Object-Based Analysis-Synthesis Coding of Image Sequences at Very Low Bit Rates

Peter Gerken

Abstract—An object-based analysis-synthesis image sequence coder for transmission bit rates between 8 and 16 kbit/s is presented. Each moving object is described by three sets of parameters defining its shape, motion, and colour. Coding is based on the source model of flexible 2-dimensional objects which move translationally in the image plane, as it has been used in an implementation for 64 kbit/s ISDN videophone by Hötter. In order to cut down the bit rate from 64 kbit/s to 8 kbit/s, QCIF image resolution instead of CIF resolution is applied. Image analysis and coding of object parameters have been adapted to the reduced resolution and to the changed parameter statistics, respectively. In addition to Hötter’s coder, predictive coding is used for encoding polygons and splines to improve the coding efficiency of shapes. Vector quantization is applied instead of DCT for coding the luminance and chrominance parameters of the object textures. Uncovered background regions are encoded by applying adaptive prediction from either the neighbouring static background or a special background memory. Experimental results obtained by computer simulations, confirm that the presented coding algorithm is a promising scheme for the application at extremely low transmission bit rates. This is shown by comparing the picture qualities obtained with the presented algorithm and a block-based hybrid-DCT scheme corresponding to H.261/RM8 at 11 kbit/s.

I. INTRODUCTION

OBJECT-BASED analysis-synthesis coding (OBASC) has been proposed for the transmission of image sequences at very low transmission bit rates in the range of 8–64 kbit/s. It has been shown, that for typical videophone scenes this coding scheme can avoid the so called blocking and mosquito artifacts of a block-based hybrid-DCT coding as used in the CCITT Recommendation H.261/RM8 [1], [2].

A general description of OBASC is given in [3]. The coder consists of 5 main components: image analysis, parameter coding, parameter decoding, image synthesis, and a memory for the object parameters. By the image analysis, the images of a sequence are decomposed into moving objects each described by three sets of parameters defining the shape, the motion, and the colour of the object. The colour parameters consist of the luminance and chrominance information of the object texture. The goal of this scheme is to synthesize as many objects as possible by transmitting only shape and motion parameters and using colour parameters which have already been transmitted with previous images and therefore are known at the receiver side. Colour parameters are transmitted only for those objects which can not sufficiently be synthesized.

A first implementation of an object-based analysis-synthesis coder has been presented by Hötter [4], [5]. It is based on a source model of flexible 2-dimensional objects which move translationally in the image plane. This implementation, which has been optimized for ISDN videophone applications using an image resolution corresponding to CIF [2] and a transmission bit rate of 64 kbit/s, proved to work very well under these conditions.

In this paper, an object-based analysis-synthesis coder for transmission bit rates between 8 and 16 kbit/s is described. With regard to the very low bit rates, the image resolution is reduced to QCIF [2]. However, direct application of Hötter’s implementation is not efficient for the following reasons. Image analysis is no longer optimum for the reduced image resolution. Especially, the range of the object motion, measured by its displacements, has changed. This requires an adaptation of the applied hierarchical displacement estimation algorithm. Also, coding of the three parameter sets is no longer optimum. It has to be adapted to the changed parameter statistics. In addition, shape parameter coding shall be improved by applying temporal prediction for coding the kind of approximation function polygon or spline. To achieve a higher coding efficiency, the concept of colour parameter coding shall be changed to vector quantization. Furthermore, uncovered background prediction is applied.

In Section II of this paper, the basic principle of object-based analysis-synthesis coding using the source model of flexible 2-dimensional objects which move translationally in the image plane is described. Sections III and IV describe the adaptations of the image analysis and the parameter coding parts, respectively. Experimental results obtained by computer simulations are presented in Section V.

II. PRINCIPLE OF OBJECT-BASED ANALYSIS-SYNTHESIS CODING

The principle of object-based analysis-synthesis coding shall be explained using the block diagram in Fig. 1. Image analysis decomposes each image of a sequence into differently moving objects. For each object, three sets of parameters are determined describing its shape, motion and colour. In addition, each object is classified whether it complies with the underlying source model, i.e. whether the changes of the colour parameters can be described only by that object motion which is allowed by the source model, or not, i.e. the source model fails. Therefore, the first class is called model compliance (MC) objects, the second one model failure (MF-) objects.

Manuscript received November 19, 1993; revised March 18, 1994 and May 1, 1994. It was recommended by Kiyoharu Aizawa.

The author is with the Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, Universität Hannover, Appelstr. 9A, D-30167 Hannover, Germany.

IEEE Log Number 9403535.
The three parameter sets of an object and its object class are forwarded to a parameter coding unit. For MC-objects only shape and motion parameters, for MF-objects shape and colour parameters are coded. A special situation appears when for a part of an MC-object its texture is not known. This is the case of regions where background is uncovered. For these parts of an MC-object its texture is not known. This is the shape and motion parameters, for MF-objects shape and colour parameters have to be coded and transmitted. After transmitting and decoding, the parameters are stored in a memory. They are used for image synthesis as well as for predictive coding of the parameter sets of the consecutive input image. The image synthesis unit reconstructs the image from the stored parameters and provides it for display and for image analysis. MC-objects are synthesized from colour parameters which have already been transmitted with previous images using conventional waveform coding techniques, but considering that the objects are arbitrarily shaped. For example, in the implementation of [4], [5] an algorithm is used with adaptive selection between block-based DCT coding and spatial DPCM coding taking into account arbitrary shapes [6].

III. IMAGE ANALYSIS

By the image analysis the current input image of a sequence is decomposed into differently moving objects. This is performed by segmenting the input image using the previously synthesized image. In the first step, significant changes between the current input and the previous synthesized image are detected. By this, static background, for which is assumed that changes are only due to camera noise, is distinguished from other image regions. In the second step, for each connected region, which is considered to be an MC-object, a set of motion parameters is calculated, since it is assumed that these changes are only caused by motion. Due to the underlying source model, displacement estimation is performed. The segmentation algorithm is again applied to the input image, now using the motion compensated version of the previous synthesized image. By this, those regions are found where significant luminance differences still remain after motion compensation. This can have two reasons: Either the motion in these regions is more complex than described by the source model, or the changes can not be described by motion at all, e.g. when new information enters the image like in situations of opening eyes or an opening mouth. Connected regions of this type are considered to be MF-objects. By evaluation of the motion parameters, also background being uncovered by moving objects is found in this step. These regions are excluded from the MC-objects. Substantial components of the image analysis are the analysis of the object motion and the analysis of the object shape. Both components have been adapted to QCIF image resolution and the extremely low transmission bit rates.

A. Motion Analysis

For motion analysis, hierarchical block matching [7] is used, because it is necessary to estimate the true motion rather than only minimizing displaced frame differences. Displacements are estimated using different measurement window sizes, signal bandwidths, and maximum update displacements on several hierarchy levels. In the first level, large measurement windows are applied to lowpass-filtered versions of the current input image and the previous synthesized image. By this, the displacements are estimated roughly but reliably with regard to the true motion, because the number of false estimates caused by high-frequency structures are reduced. In the second level, the accuracy of the displacement vectors is refined using smaller measurement windows and increased signal bandwidths. In order to maintain the reliability from the first level, the maximum update displacement is kept smaller than the maximum displacement of the first level. This procedure can iteratively be repeated. Optimized versions of this algorithm are already known for an image resolution corresponding to CIF and transmission bit rates of 64 kbit/s and above using three hierarchy levels [4], [5], [8], but not for QCIF image resolution and lower transmission bit rates. The parameter values used in [4], [5] are summarized in Table I.

<table>
<thead>
<tr>
<th>Hierarchy level</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max. update displacement</td>
<td>+/-7 pel</td>
<td>+/-3 pel</td>
<td>+/-0.5 pel</td>
</tr>
<tr>
<td>Measurement window size</td>
<td>64x64 pel</td>
<td>16x16 pel</td>
<td>16x16 pel</td>
</tr>
<tr>
<td>Signal bandwidth reduction factor</td>
<td>1/4</td>
<td>1/2</td>
<td>1/1</td>
</tr>
<tr>
<td>Spatial vector field resolution</td>
<td>32x32 pel</td>
<td>16x16 pel</td>
<td>16x16 pel</td>
</tr>
</tbody>
</table>

B. Shape Analysis

For shape analysis, a combined polygon-spline motion is used in [4], [5]. For a polygon piece two vertices the approximation is successively reinserting further vertices until the required approximation is reached everywhere. After polygon approx.
The three parameter sets of an object and its object class are forwarded to a parameter coding unit. For MC-objects only shape and motion parameters, for MF-objects shape and colour parameters are coded. A special situation appears when a part of an MC-object its texture is not known. This is the case of regions where background is uncovered. For these regions, colour parameters have to be coded and transmitted. After transmission and decoding, the parameters are stored in a memory. They are used for image synthesis as well as for predictive coding of the parameter sets of the consecutive input image. The image synthesis unit reconstructs the image from the stored parameters and provides it for display and for image analysis. MC-objects are synthesized from colour parameters which have already been transmitted with previous images. The shape and motion parameters of the current image, uncovered background regions as parts of an MC-object where the object texture is not known are synthesized using the shape parameters also of the current one. Since the displacements are estimated roughly but reliably with regard to the true motion, because the number of false estimates caused by high-frequency structures are reduced. In the second step, the accuracy of the displacement vectors is refined using smaller measurement windows and increased signal bandwidths. In order to maintain the reliability from the first level, the maximum update displacement is kept smaller than the maximum displacement of the first level. This procedure can iteratively be repeated. Optimized versions of this algorithm are already known for an image resolution corresponding to QCIF transmission bit rates of 64 kbit/s even allows a maximum displacement of +/- 4.5 picture elements in CIF resolution, this is sufficient for videophone scenes supposing moderate motion as the CCITT reference coder for 64 kbit/s. Both components have been adapted to QCIF image resolution.

**A. Motion Analysis**

For motion analysis, the segmentation algorithm is again applied to the input image, now using the motion compensated version of the previous synthesized image. By this, those regions are found where significant luminance differences still remain after motion compensation. This can have two reasons: Either the motion in these regions is more complex than described by the source model, or the changes are not described by motion at all, e.g., when new information enters the image like in situations of opening eyes or an opening mouth. Connected regions of this type are considered to be MF-objects. By evaluation of the motion parameters, also background being uncovered by moving objects is found in this step. These regions are excluded from the MC-objects.

**Substantial components of the image analysis are the analysis of the object motion and the analysis of the object shape. Both components have been adapted to QCIF image resolution and the extremely low transmission bit rates.**

**III. IMAGE ANALYSIS**

By the image analysis the current input image of a sequence is decomposed into differently moving objects. This is performed by segmenting the input image using the previous synthesized image. In the first step, significant changes between the current input and the previous synthesized image are detected. By this, static background, for which is assumed that changes are only due to camera noise, is distinguished from other image regions. In the second step, for each connected changed region, which is considered to be an MC-object, a set of motion parameters is calculated, since it is assumed now that changes are only caused by motion. Due to the underlying source model, displacement estimation is performed. The segmentation algorithm is again applied to the input image, now using the motion compensated version of the previous synthesized image. By this, those regions are found where significant luminance differences still remain after motion compensation. This can have two reasons: Either the motion in these regions is more complex than described by the source model, or the changes are not described by motion at all, e.g., when new information enters the image like in situations of opening eyes or an opening mouth. Connected regions of this type are considered to be MF-objects. By evaluation of the motion parameters, also background being uncovered by moving objects is found in this step. These regions are excluded from the MC-objects.

**TABLE I**

<table>
<thead>
<tr>
<th>Hierarchy level</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max. update displacement</td>
<td>+/- 7 pel</td>
<td>+/- 3 pel</td>
<td>+/- 0.5 pel</td>
</tr>
<tr>
<td>Measurement window size</td>
<td>64x64 pel</td>
<td>16x16 pel</td>
<td>16x16 pel</td>
</tr>
<tr>
<td>Signal bandwidth reduction factor</td>
<td>1/4</td>
<td>1/2</td>
<td>1/1</td>
</tr>
<tr>
<td>Spatial vector field resolution</td>
<td>32x32 pel</td>
<td>16x16 pel</td>
<td>16x16 pel</td>
</tr>
</tbody>
</table>

**TABLE II**

<table>
<thead>
<tr>
<th>Hierarchy level</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max. update displacement</td>
<td>+/- 3 pel</td>
<td>+/- 1 pel</td>
<td>+/- 0.5 pel</td>
</tr>
<tr>
<td>Measurement window size</td>
<td>32x32 pel</td>
<td>16x16 pel</td>
<td>16x16 pel</td>
</tr>
<tr>
<td>Signal bandwidth reduction factor</td>
<td>1/2</td>
<td>1/1</td>
<td>1/1</td>
</tr>
<tr>
<td>Spatial vector field resolution</td>
<td>16x16 pel</td>
<td>16x16 pel</td>
<td>16x16 pel</td>
</tr>
</tbody>
</table>
has been finished, for each pair of two consecutive vertices it is tested whether a spline approximation also reaches the required approximation accuracy. If so, spline approximation substitutes polygon approximation between these two vertices. By this combined technique, the different drawbacks of both single methods are avoided [4]. Fig. 2 gives an example for a very low approximation accuracy. If possible, a shape is taken from the previous image and motion compensated. Those vertices which violate the approximation accuracy requirement are then rejected. The remaining vertices are used for the initial approximation instead of the 4-vertices polygon.

In the coder presented here, this algorithm of a combined polygon-spline approximation is only applied to MC-objects. The approximation accuracy had to be adapted to the extremely low transmission bit rates and the lower image resolution compared to [4], [5]. On the one hand, high accuracy guarantees a good synthesis of the image signal near object boundaries but requires a high data rate for shape parameters. On the other hand, low accuracy requires only a small data rate for shape parameters but leads to synthesis errors near object boundaries. Thus, a compromise had to be found for the accuracy of shape approximation by subjectively evaluating the synthesis quality near object boundaries and measuring the data rate for shape parameters. A maximum distance between approximation and real shape of two picture elements outside of an MC-object and one picture element inside has been found to keep the synthesis errors still acceptable, while the data rate required for MC-object shape parameters is kept reasonably low.

Concerning MF-object shapes, it could be found in computer simulations that shape parameter coding requires a too high data rate when applying the same algorithm as for MC-object shapes. Thus, the remaining data rate for colour parameter coding is too low for providing a sufficient image quality in MF-objects. In order to overcome this problem, the data rate required for MF-object shapes had to be substantially reduced. For this purpose, the number of vertices has been fixed to 4, and spline approximation is the only performed approximation method. By this, neither the number of vertices nor the selected approximation method need to be transmitted to the receiver any longer. Spline approximation has been chosen because it fits the MF-object shapes, which are usually convex, better than polygon approximation. Also with the combined polygon-spline technique, spline approximation would be mostly selected.

IV. PARAMETER CODING

A. Shape Parameter Coding

The shape parameters which must be coded are the coordinates of the vertices and the kind of approximation function polygon or spline for each pair of consecutive vertices. In case of predictable MC-object shapes, the order of maintained, rejected, and inserted vertices are coded instead of the coordinates of all vertices. Only the coordinates of the inserted vertices must then still be coded. This method of predictive coding for the vertices is already used in [4], [5]. The kind of approximation function, however, is coded without prediction. Instead, the runlengths of consecutive polygon pieces and consecutive spline pieces are coded, although there are significant temporal statistical dependencies of the selected kind of approximation function.

In the coder presented here, these dependencies are exploited by predictive coding of the approximation function. This requires prediction values of the approximation function for each combination of rejected and inserted vertices between pairs of maintained vertices. Because only the maintained vertices are defined in the previous as well as in the current image. Furthermore, the combinations of rejected and inserted vertices are arbitrary and their number is very huge. Statistical evaluations of several videophone sequences have been performed in order to get the most likely combination of polygons and splines for each combination of rejected and inserted vertices, where also the number of vertex combinations had to be reduced. This method, the number of DCT coefficients which must be coded would exceed the number of pixels inside the MF-object. Using QICF image resolution, this results in a data rate (including chrominance parts) of about 55 kbit/s for the MF-object. Without any vector quantization, the required data rate is significantly below that required by any of the above described hybrid DCT schemes. Note that the required data rate for the MF-object requires a lower data rate for the same image quality than coding of the DCT coefficients of the complete block. The small part that belongs to an MF-object, the higher the probability that DPCM coding is selected. By introducing this alternative, an improvement is already achieved. However, for QCIF image resolution, the data rate can not be cut down below 2 bits per pixel element of the MF-objects.

Shape adapted DCT coding [10] can be taken as a reference, because it can be considered optimum in the sense that no unnecessary picture elements are coded, i.e. the number of coded transform coefficients equals the number of picture elements to be coded. But also with this algorithm, the data rate for the colour parameters can not be reduced to below 2 bits per pixel element. Since due to the low resolution the image signal is strongly decorrelated, DCT coding can not yield the necessary data rate among luminance and chrominance parameter. Therefore, spatial DPCM coding has been used for the luminance components as well as for the chrominance components. Vectors of 2 x 2 x 32 to 1024 entries are generated by training of one videophone based application. In [4], [5], an algorithm is used which adaptively selects between DCT transform coding and spatial DPCM coding [6]. For each block of 8 x 8 picture elements outside of a block, the probability that DPCM coding of only those picture elements which belong to the MF-object requires a lower data rate for the same image quality than coding of the DCT coefficients of the complete block. The small part that belongs to an MF-object, the higher the probability that DPCM coding is selected. By introducing this alternative, an improvement is already achieved. However, for QCIF image resolution, the data rate can not be cut down below 2 bits per pixel element of the MF-objects.

B. Motion Parameter Coding

As motion parameters, displacement vectors must be coded. Since there is a significant statistical dependency among neighbouring vectors, spatially predictive coding is performed in [4], [5] using optimal 3-tap predictors for the x- and y-components calculated once per frame. Since the predictor coefficients are not known at the receiver side, they must be transmitted as side information. For a transmission bit rate of 64 kbit/s, where about 1100 bit/frame are used for motion parameters [5], the data rate of about 560 bit/frame necessary for this side information is negligible. However, for a transmission bit rate of 8 kbit/s as an example, where a reasonable distribution of the transmission bit rate allows a data rate of only 150-200 bit/frame for motion parameters, the amount of this side information would not be tolerable.

In order to reduce the data rate for motion parameters, the coder presented here uses fixed predictor coefficients so that transmission of predictor coefficients is no longer necessary. Prediction from only the previously transmitted displacement vector, i.e. 1-tap prediction, is performed. By this, the desired data rate of 150-200 bit/frame has already been achieved. However, it is not the predictor order which can further reduce the data rate, but may be subject of future investigations.

Prediction errors for displacement vectors undergo an adaptive arithmetic coding joining zero-vectors to runlengths. Arithmetic coding has been adjusted to the maximum displacement.

C. Colour Parameter Coding

Colour Parameter Coding in MF-Objects

In block-based coding, usually hybrid-DCT coding of square blocks is employed. In object-based coding, however, spatially arbitrary object shapes are arbitrary. In this case, the number of DCT coefficients which must be coded would exceed the number of picture elements inside the MF-object. Using QCIF image resolution, this results in a data rate (including chrominance parts) of about 55 kbit/s for the MF-object. Without any vector quantization, the required data rate is significantly below that required by any of the above described hybrid DCT schemes. Note that the required data rate for the MF-object requires a lower data rate for the same image quality than coding of the DCT coefficients of the complete block. The small part that belongs to an MF-object, the higher the probability that DPCM coding is selected. By introducing this alternative, an improvement is already achieved. However, for QCIF image resolution, the data rate can not be cut down below 2 bits per pixel element of the MF-objects.

Shape adapted DCT coding [10] can be taken as a reference, because it can be considered optimum in the sense that no unnecessary picture elements are coded, i.e. the number of coded transform coefficients equals the number of picture elements to be coded. But also with this algorithm, the data rate for the colour parameters can not be reduced to below 2 bits per pixel element. Since due to the low resolution the image signal is strongly decorrelated, DCT coding can not yield the necessary data rate among luminance and chrominance parameter. Therefore, spatial DPCM coding has been used for the luminance components as well as for the chrominance components. Vectors of 2 x 2 x 32 to 1024 entries are generated by training of one videophone based application. In [4], [5], an algorithm is used which adaptively selects between DCT transform coding and spatial DPCM coding [6]. For each block of 8 x 8 picture elements outside of a block, the probability that DPCM coding of only those picture elements which belong to the MF-object requires a lower data rate for the same image quality than coding of the DCT coefficients of the complete block. The small part that belongs to an MF-object, the higher the probability that DPCM coding is selected. By introducing this alternative, an improvement is already achieved. However, for QCIF image resolution, the data rate can not be cut down below 2 bits per pixel element of the MF-objects.
this is controlled by memory and thus can be used for prediction at the second or the first time, only spatial prediction is performed. The de­part of the background and temporal prediction from a special selects between spatial prediction from the neighbouring static ground regions has been developed. The algorithm adaptively enough. Additionally, certain possibilities of prediction are not uncovered background regions are usually long and narrow.

Fig. 3. Adaptive prediction in uncovered background.

Colour Parameter Coding in Uncovered Background Regions

In [4], [5], uncovered background regions are processed in the same way as MF-objects. This is not optimal because uncovered background regions are usually long and narrow. Therefore, block-based coding, as used there, is not efficient enough. Additionally, certain possibilities of prediction are not taken into account.

For the coder presented here, an algorithm with adaptive predictive coding of colour parameters in uncovered background regions has been developed. The algorithm adaptively selects between spatial prediction from the neighbouring static part of the background and temporal prediction from a special background memory. The information about the selected prediction mode must be transmitted to the receiver as a coding attribute. When a part of the background appears in a sequence for the first time, only spatial prediction is performed. The de­coded colour parameters are then written into the background memory and thus can be used for prediction at the second or following appearances of the same part. This is controlled by the shape parameters of the MC-objects in the previous and the current image. Thus, all parts of the background which have been once visible during a sequence are available for temporal prediction. From the second appearance on, adaptive selection of the appropriate prediction mode can be performed. This is illustrated in Fig. 3.

Currently, prediction errors are scalar quantized. Zero values are connected to runlengths. Prediction errors and runlengths are entropy coded as combined events.

With the presented algorithm, the data rate required for the colour parameters in uncovered background regions could be reduced to 0.6 bits per picture element compared to about 1.6 bit per picture element without this method (test sequence “Miss America”). Further data rate reduction has been achieved by excluding uncovered background from colour parameter coding when the required image quality is achieved even without prediction error transmission. In these cases, a coding attribute is transmitted instead of quantized prediction errors.

Currently, both coding attributes (prediction mode and prediction error coding yes/no) are determined once per frame. This is done in order to keep the amount of side information small. Applying the coding attribute “prediction error coding yes/no” to complete images causes that, due to the well-working prediction algorithm, in sequences without highly structured background, like the test sequence “Miss America”, colour parameters in uncovered background regions must be coded only for a few images. Whether the overall data rate further decreases when uncovered background is divided into several parts and the coding attributes are determined more than once per frame, may be subject of further investigations.

V. EXPERIMENTAL RESULTS

The object-based analysis-synthesis coder has been investigated by computer simulations using several typical videophone sequences. Results will be given for the test sequence “Miss America”, because this sequence requires the highest data rates, i.e. about 500 bit/frame, for shape and motion parameters which in turn leaves over the smallest data rate for colour parameters. For the test sequence “Claire” as a further example, this is less critical. The spatial image resolution has been 176 x 144 picture elements for the luminance component and 88 x 72 picture elements for the chrominance components (QCIF image resolution). The frame rate has been 8.33 Hz.

Instead of fixing an overall data rate, a minimum image quality was fixed for both the MC- and the MF-objects. 36 dB has been chosen as the required Peak-SNR value of the quantization error in the MF-objects. A maximum synthesis error variance of 6dBSS, corresponding to about 40 dB, has been chosen for MC-objects excluding regions of geometrical distortions where according to [4] higher synthesis errors are intentionally allowed. As the first frame of the reconstructed sequence the first frame of the original sequence was taken. Coding of the first frame is under current study. The resulting data rates $R_2$, $R_3$, and $R_4$ for the three sets of shape, motion, and colour parameters have been measured before and after adaptation. They are shown in Table III together with the target data rates for 8 kbit/s.

As it is demonstrated there, the required data rates for the three parameter sets could be substantially reduced by the introduced changes. The coder can already code image sequences at a transmission bit rate of about 11 kbit/s with an image quality acceptable for the given applications. This image quality is intended to be maintained even when further improvements of the coding algorithm still necessary to cut down the overall data rate will have been accomplished.

Fig. 4 proves the achieved image quality to be acceptable for the extremely low transmission bit rates given here. Furthermore, a comparison with RM8 [1] is drawn where the quantizer step size has been selected so that the same transmission bit rate results. Also here, the first frame of the original sequence was taken as the first reconstructed frame. It can be seen that, in opposite to RM8, the presented coder causes only slight degradations which can be assessed to be subjectively not annoying.

VI. CONCLUSIONS

An object-based analysis-synthesis image sequence coder for transmission bit rates between 8 and 16 kbit/s has been presented. An implementation for 64 kbit/s ISDN videophone applications has been taken as a basis. It uses the source model of flexible 2-dimensional objects which move translationally in the image plane. QCIF image resolution has been taken for the input images instead of CIF used in the 64 kbit/s environment.

Parameter values of the displacement estimation algorithm have been adapted to the reduced image resolution. Especially, the range of displacements has been reduced to 4.5 pixel elements, which can be considered to be sufficient for typical videophone scenes.

Coding of the three object parameter sets has been adapted to the changed parameter statistics. For MC-object shapes, temporal prediction has been extended to the kind of approximation function polygon or spline, exploiting the temporal statistical dependencies of these parameters. By this, the data rate required for MC-object shape parameters is reduced by 25% to below 200 bit/frame. The data rate required for MF-object shape parameters has been reduced from about 250
bit/frame to below 150 bit/frame by limiting the number of vertices to 4.

For motion parameter coding, fixed predictor coefficients have been used instead of an optimal predictor, thus avoiding transmission of side information, and entropy coding has been adapted to the changed range of displacements. By this, the data rate required for motion parameters has been reduced from about 250 bit/frame to below 200 bit/frame.

For colour parameter coding in MF-objects, vector quantization has been introduced instead of DCT. Compound vector quantization of the colour components turned out to reduce the data rate for colour parameters from 1500 bit/frame to 800 bit/frame when compared to hybrid DCT coding. Instead of applying the same algorithm for coding colour parameters in areas of uncovered background, adaptive predictive coding has been implemented, which selects between spatial prediction from the neighbouring static background or temporal prediction from a background memory, in which all background parts once visible in a sequence are stored. This reduces the data rate for colour parameters in uncovered background from 1.6 to 0.6 bit/pel.

The complete coding algorithm has been simulated in computer experiments. At an overall data rate of about 11 kbit/s and QCIF image resolution, an image quality is achieved which is similar to that obtained with 64 kbit/s and CIF if neglecting the change of image resolution. Compared to RM8, scaled down to the same transmission bit rate, image quality is substantially improved. 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.

In the current state of investigations, the presented coding algorithm can be applied to typical videophone scenes where one talking person is moderately moving in front of a static background. Current investigations concern sequences with moving background as usual in mobile environments and sequences with scene contents other than talking persons. Since the objects are considered to be flexible and 2-dimensional, there is no restriction to the scene contents by the underlying source model. Also, a fallback mode for cases in which the source model is not applicable will be introduced. With these extensions, the coding algorithm may be used for a wide range of applications.

ACKNOWLEDGMENT

The author wishes to thank Prof. Dr.-Ing. H. G. Musmann and Dipl.-Ing. D. Lappe for encouraging this work and many helpful discussions. This work has been supported by Robert Bosch GmbH, Germany.

REFERENCES


Peter Gerken was born near Hamburg, Germany, in 1960. He received the Dipl.-Ing. degree in electrical engineering from the University of Hannover, Germany, in 1985. Since 1985 he has been with the Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung at the University of Hannover, where he has been involved in several projects of image coding. His current research interest is in the field of object-based analysis-synthesis coding for very low bit rate applications.