OBJECT–BASED ANALYSIS–SYNTHESIS CODING OF IMAGE SEQUENCES FOR MOBILE MULTIMEDIA APPLICATIONS

Peter GERKEN
Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung
Universität Hannover, Appelstr. 9 A, D–30167 Hannover, Germany

Summary

Mobile multimedia communications require the transmission of image sequences with extremely low bit rates. For this application, an object–based analysis–synthesis coder is presented, which transmits moving images with bit rates around 16 kbit/s.

The principle of such coding schemes, known from [1], is shown as block diagram in Fig. 1. An image analysis decomposes the images of a sequence into moving objects. Each object is described by three sets of parameters, namely shape, motion and colour parameters. The latter consist of the luminance and chrominance values of the object surface. These parameter sets are forwarded to a parameter coding unit for coding and transmission. The overall transmission bit rate consists of three parts R_M, R_A, R_S used for shape, motion and colour parameters, respectively. After decoding, the parameters are written into a memory where they are provided for the image synthesis. The synthesized image is displayed and used for the analysis of the consecutive image.

The goal of this approach is to synthesize as many moving objects as possible from colour parameters which have already been transmitted with previous images and therefore are known at the receiver side. In this case, only shape and motion parameters of the input image are transmitted. This will lead to a good image quality whenever the source model the image analysis is based on describes the moving objects sufficiently accurate, these objects are therefore called model compliance (MC–) objects. Whenever the synthesis based on shape and motion parameters only is not successful colour parameters of the input image must be transmitted. These regions are called model failure (MF–) objects. At low bit rates colour parameter coding requires most of the transmission bit rate. Therefore, the area of the MF–objects must be kept as small as possible.

Image analysis is based on a source model which makes certain assumptions about the objects and their motion. Depending on the applied source model, objects may be either 2– or 3–dimensional, flexible or rigid, and their motion may be either 2– or 3–dimensional. A first implementation of an object–based analysis–synthesis coder for 64 kbit/s ISDN videophone application using an image resolution corresponding to CIF was presented by Hötter [2,3]. It applies the source model of flexible 2–dimensional objects which move translationally in the image plane. The same source model is used here for mobile multimedia applications. Compared to source models with 3–dimensional objects this source model is less complex.

Due to the lower transmission bit rate and the different image resolution corresponding to QCIF some algorithms and parameters of modules used in the coder of [2,3] had to be changed

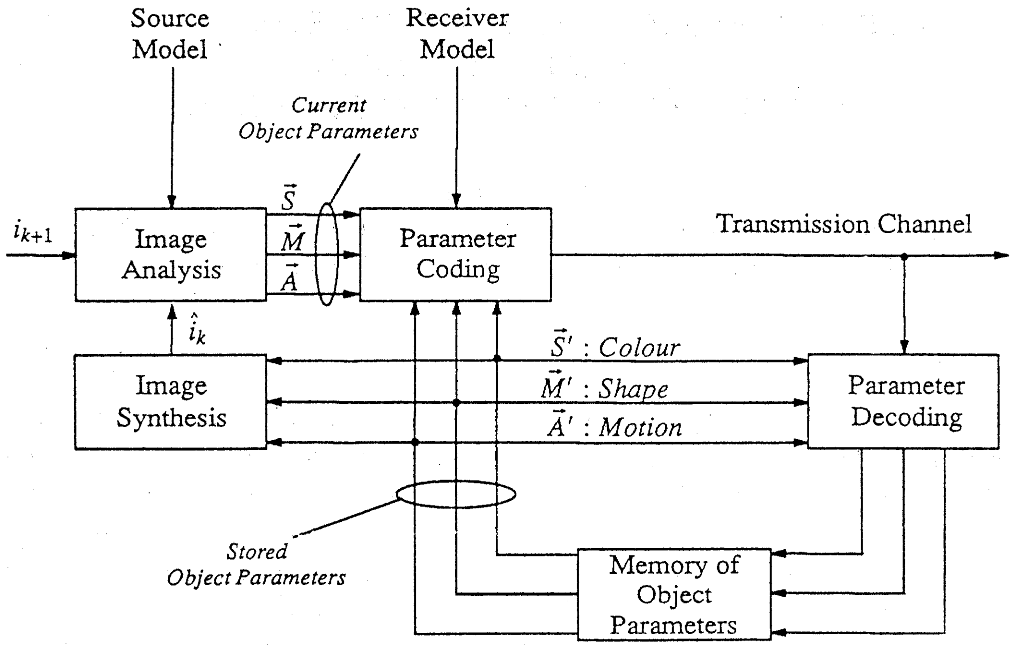

Fig. 1: Block diagram of an object–based analysis–synthesis coder [1]

in order to yield the necessary bit rate reduction and to achieve a better image analysis and by this a better synthesis of MC–objects.

For an improvement of the image analysis, motion description had to be adapted. Due to the underlying source model, a set of displacement vectors must be estimated for each object. Hierarchical block matching [4] which is known to estimate the true motion of an object very well rather than only minimizing the prediction error is used here. A search strategy is applied using different measurement window sizes, signal bandwidths and maximum update displacements on several hierarchy levels. Optimized versions of this algorithm are already known [2,3,5], but not for QCIF image resolution and the transmission bit rates given here. The hierarchical block matching procedure has been optimized for the given constraints. Table 1 summarizes the optimized parameter values. A maximum displacement of +/– 4.5 picture elements has been selected. Corresponding to +/– 9 picture elements in CIF resolution, this is sufficient for videophone scenes supposing moderate motion as the CCITT reference coder for 64 kbit/s even has a maximum displacement of only +/– 7 picture elements [6].

Table 1: Optimized parameter values for hierarchical displacement estimation

Hierarchy level	1	2	3
Max. update displacement	+/– 3 pel	+/– 1 pel	+/– 0.5 pel
Measurement window size	32x32 pel	16x16 pel	16x16 pel
Signal bandwidth reduction factor	1/2	1/1	1/1
Spatial vector field resolution	16x16 pel	16x16 pel	16x16 pel

For shape parameter description, a combined polygon–spline approximation is used for MC–objects like in [2,3]. Approximation accuracy had to be optimized for QCIF image resolution and the transmission bit rates given here. A maximum distance of 2 picture elements to outside an MC–object and 1 picture element to inside has been found optimum. With these values synthesis errors at object boundaries are still tolerable while the bit rate required for MC shape parameters is not unnecessarily high. MF–shape description has been simplified compared to [2,3] in order to reduce the necessary bit rate. The number of vertices has been limited to 4. Only spline approximation is applied.

Shape parameter coding has been modified compared to [2,3]. In order to reduce the bit rate for MC–shapes, prediction of the order of succession of polygon and spline pieces from the preceding image has been introduced. Instead of the order itself, the prediction error is coded now which requires a considerably lower bit rate. MF–shape parameter coding is simplified as information about the approximation type (polygon/spline) is no longer necessary to be coded.

For motion parameter coding, predictive coding is used. Compared to [2,3], where optimal predictor coefficients are calculated and transmitted once per image, fixed predictor coefficients are used here, because at transmission bit rates of 8 or 16 kbit/s the amount of side information would not be tolerable. Furthermore, coding has been adapted to the changed displacement vector ensemble.

For colour parameter coding, conventional hybrid schemes use transform coding with 2D–block DCT. In object–based analysis–synthesis coding, the MF–object shapes are arbitrary. Therefore, more picture elements would be coded with a block–based DCT than necessary. For this reason, an algorithm is used in [2,3] which adaptively selects between DCT transform coding of a complete block of 8x8 picture elements and spatial predictive coding of an arbitrary number of picture elements within a block [7]. Generally, spatial correlation of the image signal is lower at QCIF resolution compared to CIF. Therefore, DCT transform coding achieves less coding gain at QCIF. Even with shape–adapted DCT transform coding [8], which can be considered optimum in the sense that no unnecessary picture elements are coded, the necessary bit rate reduction is not achieved. For this reason, spatial vector quantization is used here for colour parameter coding. Although not optimized, first experimental results with this algorithm yield bit rates below those of the transform based algorithms. Optimization of spatial vector quantization is under current investigation, in order to further reduce the amount of bits necessary for colour parameter coding. Special colour parameter coding for uncovered background has been introduced. Compared to [2,3] where uncovered background is treated in the same way as MF–objects adaptive predictive coding is performed using either the information from the neighbouring static background or from a background memory as reference.

Experimental results, obtained by computer simulations, will be presented at the Workshop. They will reflect the status of research. Currently, typical videophone sequences can be coded at 16 kbit/s with an image quality acceptable for the given application. Bit rates achieved with and without the introduced changes are shown in Table 2. Further optimization is intended to cut down the necessary transmission bit rate to 8 kbit/s while maintaining the image quality at the reached level.

Table 2: Bit rates R_M, R_A, R_S
(Test sequence "Miss America", image resolution QCIF, frame rate 8.33 Hz)

	Achieved bit rates without changes	with changes	Available bit rates at 8 kbit/s	
R_M	500	340	300	bit/frame
R_A	350	190	150	bit/frame
R_S	2000	980	550	bit/frame
R_M+R_A+R_S	24	13	8	kbit/s

Acknowledgement

This work has been supported by the Robert Bosch GmbH, Germany.

References

[1] H.G. Musmann, M. Hötter, J. Ostermann, "Object–Oriented Analysis–Synthesis Coding of Moving Images", Signal Processing: Image Communication, Vol. 1, No. 2, Oct. 1989, pp. 117–138.

[2] M. Hötter, "Object–Oriented Analysis–Synthesis Coding Based on Moving Two–Dimensional Objects", Signal Processing: Image Communication, Vol. 2, No. 4, Dec. 1990, pp. 409–428.

[3] M. Hötter, "Optimization and Efficiency of an Object–Oriented Analysis–Synthesis Coder", submitted for publication to IEEE Trans. Circ. Syst. Vid. Tec..

[4] M. Bierling, "Displacement Estimation by Hierarchical Blockmatching", Visual Commun. and Image Proc. '88, Proc. SPIE 1001, Cambridge, MA, Nov. 1988, pp. 942–951.

[5] P. Gerken, H. Amor, "Hierarchical Displacement Estimation with Low Computational Effort for Applications in 64 kbit/s Moving Video Coding", Proc. 2nd Int. Workshop on 64 kbit/s Coding of Moving Video, Hannover, Germany, Sept. 1989, pp. 3.3.1–3.3.2.

[6] "Description of Ref. Model 8 (RM8)", CCITT Doc. 525, June 1989.

[7] H. Schiller, M. Hötter, "Investigations on Colour Coding in an Object–Oriented Analysis–Synthesis Coder", accepted for publication in Signal Processing: Image Communication.

[8] M. Gilge, T. Engelhard, R. Mehlan, "Coding of Arbitrarily Shaped Image Segments Based on a Generalized Orthogonal Transform", Signal Processing: Image Communication, Vol. 1, No. 2, Oct. 1989, pp. 153–180.