AN OPTIMAL SEGMENTATION ENCODING SCHEME IN THE RATE DISTORTION SENSE

Guido M. Schuster[1] *Aggelos K. Katsaggelos*[2]

Northwestern University
Department of Electrical Engineering and Computer Science
McCormick School of Engineering and Applied Science
Evanston, IL 60208-3118, USA
[1]gschuster@nwu.edu
[2]aggk@eecs.nwu.edu

ABSTRACT

In this paper, we present two fast and efficient methods for the lossy encoding of object boundaries which are given as 8-connect chain codes. We approximate the boundary by a polygon and consider the problem of finding the polygon which leads to the smallest distortion for a given number of bits. We introduce two different approaches to find the best polygonal approximation when an additive distortion measure is employed. The first approach is based on Lagrangian relaxation and a shortest path algorithm. This scheme results in solutions which belong to the convex hull of the operational rate-distortion curve. The second algorithm is based on a tree pruning scheme and finds all optimal solutions. We present results of the proposed algorithms using an object from the “Miss America” sequence.

1. INTRODUCTION

A major problem in object oriented video coding [1] is the efficient encoding of object boundaries. There are two common approaches for encoding the segmentation information: a lossy approach, which is based on a spline approximation of the boundary [2], and a lossless approach which is based on chain codes [3]. The proposed boundary encoding scheme is a lossy scheme which can be considered a combination of the spline and the chain code approaches since the boundary is approximated by a polygon and its vertices are encoded relative to each other. The approximation of the boundary by a polygon is similar to the spline approach, whereas the encoding of the vertices is achieved with a chain code-like scheme.

The encoding of planar curves is an important problem in many different fields, such as CAD, object recognition, object oriented video coding, etc.. This research is motivated by object oriented video coding, but the developed algorithms can also be used for other applications. Chain coding was originally proposed by Freeman [3] and has attracted considerable attention over the last thirty years [4, 5, 6, 7, 8]. The most common chain code is the 8-connect chain code which is based on a rectangular grid superimposed on a planar curve. The curve is quantized using the grid intersection scheme [3] and the quantized curve is represented using a string of increments. Since the planar curve is assumed to be continuous, the increments between grid points are limited to the 8 grid neighbors, and hence an increment can be represented by 3 bits. There have been many extensions to this basic scheme such as the generalized chain codes [4], where the coding efficiency has been improved by using links of different length and different angular resolution. In [7] a scheme is presented which utilizes patterns in a chain code string to increase the coding efficiency and in [8] differential chain codes are presented, which employ the statistical dependency between successive links. There has also been interest in the theoretical performance of chain codes. In [5] the performance of different quantization schemes is compared, whereas in [6] the rate distortion characteristics of certain chain codes are studied.

In this paper, we are not concerned with the quantization of the continuous curve, since we assume that the segmentation is given with pixel accuracy. According to the 8-connect chain code encoding scheme, a pixel on a boundary is selected as the starting point of the chain code. Every future pixel can then be described as being one of the 8 closest neighbors of the current pixel. Hence 3 bits per boundary pixel are required to encode this incremental information between two pixels. The string of these increments forms the 8-connect chain code of the boundary. Most boundaries contain many straight lines or lines with a very small curvature, which result in runs of the same increment. Therefore a run-length encoding scheme can be used to encode the 8-connect chain code even more efficiently. Clearly the larger the number of straight lines a boundary contains, the more efficient a chain code/run-length scheme is. This is the idea behind some preprocessing algorithms [9] which are used to “straighten” the boundary, i.e., these algorithms lead to a new boundary which can be encoded more efficiently. Clearly this preprocessing introduces an error in the boundary representation, but as long as the error is small enough, the visual distortion is considered insignificant. Such approaches, however have at best an indirect control of the amount of the distortion introduced in the representation of the boundary.

In this paper, we present an algorithm where the preprocessing step and the chain code/run-length encoding are combined into an optimal lossy segmentation encoding scheme. The proposed approach offers complete control over the tradeoff between distortion and bit rate. Note that this is achieved in an optimal fashion, resulting in a very efficient encoding scheme.

In section 2. we define the problem and introduce the required notation. In section 3. we use Lagrangian relaxation to derive an algorithm for finding the polygon which leads to the smallest distortion for a given maximum bit rate. Like every Lagrangian relaxation based algorithm, this algorithm only finds the solutions which belong to the convex hull of the operational rate-distortion curve. In section 4. we propose a tree pruning algorithm which can find all solutions, i.e., it also finds solutions which do not belong to the convex hull. In section 5. we present results of the proposed algorithms and in section 6. we summarize the paper.

2. PROBLEM FORMULATION

The main idea behind the proposed approach is to approximate the boundary by a polygon, and to encode the polygons vertices instead of the original boundary. Since we assume that the original boundary is represented with pixel accuracy, it can be losslessly encoded by an 8-connect chain-code. We propose to approximate the boundary with a low order polygon which can be encoded efficiently and also reduces some of the noise along the object boundary which is a result of the segmentation algorithm.

The following notation will be used. Let $B = \{b_0, \ldots, b_{N_B-1}\}$ denote the connected boundary which is an ordered set, where b_j is the j-th point of B and N_B is the total number of points in B. Note that in the case of a closed boundary, $b_0 = b_{N_B-1}$. Let $P = \{p_0, \ldots, p_{N_P-1}\}$ denote the polygon used to approximate B which is an ordered set, where p_k is the k-th vertex of P, N_P is the total number of vertices in P and the k-th edge starts at p_{k-1} and ends at p_k. Since P is an ordered set, the ordering rule and the set of vertices uniquely define the polygon.

We assume that the vertices of the polygon are encoded differentially which is for example the case when a single ring chain code or a generalized chain code [4] is used. In other words, only the difference between two consecutive vertices is encoded. We denote by $r(p_{k-1}, p_k)$ the required bit rate for that, hence the bit rate $R(p_0, \ldots, p_{N_P-1})$ for the entire polygon is,

$$R(p_0, \ldots, p_{N_P-1}) = \sum_{k=0}^{N_P-1} r(p_{k-1}, p_k), \tag{1}$$

where $r(p_{-1}, p_0)$ is set equal to the number of bits needed to encode the absolute position of the first vertex. In the case of a closed boundary, i.e., the first vertex is identical to the last one, the rate $r(p_{N_P-2}, p_{N_P-1})$ is set to zero since the last vertex does not need to be encoded. Note that this rate depends on the specific vertex encoding scheme and we presented one such scheme in [10].

In general the polygon which is used to approximate the boundary should be permitted to place its vertices anywhere in the plane. In this paper, we restrict the vertices to belong to the original boundary ($p_k \in B$), so that we can employ a fast polygon selection algorithm. This restriction results in the following fact, which we employ to derive low complexity optimization algorithms.

The k-th polygon edge which connects two consecutive vertices p_{k-1} and p_k is an approximation to the partial boundary $\{b_j = p_{k-1}, b_{j+1}, \ldots, b_{j+l} = p_k\}$ which contains $l+1$ boundary points. Therefore, we can measure the quality of this approximation by an edge distortion measure which we denote by $d(p_{k-1}, p_k)$. There are several different distortion measures which can be employed. In this paper we focus on distortion measures which are additive, in other words, the polygon distortion $D(p_0, \ldots, p_{N_P-1})$ equals the sum of the edge distortions,

$$D(p_0, \ldots, p_{N_P-1}) = \sum_{k=0}^{N_P-1} d(p_{k-1}, p_k), \tag{2}$$

where $d(p_{-1}, p_0)$ is set equal to zero. Two examples of additive distortion measures are the absolute area between the polygon and the boundary and the sum of the squared distance from the boundary to the polygon. In [10] we presented a scheme where the maximum absolute distance between a boundary point and the approximation is used as the distortion measure. Note that this measure is not additive and therefore a different set of algorithms has been proposed in [10] for this problem.

In the reminder of the paper we introduce two algorithms which solve the following constrained optimization problem,

$$\min D(p_0, \ldots, p_{N_P-1}), \quad \text{s.t.:} \quad R(p_0, \ldots, p_{N_P-1}) \leq R_{max}, \tag{3}$$

where R_{max} is the maximum bit rate permitted for the encoding of the boundary. Note that both algorithms are symmetric in the rate and the distortion and therefore they can also be used to solve the dual problem,

$$\min R(p_0, \ldots, p_{N_P-1}), \quad \text{s.t.:} \quad D(p_0, \ldots, p_{N_P-1}) \leq D_{max}, \tag{4}$$

where D_{max} is the maximum distortion permitted.

3. LAGRANGIAN RELAXATION APPROACH

In this section we derive a solution to problem (3) which is based on Lagrangian relaxation and the shortest path algorithm presented in [10]. We first define the Lagrangian cost function

$$J_\lambda(p_0, \ldots, p_{N_P-1}) = D(p_0, \ldots, p_{N_P-1}) + \lambda * R(p_0, \ldots, p_{N_P-1}), \tag{5}$$

where λ is called the Lagrangian multiplier.

It has been shown in [11, 12] that if there is a λ^* such that,

$$\{p_0^*, \ldots, p_{N_P-1}^*\} = \arg\min J_{\lambda^*}(p_0, \ldots, p_{N_P-1}), \tag{6}$$

and which leads to $R(p_0, \ldots, p_{N_P-1}) = R_{max}$, then $\{p_0^*, \ldots, p_{N_P-1}^*\}$ is also an optimal solution to (3). It is well known that when λ sweeps from zero to infinity, the solution to problem (6) traces out the convex hull of the operational rate distortion curve, which is a non-increasing function. Hence bisection [13] or the fast convex search we presented in [14] can be used to find λ^*.

Therefore the problem at hand is to find the optimal solution to problem (6). For a given λ, the weight function $w(u,v) : E \rightarrow \mathcal{R}$, which is used in the shortest path algorithm presented in [10], needs to be redefined as follows,

$$w(u,v) = d(u,v) + \lambda * r(u,v), \tag{7}$$

where $E = \{(b_i, b_j) \in B^2 : i < j\}$ is the edge set of the underlying graph. Since the shortest path algorithm results in the polygon which minimizes the following sum,

$$\sum_{k=0}^{N_P-1} w(p_{k-1}, p_k), \tag{8}$$

this polygon is the optimal solution to the relaxed problem of Eq. (6).

The time complexity of the Lagrangian approach strongly depends on the employed distortion measure. If the sum of the squared distance is used, then the time complexity is $\Theta(N_B^3)$ [10]. The above proposed approach finds the optimal solutions which constitute the vertices of the convex hull. Clearly there are other optimal solutions which are above the convex hull and in the next section we present a pruning algorithm which finds all the optimal solutions.

4. PRUNING APPROACH

The second algorithm for solving the problem of Eq. (3) is based on a tree pruning approach which finds the rate distortion points of all optimal solutions. When the current vertex is selected , the previous vertices do not influence the selection of the future vertices. This is the basic concept behind the shortest path algorithm proposed in [10] which we also employed for the Lagrangian relaxation approach proposed in the previous section.

For a given boundary point under consideration for a vertex, every previous boundary point could have been the last vertex used for the polygon approximation. Therefore, the rate and distortion for every previous boundary point are calculated under the assumption that the previous boundary point was used as the previous vertex. These calculations lead to a set of nodes, each representing the hypothesis that the current boundary point is a vertex but with different boundary points as previous vertices.

We can introduce a pruning procedure to reduce the number of nodes for each boundary point. If there are two nodes j and i such that $D(j) \geq D(i)$ and $R(j) \geq R(i)$, where $D(n)$ is the distortion and $R(n)$ is the rate up to and including node n, then it is clear that node j cannot belong to the optimal solution. This follows since node i has a lower distortion and a lower rate than node j, but both represents the same boundary point as selected vertex. Hence node j is pruned from the decision tree. Since the pruned nodes need not be considered in the future of the optimization process, the more nodes pruned, the faster the algorithm will perform. A straightforward implementation of the pruning operation has a quadratic time complexity in the number of nodes to be pruned. Since the number of nodes to be pruned depends on previous pruning results, the time complexity of the entire approach depends on the boundary, the distortion measure and the vertex encoding scheme. As with most integer programming algorithms, one can construct an example where the pruning scheme fails completely, which results in an exponential time complexity. In general though, the pruning is extremely efficient in cutting down the complexity of the algorithm and in fact, this scheme and the previously discussed Lagrangian approach take about the same amount of time for the experimental results we will present in section 5.. If the complete set of optimal rate distortion points is not of interest, but only the problem of Eq. (3) needs to be solved, additional pruning can be achieved by removing all nodes which contain a rate higher than R_{max}. This leads effectively to all optimal rate distortion points below and including the line $R(D) = R_{max}$.

Each of the remaining nodes represents the current boundary point, but with different rate-distortion characteristics. In other words, the remaining nodes represent the set of all optimal solutions for the encoding of the boundary up to and including the current boundary point. These nodes make up the admissible nodes for this boundary point, when this boundary point is considered as a previous vertex in the future of the optimization process.

Fig. 1 shows a simple example of the algorithm. In the left upper corner is the boundary which must be encoded. The adjacent pixels are labeled 0, 1, 2, 3 and 4 and they simply form a square of sidelength 1. Note that point 4 is the same as point 0, and therefore it does not need to be transmitted, but still a distortion occurs between the last vertex of the polygon and point 4 and it is therefore included in the closed boundary. Fig. 1 shows the complete decision tree. This tree reflects the fact that given the boundary point used as the previous vertex, the rate and the distortion for encoding the polygon up to and including that vertex, the selection of future vertices is independent of the selection made for previous vertices. In Fig. 1 the boundary point index, the rate and the distortion are indicated in the following fashion: "index/rate/distortion". In this example, the sum of the squared distance between the boundary and the polygon is used as the distortion measure and the vertex encoding scheme proposed in [10] is employed.

There are two possible transitions from a given node. The upward transition, which indicates that this node is used as a previous vertex and the downward transition which indicates that this node is not used as a previous vertex. The downward transitions carry a weight of zero, but the upward transitions result in the addition of r (previous vertex, current vertex) to the rate and d (previous vertex, current vertex) to the distortion. Note that the employed vertex encoding scheme requires 4 bits for each transition in this example. The epochs, which correspond to the boundary points, are indicated at the bottom of the tree. The boxes indicate the new nodes per boundary point and therefore the pruning procedure is only applied to those nodes.

Consider the box at epoch 3. There are two nodes, (both with description 3/8/0.5) which require the same rate and lead to the same distortion to reach boundary point 3. Therefore one of the two can be pruned (indicated by an empty circle) since both will lead to the same collection of future paths. By pruning a node, the collection of future paths gets reduced. Clearly, the more nodes that can be pruned, the faster the algorithm is.

At the last epoch, boundary point 4, 3 nodes can be pruned and only 4 final nodes remain which represent the 4 optimal solutions to the boundary approximation. These 4 optimal solutions are also displayed as an operational rate distortion curve in the lower left corner of Fig. 1. The path (0,1,2,3,4) which leads to 12 bits and no distortion is the highest quality approximation which is basically the chain code of the original boundary. The path (0,1,2,2,4) approximates the box by a triangle which requires 8 bits and leads to a distortion of 0.5. The path (0,0,2,2,4) approximates the box by a diagonal line which requires 4 bits and leads to a distortion of 1. Finally the path (0,0,0,0,4) does not require any bits, since it approximates the box by its starting point, but it leads to a maximum distortion of 4.

5. EXPERIMENTAL RESULTS

In this section we present results of the proposed algorithms using an object from the "Miss America" sequence which is shown in Fig. 2. The mean squared distance is the distortion measure and the vertex encoding scheme is the one presented in [10]. To highlight the difference between the two approaches, we set R_{max} from Eq. (3) to 200 bits. In Fig. 3 the operational rate distortion curve is displayed and the possible Lagrangian solutions are indicated. The Lagrangian solution which satisfies $R \leq 200$ bits results in $R = 169$ bits and $D = 0.1$, (for a $\lambda^* = 0.002$), whereas the pruning approach results in $R = 200$ bits and $D = 0.05$. Both solutions are shown in Fig. 2 and it is clear that the pruning approach results in a better approximation of the original boundary.

6. SUMMARY

We presented two different approaches for an optimal segmentation encoding scheme in the rate distortion sense for additive distortion measures. The main advantage of these schemes is the complete control of the optimal trade-off between the rate and the distortion. The first approach is based on Lagrangian relaxation and a shortest path algorithm whereas the second approach is based on a tree-pruning scheme. We compared the proposed approaches experimentally using an object from the "Miss America" sequence.

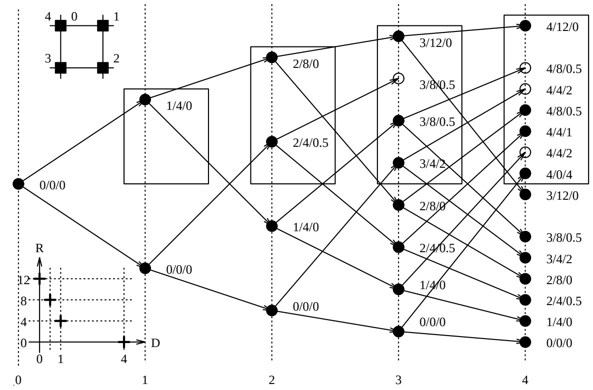

Figure 1. Pruned decision tree

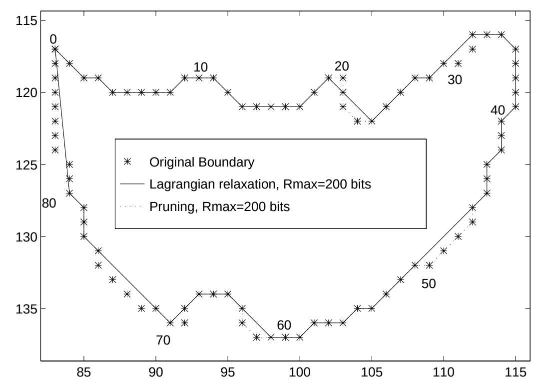

Figure 2. Lagrangian relaxation vs. pruning

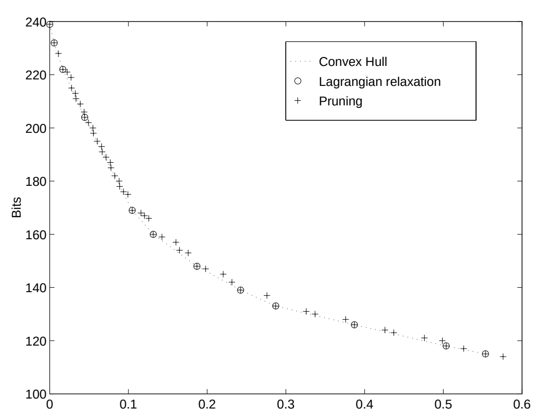

Figure 3. Operational rate distortion curve

REFERENCES

[1] H. Musmann, M. Hötter, and J. Ostermann, "Object-oriented analysis-synthesis coding of moving images," *Signal Processing: Image Communication*, vol. 1, pp. 117–138, Oct. 1989.

[2] A. K. Jain, *Fundamentals of Image Processing*. Prentice-Hall, 1989.

[3] H. Freeman, "On the encoding of arbitrary geometric configurations," *IRE Trans. Electron. Comput.*, vol. EC-10, pp. 260–268, June 1961.

[4] J. Saghri and H. Freeman, "Analysis of the precision of generalized chain codes for the representation of planar curves," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-3, pp. 533–539, Sept. 1981.

[5] J. Koplowitz, "On the performance of chain codes for quantization of line drawings," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-3, pp. 180–185, Mar. 1981.

[6] D. Neuhoff and K. Castor, "A rate and distortion analysis of chain codes for line drawings," *IEEE Transactions on Information Theory*, vol. IT-31, pp. 53–68, Jan. 1985.

[7] T. Kaneko and M. Okudaira, "Encoding of arbitary curves based on the chain code representation," *IEEE Transations on Communications*, vol. COM-33, pp. 697–707, July 1985.

[8] R. Prasad, J. W. Vieveen, J. H. Bons, and J. C. Arnbak, "Relative vector probabilities in differential chain coded line-drawings," in *Proc. IEEE Pacific Rim Conference on Communication, Computers and Signal Processing*, (Victoria, Canada), pp. 138–142, June 1989.

[9] T. Özçelik, *A Very Low Bit Rate Video Codec*. PhD thesis, Dept. EECS, Northwestern University, Dec. 1994.

[10] G. M. Schuster and A. K. Katsaggelos, "An optimal lossy segmentation scheme," in *Proc. VCIP 96*, Mar. 1996.

[11] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 36, pp. 1445–1453, Sept. 1988.

[12] H. Everett, "Generalized Lagrange multiplier method for solving problems of optimum allocation of resources," *Operations Research*, vol. 11, pp. 399–417, 1963.

[13] C. F. Gerald and P. O. Wheatley, *Applied Numerical Analysis*. Addison Wesley, fourth ed., 1990.

[14] G. M. Schuster and A. K. Katsaggelos, "Fast and efficient mode and quantizer selection in the rate distortion sense for H.263," in *Proc. VCIP 96*, Mar. 1996.