T(i-o)})$$

### 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 $D_{max}$. 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,

$$\min R_{tot}, \text{ subject to } D_{LLi} \leq D_{S_i} - D_{Ti} \leq D_{LRi}, \text{ and}$$
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,

\[
\min |D_{\text{max}}| \quad \text{subject to } R_{\text{tot}} \leq R_{\text{max}},
\]

where \( R_{\text{max}} \) 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 (\( a = 1 \)) 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 \((p)\) \((i \leq p)\), and the last two control points in the distance signal approximation \((j)\) and \((q)\) \((j \leq q)\), 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 \((i,j,p,q) \rightarrow (p,j,s,q)\), \((i,j,p,q) \rightarrow (i,q,p,t)\), and \((i,j,p,q) \rightarrow (p,q,s,t)\).

![Figure 4 Example of states and edges](image)

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 a feasible solution, the approximation sequence, which minimizes the total rate, as defined in (4), will not give any source distortion greater than \( D_{\text{max}} \).

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 \( O(N^4) \).

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 \( D_{\text{max}} \), we can find the approximation sequence which results in \( R^*(D_{\text{max}}) \), the minimum rate for encoding the combined sources, where
(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.

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