

# OPTIMIZATION OF WIREFRAME MODEL ADAPTATION AND MOTION ESTIMATION IN A RATE-DISTORTION FRAMEWORK

Dimitrios Tzovaras, Stavros Vachtsevanos and Michael G. Strintzis

Electrical and Computer Engineering Department  
Information Processing Laboratory  
Aristotle University of Thessaloniki  
Thessaloniki 54006, Greece  
phone: (+30-31) 996-359, FAX: (+30-31) 996-398  
e-mail : tzovaras@dion.ee.auth.gr

## ABSTRACT

A rate-distortion framework is used to define a very low bit rate coding scheme based on wireframe model adaptation and optimized selection of motion estimators. This technique achieves maximum reconstructed image quality under the constraint of a target bitrate for the coding of the motion vector field, the wireframe representation information and the prediction error. Experimental results illustrating the performance of the proposed techniques in very low bit rate image sequence coding application areas are presented and evaluated.

## 1 INTRODUCTION

A great deal of attention has recently been directed to the area of very low bitrate image sequence coding. This is mainly due to the maturity of new applications such as videophone services, interactive multimedia databases, remote sensing and many others. The key requirement for these applications is lower demand for channel capacity for transmission or storage, in order to use the already existing communication networks.

The transmission of full motion video through limited capacity channels is critically dependent on the ability of the compression schemes to achieve target bit rates while still maintaining acceptable visual quality. In order to achieve this, motion estimation and motion compensated prediction are frequently used, so as to reduce temporal redundancies in image sequences. Block-based motion estimation techniques have been extensively studied and applied for very low bitrate coding. However, the performance of these techniques in such low bitrates is restricted by well known limitations such as the block and mosquito artifacts. More sophisticated techniques are required to maintain the distortion below the level of annoyance.

In most model-based video compression algorithms, it is desirable to select both the adaptation of a wireframe to an image (which segments an image into objects) and the motion estimation method for each object, adaptively so as to best represent the motion of each part of the image. For example, in a typical videophone sequence, rough object subdivision combined with a trans-

lational or affine 2D motion estimation model would suffice for the description of the motion of most parts of the foreground object while much finer object subdivision, perhaps down to the size of very small triangles and more sophisticated 3D motion models would be best suited for the description of mouth and eye motion.

An elegant framework for the definition of such a strategy is provided by the classical rate-distortion (R-D) constrained minimization procedure. Recently, rate-distortion optimization was used for the development of efficient motion and disparity estimation strategies [1]. The present paper, extending the results of [2], investigates the use of this methodology for wireframe model adaptation and hybrid motion field estimation under the constraint of a target bitrate for the coding of the vector and wireframe information. Wireframe model adaptation is performed using rate-distortion criteria and is fused with motion estimation by selecting for each triangle of the wireframe, the optimum motion estimator from a predetermined set of candidate motion estimators.

## 2 WIREFRAME MODEL ADAPTATION

For the generation of the 2D model object, the scene information is modeled using a wire mesh. We shall generate a surface model of the form

$$\mathbf{P} = S(\hat{\mathbf{P}}) , \quad (1)$$

where  $\hat{\mathbf{P}}$  is a set of 2D “control” points  $\hat{\mathbf{P}} = \{(x_i, y_i), i = 1, \dots, N\}$  that determine the shape of the surface and  $\mathbf{P} = (x, y)$  is a 2D point in the image plane. An initial choice for  $\hat{\mathbf{P}}$  is the regular tessellation.

The wireframe model is then assumed to consist of sub-objects (triangles) which exhibit different motion. The wireframe information is assumed to be completely described by a quadtree, whose nodes indicate if a triangle is split or not. The refinement of the object segmentation procedure (i.e. the wireframe adaptation) is performed by splitting the triangles, based on a rate-distortion optimization procedure that will be described in detail in the following section. Splitting is based on

the procedure illustrated in Figure 1. For the estimation of the object motion a hybrid motion scheme is proposed which is also optimized based on rate-distortion optimization criteria.

### 3 HYBRID MOTION ESTIMATION

Modeling of the motion information by translation, zoom and pan, or a 3D rotation and translation, has been used in translational, affine and 3D motion estimators. Experimental results have shown that affine 2D motion or 3D motion models may represent efficiently the displacement occurring in typical scenes ; however most parts of the image may be coded very satisfactorily using only translational motion (e.g. the background). Moreover, the complexity of the affine and 3D motion estimation algorithms is higher than the complexity of the translational scheme.

Based on the above observations, we propose the use of all these models for the motion compensated coding of the objects of a scene, within a rate-distortion framework optimizing both the wireframe adaptation and the motion estimation. The alternatives are :

1. The motion of the object is insignificant. No motion vector is transmitted and the previous estimate for this frame is considered sufficient.
2. Translational motion is used to compensate the motion of an object. A two-component motion vector is transmitted.
3. An affine 2D motion model is used to represent the motion of an object. The six model parameters are transmitted.
4. A 3D motion model represents best the motion of an object in the scene. The eight motion model parameters are transmitted.

### 4 WIREFRAME MODEL ADAPTATION USING RATE DISTORTION OPTIMIZATION

Let a wireframe  $\mathcal{W}$  adapted to the image plane consisting of  $N_s$  triangles  $s_i : \mathcal{W} = \{s_i, i = 1, \dots, N_s\}$ . Each triangle is assumed to be coded using one of the motion estimation techniques described in the previous section. For each candidate motion estimator  $M_j \in \mathcal{M}$ ,  $\mathcal{M} = \{M_j, j = 1, \dots, M\}$ , let  $\mathcal{V}^{(j)} = \{\mathbf{v}_i^{(j)}, i = 1, \dots, N_s\}$  for  $j = 1, \dots, M$ , be the corresponding set of triangle motion vectors. The general joint vector field estimation and wireframe adaptation algorithm aims to minimize the distortion  $D$  of the reconstructed image sequence, under a constraint  $R_{budget}$  on the rate for the transmission of the vector field and the corresponding wireframe information. This corresponds to the following

constrained optimization problem :

$$\min_{\{\mathbf{v}_i^{(j)} \in \mathcal{V}^{(j)}, j=1, \dots, M \ \& \ s_i \in \mathcal{W}\}} \sum_{i=1}^{N_s} D(\mathbf{v}_i, s_i), \quad (2)$$

subject to

$$\sum_{i=1}^{N_s} R(\mathbf{v}_i, s_i) \leq R_{budget},$$

where  $N_s$  is the total number of objects (triangles) in the image,  $D(\mathbf{v}_i, s_i)$  is the contribution of the decision  $(\mathbf{v}_i, s_i)$  to the distortion function and  $R(\mathbf{v}_i, s_i)$  is the contribution of the same to the total rate or cost of the transmission of the motion vectors and the wireframe information.

The methodology in [3] permits the transformation of the above into an unconstrained optimization problem. In fact, as shown in [3] (the proof is also contained in [4]), the solution  $\{\mathbf{v}_i^*(\lambda), s_i^*(\lambda), i = 1, \dots, N_s\}$  of the problem of unconstrained minimization of

$$J(\lambda) = \sum_{i=1}^{N_s} D(\mathbf{v}_i(\lambda), s_i(\lambda)) + \lambda \sum_{i=1}^{N_s} R(\mathbf{v}_i(\lambda), s_i(\lambda)), \quad (3)$$

is also a solution of (2) if

$$R_{budget} = \sum_{i=1}^{N_s} R(\mathbf{v}_i^*(\lambda), s_i^*(\lambda)). \quad (4)$$

The problem therefore, reduces to ensuring that (4) has a solution for  $\{(\mathbf{v}_i(\lambda), s_i(\lambda), i = 1, \dots, N_s)\}$  and determining this solution. This was investigated from a general viewpoint in [4], where it was shown that  $R(\mathbf{v}(\lambda), s(\lambda))$  and  $D(\mathbf{v}(\lambda), s(\lambda))$  are monotonic functions of the Lagrange multiplier  $\lambda$ , which may be interpreted as a quality index, with values ranging from 0 (highest rate, lowest distortion) to  $\infty$  (lowest rate, highest distortion). Further investigation in [5] proved that the solution of (4) may be obtained using any fast convex algorithm such as the bisection algorithm. One such algorithm, which gave very good results in [5] is also adopted in the present paper. The overall procedure for the determination of the optimal wireframe model and motion estimator consists of the following steps.

#### 4.1 Initialization

In the present work the wireframe model  $\mathcal{W}$  is assumed to be completely described by a quadtree  $T = \{n_j^i, i = 0, \dots, \log(K), j = 1, \dots, 2^i\}$ , (i.e.  $n_j^i$  is the  $j$ th node at scale  $i$ ) where  $K \times K$  is the dimension of the image. Using the quadtree information the corresponding wireframe model can be created. Associated with each node  $n_j^i$  of the quadtree  $T$  is a data structure of the form :  $\{\tilde{R}_j^i, \tilde{D}_j^i, \tilde{J}_j^i, split(n_j^i)\}$  where  $\tilde{R}_j^i$  is the entropy associated with the current node,  $\tilde{D}_j^i$  is the distortion corresponding to the reconstruction of the triangle associated

with the current quadtree node,  $\tilde{J}_j^i$  is the corresponding Lagrangian cost and  $split(n_j^i)$  is a bit indicating if the current node is split or not.

The proposed joint motion estimation and wireframe adaptation algorithm begins by gathering the motion estimator set dependent  $(R_m(n), D_m(n))$  values for each node  $n \in T$ , and for each  $m \in \mathcal{M}(n)$ , to generate the  $R$  versus  $D$  values for each node. In the following, the two stages of the rate-distortion optimization algorithm are presented i.e. (a) finding the optimal solution for a given operating slope  $\lambda$  and (b) determining the optimal slope  $\lambda$ . Stage (a) of the algorithm is run for a given slope value  $\lambda$  and could be considered a subroutine called by stage (b) of the algorithm.

#### 4.2 Finding the optimal solution for a given operating slope $\lambda$

For the current  $\lambda$ , populate all the nodes  $n_j^i$  of the tree with their minimum Lagrangian costs  $J_j^i(\lambda)$ , i.e.:

$$J_j^i(\lambda) = \min_{m \in \mathcal{M}} [D_m(\lambda) + \lambda R_m(\lambda)].$$

The cost is minimized with respect to the choice of motion estimation for node  $n_j^i$  using the technique in [2]. According to this technique the procedure is initialized by considering the optimal motion estimator for each possible object, for the specific value of  $\lambda$ . The segmentation algorithm starts then by considering the whole image as a node of the quadtree which is split in four if

$$J_j^i(\lambda) > \tilde{J}_{2j}^{i+1}(\lambda) + \tilde{J}_{2j-1}^{i+1}(\lambda) + \tilde{J}_{2j-2}^{i+1}(\lambda) + \tilde{J}_{2j-3}^{i+1}(\lambda) \quad (5)$$

The process is repeated until no further splitting of the triangles is selected.

#### 4.3 Determining the optimal slope $\lambda$

First two values  $\lambda_l < \lambda_u$  of  $\lambda$  are found so that

$$\sum_{i=1}^{N_s} R(\mathbf{v}_i(\lambda_u), s_i(\lambda_u)) \leq R_{budget} \leq \sum_{i=1}^{N_s} R(\mathbf{v}_i(\lambda_l), s_i(\lambda_l)).$$

Note that the initial wireframe model  $\mathcal{W}(\lambda_u)$  of the first frame of the image sequence, selects the whole image to be a single object, while for the subsequent frames  $\mathcal{W}(\lambda_u)$  is the optimal wireframe model corresponding to the previous in time frame. Similarly,  $\mathcal{W}(\lambda_l)$  corresponds to the wireframe model resulting by full splitting of the wireframe until the minimum allowed object (triangle) size is reached.

For the coding of a sequence of frames, the values of  $\lambda_l, \lambda_u$  are chosen to be  $\lambda_l = 0, \lambda_u = \infty$  for the initial frame and  $\lambda_l = 0.8\lambda^*, \lambda_u = 1.2\lambda^*$  for subsequent frames, where  $\lambda^*$  is the solution of (4) for the previous frame. The bracketing interval is then successively decreased in size by applying the bisection algorithm.

## 5 APPLICATION TO VERY LOW BIT RATE IMAGE SEQUENCE CODING

In order to evaluate the performance of the proposed approach for very low bitrate coding, the algorithm was applied to the typical QCIF sequences ‘‘Salesman’’ and ‘‘Foreman’’. The frame rate of the sequences was 10 *frames/sec*.

For very low bitrate video coding applications the proposed algorithm was applied to groups of pictures (GOP), with the bit allocation assigned adaptively to each frame of the sequence in order to optimize the transmission bitstream. More specifically, the total bitrate is given for the coding of the whole GOP and the rate is allocated to each frame according to the frame difference between the current frame and the preceding in time frame. Thus, if  $R_{target}$  is the target bitrate for the coding of the whole GOP and

$$FD_i = \sum_{k_1}^{N_1} \sum_{k_2}^{N_2} |im_i(k_1, k_2) - im_{i-1}(k_1, k_2)|$$

is the frame difference between frames  $i$  and  $i - 1$ , and  $N_1, N_2$  are the image dimensions, the bitrate is allocated as follows

$$R_{budget}^{(i)} = R_{target} \frac{FD_i}{\sum_{i=1}^{10} FD_i}.$$

Then for each frame, a percentage of the total budget  $R_{budget}^{(i)}$  (60 % in our experiments) is allocated to the coding of the motion vector and wireframe information, and the remaining rate is allocated to the coding of the prediction error image.

The proposed coder which includes optimization of the wireframe representation, the motion estimation and the error transmission was then applied for the coding of the ‘‘Salesman’’ and ‘‘Foreman’’ QCIF sequences. Two target bit rates were tested : 28.8 *Kbits/sec* and 64 *Kbits/sec*. Figures 2 and 3 compare the performance of the proposed scheme with the technique in [2] in terms of PSNR versus bitrate for the coding of the first fifty frames of ‘‘Salesman’’ and ‘‘Foreman’’. The technique in [2] performs rate-distortion optimization of a block-based segmentation and the corresponding motion estimation for each object. As seen, the proposed technique overperforms slightly the technique in [2] in terms of PSNR versus frame number for the same bitrate. Note that the coder performance remains stable providing a high quality image without the intervention of a new intra-coded image before the end of the whole sequence.

## 6 CONCLUSIONS

A rate-distortion framework was used to define a very low bit rate coding scheme based on wireframe model adaptation and optimized selection of motion estimation and transmission of the prediction error. This technique achieved maximum reconstructed image quality

under the constraint of a target bitrate for the coding of the vector and wireframe model description information. Experimental results in application to the coding of typical videophone sequences have demonstrated the excellent performance of the proposed very low bit-rate video coding scheme.

## 7 ACKNOWLEDGEMENT

This work was supported by the EU CEC Projects PANORAMA (ACTS project 092) and VIDAS (ACTS project 057).

## References

- [1] D. Tzovaras and M. G. Strintzis, "Motion and Disparity Estimation Using Rate Distortion Theory for Very Low Bit Rate and Multiview Image Sequence Coding," in *VCIP '97*, San Jose, Feb. 1997.
- [2] D. Tzovaras, S. Vachtsevanos, and M. G. Strintzis, "Optimization of Quadtree Segmentation and Hybrid 2D and 3D Motion Estimation in a Rate-Distortion Framework," *IEEE Journal on Selected Areas in Communications, Special issue on Very Low Bit-rate Coding*, vol. 15, no. 9, pp. 1726-1738, Dec. 1997.
- [3] H. Everett, "Generalized Lagrange Multiplier Method for Solving Problems of Optimum Allocation of Resources," *Operation Res.*, vol. 11, pp. 399-417, 1963.
- [4] Y. Shoham and A. Gersho, "Efficient Bit Allocation for an Arbitrary Set of Quantizers," *IEEE Trans. on Acoust., Speech, Signal Processing*, vol. 36, pp. 1445-1453, Sep. 1988.
- [5] K. Ramchandran and M. Vetterli, "Best Wavelet Packet Bases in a Rate-Distortion Sense," *IEEE Trans. on Image Processing*, vol. 2, pp. 160-175, Apr. 1993.

| Method        | Percentage<br>of Selection (%) |
|---------------|--------------------------------|
| No motion     | 38.57                          |
| Translational | 47.26                          |
| Affine 2D     | 10.43                          |
| 3D motion     | 3.74                           |

Table 1: Average percentage of selection of the candidate motion estimation methods used for the coding of "Salesman", "Foreman" and "Tunnel".

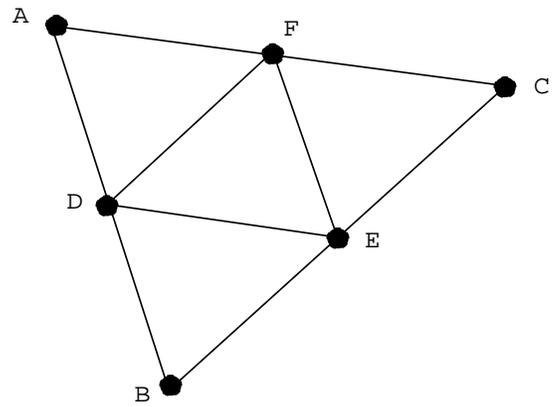


Figure 1: The method used for splitting triangle ABC to triangles ADF, BDE, CFE and DEF.

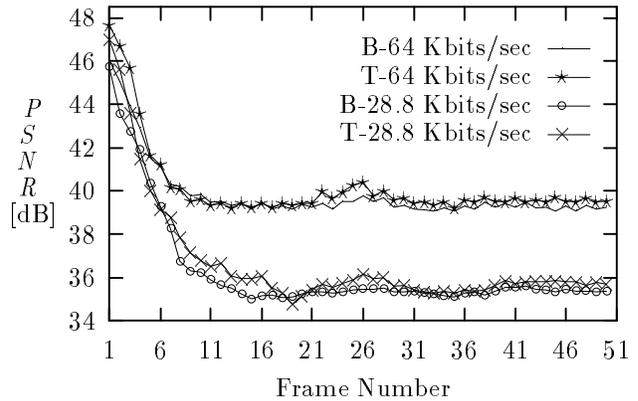


Figure 2: Comparison of the performance of the proposed rate-distortion optimized hybrid coder (indicated with the prefix T-) with the block-based technique described in [2] (indicated with the prefix B-) for the coding of "Salesman" at 64 Kbits/sec and 28.8 Kbits/sec.

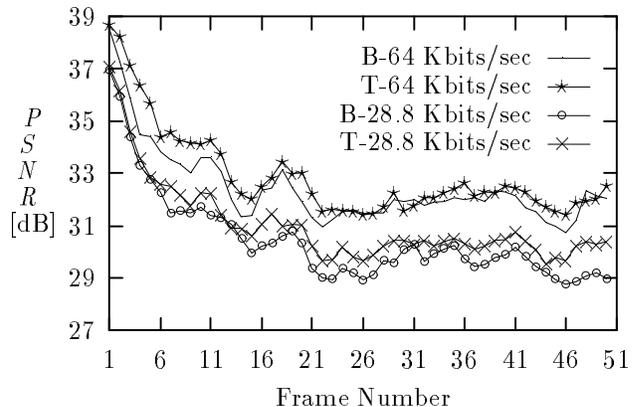


Figure 3: Comparison of the performance of the proposed rate-distortion optimized hybrid coder (indicated with the prefix T- in the diagram) with the block-based technique described in [2] (indicated with the prefix B-) for the coding of "Foreman" at 64 Kbits/sec and 28.8 Kbits/sec.