Coding of Eyes and Mouth Areas in a Knowledge–Based Analysis–Synthesis Coder

Markus Kampmann
Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung
Universität Hannover, Appelstraße 9A, 30167 Hannover, F.R.Germany
Phone: ++49–511–7625314, email: kampmann@tnt.uni–hannover.de

ABSTRACT: *In this contribution, an algorithm for coding of eyes and mouth areas in a knowledge–based analysis–synthesis coder is presented. As a first part of this algorithm, a new quality criterion for the image quality inside the facial area is proposed assuming eyes and mouth being subjectively more important. As a second part of this algorithm, codebooks for eyes and mouth are developed. For the test sequence "Miss America" (CIF,10 Hz), the application of both parts of the described algorithm reduces the data rate from 49 kbit/s to 45 kbit/s at the same subjective image quality.*

1. INTRODUCTION

For coding of moving images at low data rates, an object–based analysis–synthesis coder (OBASC) has been introduced[1]. In an OBASC, real objects are described by model objects. A model object is defined by motion, shape and color parameters, which are estimated by an automatic image analysis. By the source model of moving 3D objects [2][3], the shape of the model object is represented by a 3D wireframe. The motion parameters describe translation and rotation of the model object, the color parameters denote luminance and chrominance reflectance of the object surface. After estimation of motion and shape parameters by the image analysis, a prediction image can be synthesized using previously transmitted color, shape and motion parameters as well as the currently estimated motion and shape parameters. Areas of the real image which cannot be described by the prediction image with sufficient subjective image quality are called areas of model failures. Areas of model failures are encoded by color and 2D shape parameters that have to be transmitted. Since the transmission of these parameters is expensive in terms of data rate, the areas of model failures should be as small as possible. In [4], an OBASC is extended to a knowledge–based analysis–synthesis coder (KBASC) by use of the 3D face model *Candide* [5]. In an KBASC, the knowledge of the real object being a person is evaluated. At the beginning of the image sequence, the face model is adapted automatically to the person's face in the scene. After adaption, the wireframe with the face model is motion compensated during the image sequence. By projection of the face model onto the image plane, the location of the person's face in the image sequence is known. Evaluating the location of the face, the criterion for the detection of model failure areas distinguishes between the subjectively more important face and the less important area outside the face. Thus, the areas of model failures outside the face are reduced. This is done without subjective loss of image quality[4].

Eyes and mouth are the most important areas concerned with facial expression. Therefore, special methods for coding of eyes and mouth areas have been developed. Because eyes and mouths appear several times in an image sequence with nearly the same expression, the use of codebooks for coding of eyes and mouth has been proposed [6][7][8]. Eye and mouth patterns with different expressions are stored in codebooks at the coder and decoder. At the coder, those eye and mouth patterns from the codebooks are selected which are the best match for the eyes and mouth in the current image. The indices to the selected codebook entries are transmitted. At the decoder, the selected eye and mouth patterns are taken from the codebooks for image synthesis. *Prototype prediction* [6] has been developed for an OBASC with the source model of moving 2D objects. This method does not require knowledge about the scene contents. Areas of model failures from previously transmitted images are stored in a codebook. These areas are used for color prediction of the model failure areas of the current image. For the algorithms [7] and [8], the positions of eyes and mouth are available. In [7], a facial image coding scheme is proposed using one codebook for eye patterns and one codebook for mouth patterns. Those eye and mouth patterns which are the best match for the eyes and mouth in the current image are selected. At the decoder, the current image is synthesized by inserting the selected patterns into a full face image that is transmitted once at the beginning of the sequence and is held fixed from that on. In this coding scheme, the codebooks are created by training on a large number of images of the person and transmitted to the receiver at the beginning of the transmission. In [8], codebooks for the left eye, the right eye and the mouth are added to a face muscle model. After manual adaptation of the face model to the person's face in the scene, facial expressions are synthesized by simulating human face muscles. Because eye and mouth are difficult to model, patterns from the codebooks are used for texture updates.

In this contribution, an algorithm for coding of eyes and mouth areas in a knowledge–based analysis–synthesis coder (KBASC) is presented. This algorithm consists of two parts. As a first part, a new quality criterion for detection of model failure areas is introduced in Section 2. This new quality criterion considers the positions of eyes and mouth in the image sequence. As a second part, codebooks for eyes and mouth are introduced in Section 3.

2. KNOWLEDGE–BASED DETECTION OF MODEL FAILURES CONSIDERING EYES AND MOUTH POSITIONS

After adaptation, the face model *Candide* is motion compensated during the image sequence. Assuming, that the face model tracks the person's face in the image with good accuracy, the location of the person's face in the sequence is known. Evaluating the location of the face, the criterion used until now in KBASC for the detection of model failure areas distinguishes between the subjectively more important face and the less important area outside the face [4]. In the facial area the eyes and the mouth are the subjectively most important areas. Therefore, the new criterion for detection of model failure areas distinguishes between four different areas, the eyes and mouth, the face, the head, and the body. By projection of the face model's eyes and mouth onto the image plane the positions of eyes and mouth in the image sequence are known. Inside the facial area, a distinction between the eyes and mouth areas and the other parts of the face is carried out by allowing larger synthesis errors outside the areas of eyes and mouth.

3. CODEBOOKS FOR EYES AND MOUTH

For coding of model failures inside the eyes and mouth areas independent codebooks for the left eye, the right eye and the mouth are applied in this contribution. Knowning positions of eyes and mouth in the image sequence, eye and mouth patterns are extracted automatically from the transmitted images (Fig. 1(a)). The extracted patterns are rectangular (Fig. 1(b)). The size of the rectangular patterns is determined automatically in accordance with the size of the person's face in the image. The extracted eye and mouth pattern are stored in codebooks. These codebooks are associated with the surface of the face model and move together with the face model. Assuming, that the face model tracks the person's face in the image with good accuracy, the projections of the eye and mouth patterns from the codebooks onto the image plane match eyes and mouth of the current image by means of size and incline.

The codebooks are applied and updated as follows. After estimation of motion and shape parameters by the image analysis, a prediction image is synthesized using previously transmitted color parameters and the estimated motion and shape parameters. Afterwards, preliminary areas of model failures are determined. Patterns from the codebooks are used to improve the prediction image.

For model failure areas covering eyes and mouth areas (Fig. 2(d)), eye and mouth patterns from the codebooks are projected onto the image plane and compared with the current real image by calculation the mean squared luminance difference MSE. Eyes and mouth pattern with the lowest MSE are selected and their projections are inserted into the prediction image (Fig. 2(c)).

Areas of model failures are determined again to assess the modified prediction image. If a selected pattern does not reduce the areas of model failures, it will be removed from the prediction image again and the former prediction image content is restored. In this case areas of model failures are determined a third time. Then, color and 2D shape parameters for these final areas of model failures are transmitted (Fig. 2(e)).

If a selected pattern does not improve the prediction, i.e. a pattern similar to the eye or mouth in the current image is not available in the codebooks, a new pattern will be extracted from the transmitted image and will be stored in the codebook. If a codebook is full, the less used pattern will be replaced.

4. EXPERIMENTAL RESULTS

The described algorithms have been integrated into a knowledge–based analysis–synthesis coder (KBASC) [4] and have been applied to typical head–and–shoulders videophone sequences. For the test sequence *Miss America* (CIF,10 Hz), knowledge–based coding without the described algorithms for coding of eyes and mouth areas requires a data rate of 49 kbit/s. Applying the new quality criterion for detection of model failure areas as well as the codebooks for eyes and mouth, the area of model failures is reduced by 10%. This leads to a reduced data rate of 45 kbit/s without subjective loss of image quality.

5. CONCLUSIONS

For coding of eyes and mouth areas in a knowledge–based analysis–synthesis coder a new algorithm has been presented. This algorithm consists of two parts. As a first part, a new quality criterion for detection of model failure areas has been introduced. This new criterion distinguishes between four different areas, the eyes and mouth as the subjectively most important areas, the face, the head and the body. As a second part, independent codebooks for the left eye, the right eye and the mouth are used. For the test sequence *Miss America* (CIF,10 Hz), the application of both parts of the described algorithm reduces the data rate from 49 kbit/s to 45 kbit/s without subjective loss of image quality.

This work has been supported by the Deutsche Telekom AG.

6. REFERENCES

[1] H.G. Musmann, M. Hötter, J. Ostermann, ”Object–oriented analysis–synthesis coding of moving images”, Signal Processing: Image Communications, Vol.3, No. 2, pp.117–138, November 1989.

[2] J. Ostermann, ”Object–based analysis–synthesis Coding based on the source model of moving rigid 3D objects”, Signal Processing: Image Communications, Vol.6, pp. 143–161, May 1994.

[3] J. Ostermann, ”Object–based Analysis–Synthesis Coding (OBASC) based on the Source Model of Moving Flexible 3D Objects”, IEEE Trans. on Image Processing, Vol.3, No.5,pp.705–711, September 1994.

[4] M. Kampmann, J. Ostermann, ”Automatic Adaptation of a Face Model in a Layered Coder with an Object–based Analysis–Synthesis Layer and a Knowledge–based Layer”, accepted for publication in Signal Processing: Image Communication.

[5] R. Rydfalk, ”CANDIDE, A parameterised face”, Internal Report Lith–ISY–I–0866, Linköping University, Linköping, Sweden, 1987.

[6] M. Wollborn, ”Object–oriented analysis–synthesis coding using prototype prediction for colour update”, Picture Coding Symposium (PCS ’93), Lausanne, Suisse, No. 17.6, March 1993.

[7] W.J. Welsh, S. Searby, J.B. Waite, ”Model–based image coding”, British Telecom Technology Journal, Vol.8, No.3, pp.94–106, July 1990.

[8] S. Chao, J.A. Robinson, ”Model–Based Analysis/Synthesis Image Coding with Eye and Mouth Patch Codebooks”, Proceedings of Vision Interface, 1994, Banff, Alberta, pp.104–109, May 1994.

(a) (b)

Fig. 1: Testsequence Miss America (CIF,10Hz,frame 2): (a) Synthesized image; (b) extracted eye and mouth patterns.

(a) (b) (c)

(d) (e)

Fig. 2: Testsequence Miss America (CIF,10Hz,frame 15): (a) original frame; (b) prediction image before insertion of eye and mouth patterns; (c) prediction image after insertion of eye and mouth patterns (left eye from frame 12, mouth from frame 13); (d) areas of model failures over the original frame before insertion of eye and mouth patterns; (e) areas of model failures over the original frame after insertion of eye and mouth patterns.