Investigations on an Object–Based Layered Coding System

Michael Wollborn
Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung
Universität Hannover, Appelstraße 9A, D–30167 Hannover 1, FRG
wollborn@tnt.uni–hannover.de

ABSTRACT: *An object–based layered coder for the transmission of image sequences at very low bit rates between 8 and 64 kbit/s is presented. In layer 1, only colour parameters are transmitted, whereas in layer 2 additionally motion parameters are transmitted in order to apply motion compensated prediction. In layer 3, in addition to colour and motion parameters, shape parameters of arbitrarily shaped moving objects are transmitted, since the consideration of the object shape can improve motion compensation at moving object boundaries. For each frame, the most efficient layer is chosen by an appropriate layer selection. Compared to a block–based coder according to the ITU–T Rec. H.263, simulation results show that the bit rate for coding of colour parameters can be reduced by about 5% at the same image quality.*

1. INTRODUCTION

For the coding and transmission of image sequences at low and very low bit rates between 8 and 64 kbit/s, a hybrid coding concept has been standardized in the ITU–T Rec. H.263 [5]. Here, each image of an image sequence is subdivided into square blocks of N x N picture elements (pels). Each block is coded seperately, using 2D motion compensated prediction and DCT based transform coding. Since the motion of a block is only described by one displacement vector and due to the blockbased processing, motion compensation will cause large prediction errors especially in boundary blocks of moving objects, i.e. blocks which contain parts of a moving object as well as static background. This leads either to high bit rates for prediction error coding or to visible artifacts in these areas.

To overcome this problem, the shape of moving objects should be considered for motion estimation and compensation. Therefore, coding techniques based on the concept of object–based analysis–synthesis coding have been proposed [6][4][8][9]. Here, each image is subdivided into arbitrarily shaped moving objects, where each object is described by its colour, motion and shape parameters. The parameter sets are automatically estimated by an image analysis. Compared to the blockbased hybrid coding, in addition to colour and motion parameters now also shape parameters have to be coded and transmitted to the receiver. The use of the shape is only efficient when the additional required bit rate for coding of the shape parameters is overcompensated by the bit rate saved in colour parameter coding. Since this condition is not always complied, object–based coding may be sometimes less efficient than the blockbased approach. This drawback can be overcome when the concept of layered coding [7] is used: Starting from a bottom layer only transmitting colour parameters, in each higher layer additional parameters are coded and transmitted. The layers are combined in such a way, that a higher layer is only used if it is more efficient than the lower layer, using an appropriate layer selector.

In this paper, an object–based layered coding system for the transmission of image sequences at very low bit rates between 8 and 64 kbit/s is proposed. The system consists of three layers: In layer 1, only colour parameters of the image are transmitted. In layer 2, in addition to the colour parameters, motion parameters are transmitted in order to apply motion compensated prediction. In layer 3, in addition to colour and motion parameters, shape parameters of arbitrarily shaped moving objects are transmitted. The consideration of the object shape is used to improve motion estimation and compensation at the boundaries of moving objects. A layer selector is used to choose always the most efficient layer, taking into account the respective additional bit rate as well as the gain due to the use of the additional parameters. In order to demonstrate the efficiency of the proposed coder, it is compared with a standardized blockbased hybrid coder according to ITU–T Rec. H.263.

The paper is organized as follows: In Section 2, the implementation of the three coder layers is described. In Section 3, the selection of the appropriate layer is explained. In Section 4, experimental results are given and Section 5 concludes the paper.

2. DESCRIPTION OF CODER LAYERS

In the presented approach, layers 1 and 2 are implemented using a block–based hybrid coder according to the ITU–T Rec. H.263 [5]. This coder contains an automatic switching between layer 1 and layer 2 on a macroblock (16x16 pels) basis. In layer 3, a simplified version of an object–based analysis–synthesis coder based on a 2D flexible object model [1][3][4] is used. Therefore, besides the motion parameters, the shape of moving objects is estimated by image analysis and described by its 2D silhouette. The shape parameters are predictively coded [1], applying a polygon/spline approximation for the object shape [2]. Motion compensated prediction is performed in the same way as in layer 2, except for blocks at object boundaries. Here, for motion estimation the object shape is taken into account.

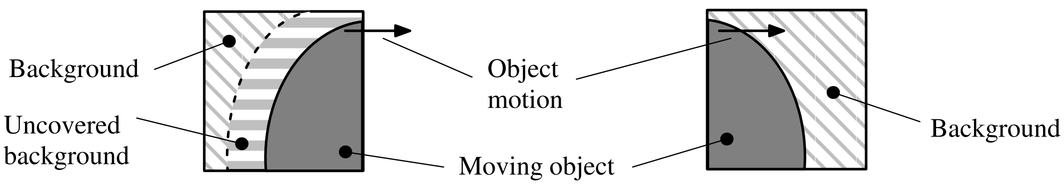

Figure 1 Subdivision of boundary blocks for prediction of background, uncovered background and object texture in layer 3

For motion compensation, the boundary blocks are subdivided into one part corresponding to the background and another part corresponding to the moving object, as can be seen in Fig. 1. Then, background parts are predicted from the preceding frame assuming zero motion. The areas inside the object as well as the uncovered background areas are predicted motion compensated, applying the estimated motion parameters of the moving object.

The coding and transmission of the prediction error is performed the same way as in layer 2, i.e. 8x8 DCT and Huffmann coding of the quantized transform coefficients is applied.

3. LAYER SELECTION

As described in Section 2, the layers 1 and 2 have been implemented according to the ITU–T Rec. H.263. This coder already contains a criterion for automatic switching between layer 1 and layer 2 on a macroblock (16x16 pels) basis. Therefore, for each macroblock that layer is chosen which yields the lower mean energy for the signal to code (which is for layer 1 colour values and for layer 2 prediction error values), taking into account the bitrate needed for motion vector coding in layer 2.

Since layer 3 should only be used, when the saved bit rate exceeds the amount of bits required for transmission of shape parameters, an appropriate criterion for switching between layer 2 and layer 3 has been developed. Because the coding of shape parameters is only efficient for the whole frame, the switching is performed on frame by frame basis. The only difference between layer 2 and layer 3 is the coding of boundary blocks. Assuming that the bit rate for motion parameters is identical in both layers, only the remaining bit rate (i.e. for colour parameters in layer 2 and for colour and shape parameters in layer 3) for coding of boundary blocks is compared for each frame. Layer 2 requires a bit rate of R_{col}^{l2} for colour parameters of boundary blocks. In layer 3, a bit rate of R_{col}^{l3} for colour parameters and an additional bit rate of R_{sha}^{l3} for shape parameters of boundary blocks is needed. The decision between layer 2 and layer 3 is taken as follows:

$$(R_{col}^{l3} + R_{sha}^{l3}) - R_{col}^{l2} \quad \begin{cases} > 0 & \rightarrow \mathit{layer} 2 \\ \leq 0 & \rightarrow \mathit{layer} 3 \end{cases}$$

4. EXPERIMENTAL RESULTS

Investigations of the proposed scheme using different test sequences showed, that the use of shape information for motion compensation is only efficient in cases where the background is at least partially structured. In the case of unstructured, i.e. uniform background the prediction error at boundary blocks is not significantly reduced by the use of shape information. This is due to the fact, that the wrong motion compensation in the static background areas, which occurs when no shape is used, does not lead to large prediction errors since the signal to compensate is unstructured. Therefore, two test sequences have been used for the simulations where at least parts of the background are structured. One is the MPEG–4 test sequence "Hall Monitor", the second is the modified MPEG–4 test sequence "Akiyo Garden". In the latter, the background of the original sequence "Akiyo" was replaced by a more structured one; this was possible since the original segmentation masks for that sequence are available.

The simulations were performed at a data rate in the range of 36 kbit/s with a reduced frame frequency of 10 Hz, the image format is QCIF. In Table 1, the average bit rates per frame for encoding colour parameters in the H.263 coder (layers 1 and 2) and for encoding colour and shape parameters in the proposed layered coder (layers 1, 2 and 3) are given. The comparison shows, that for both test sequences the object–based layered coder is superior to the block–based H.263. The results indicate the possible saving of bits using the proposed layered coder.

5. CONCLUSIONS

An object–based layered coder with three coding layers has been proposed for coding of image sequences at very low data rates between 8 and 64 kbit/s. For the first two layers, a standard block–based hybrid coder according to the ITU–T Rec. H.263 was used. For layer 3, a simplified object–based analysis–synthesis coder with the source model of 2D flexible objects was applied. While the automatic selection between layer 1 and 2 is already contained in the H.263 coder, an appropriate criterion for selection between layer 2 and 3 was developed. First simulation results show, that for test sequences with unstructured, uniform background mainly the block–based layers 1 and 2 are chosen, which are identical with the reference H.263 coder. Significant improvements are possible for test sequences with at least partially structured background. Here, the data rate for coding of colour parameters can be reduced by about 5% at the same image quality compared to H.263, already taking into account the additional data rate for shape parameters in the layered coder.

REFERENCES

[1] P. Gerken, "Object–Based Analysis–Synthesis Coding of Image Sequences at Very Low Bit Rates", *IEEE Transactions on Circuits and Systems for Video Technology: Special Issue on Very Low Bit Rate Video Coding,* Vol. 4, No. 3, pp. 228–235, June 1994.

[2] M. Hötter, "Predictive contour coding for an object–oriented analysis–synthesis coder," *IEEE International Symposium on Information Theory*, San Diego, California, U.S.A., p. 75, January 1990.

[3] M. Hötter, "Optimization and efficiency of an object–oriented analysis–synthesis coder", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 4, No. 2, April 1994.

[4] M. Hötter, "Object–oriented analysis–synthesis coding based on moving two–dimensional objects," *Signal Processing: Image Communication*, Vol. 2, No. 4, pp. 409–428, December 1990.

[5] ITU–T, "Video coding for narrow telecommunication channels at < 64 kbit/s", *Draft Recommendation H.263,* January 1995.

[6] H.G. Musmann, M. Hötter, J. Ostermann, "Object–oriented analysis–synthesis coding of moving images", Signal Processing: Image Communication, Vol. 1, Nr. 2, pp. 117–138, Oktober 1989.

[7] H.G. Musmann, "A layered coding system for very low bit rate video coding", *Signal Processing: Image Communication*, Vol. 7, No. 4–6, pp. 267–278, November 1995.

[8] J.Ostermann, "Object–oriented analysis–synthesis Coding based on the source model of moving rigid 3D objects", Signal Processing: Image Communication, Nr. 6. pp. 143–161, 1994.

[9] J.Ostermann, "Object–oriented analysis–synthesis Coding (OOASC) based on the source model of moving flexible 3D objects", IEEE Trans. on Image Processing, Vol. 3, No. 5, September 1994.

Sequence	Freq. layer 3 [%]	R_{col} H.263 [bit/frame]	$R_{col} + R_{sha}$ layered coder [bit/frame]	Saved bit rate [%]
Akiyo Garden	37	3617	3562	1.5
Hall Monitor	55	2674	2548	4.7

Table 1: Frequency of layer 3 coded frames with respect to the total frame number of 99. Average bit rate per frame for coding of colour in H.263 and colour and shape parameters in the layered coder. Saved bit rate of layered coder with respect to the average total colour rate of H.263.