Reprinted from

SIGNAL PROCESSING: IMAGE COMMUNICATION

Signal Processing: *Image Communication* 6 (1994) 143–161

Object-based analysis–synthesis coding based on the source model of moving rigid 3D objects

Jörn Ostermann

Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, Universität Hannover, Appelstraße 9A, D-30167 Hannover 1, Germany

Received 1 March 1993

ELSEVIER

SIGNAL PROCESSING

IMAGE COMMUNICATION

Theory, Techniques & Applications

A publication of the European Association for Signal Processing (EURASIP)

Editor-in-Chief
Leonardo CHIARIGLIONE
Centro Studi e Laboratori
Telecommunicazioni (CSELT)
Via Guglielmo Reiss-Romoli, 274
I-10148 Torino, Italy
Telephone: (11) 2286 120
Telefax: (11) 2286 299

Editorial Board
J. Biemond (Univ. Delft, The Netherlands)
G. Boerger (HHI, Germany)
E. Dubois (Univ. Quebec, Canada)
B. Girod (Köln, Germany)
H. Harashima (Univ. Tokyo, Japan)
C.N. Judice (Bellcore, USA)
J.-K. Kim (KAIST, South Korea)
A.B. Lippman (MIT, USA)
H.G. Musmann (Univ. Hannover, Germany)
D. Nasse (CCETT, France)
A. Netravali (ATT, USA)
Y. Ninomiya (NHK, Japan)
D. Pearson (Univ. Essex, United Kingdom)
H. Seguin (CNET, France)
L. Stenger (FI/DBP, Germany)
H. Tominaga (Waseda Univ., Japan)
M. Vetterli (Columbia Univ., USA)
L.T. Wu (ITRI, Taiwan)
H. Yasuda (NTT, Japan)

Guest Editors
S.A. Benton (MIT, USA)
T. Fujii (NTT, Japan)
J. Guichard (CNET, France)
J. Hamasaki (Univ. Tokyo, Japan)
J. Johann (Deutsche Bundespost, Germany)
D. LeGall (C-Cube, USA)
B. Liu (Princeton Univ., USA)
G. Morrison (BT Labs, United Kingdom)
S. Okubo (NTT, Japan)
T. Omachi (NEC, Japan)
W. Verbiest (Belgium)

Editorial Policy. SIGNAL PROCESSING: IMAGE COMMUNICATION is an international journal for the development of the theory and practice of image communication. Its primary objectives are the following:
- To present a forum for the advancement of the theory and practice of image communication.
- To stimulate cross-fertilization between areas similar in nature which have traditionally been separated, for example, various aspects of visual communications and information systems.
- To contribute to a rapid information exchange between the industrial and academic environments.

The editorial policy and the technical content of the journal are the responsibility of the Editor-in-Chief and the Editorial Board. The Journal is self-supporting from subscription income and contains a minimum amount of advertisements. Advertisements are subject to the prior approval of the Editor-in-Chief. The journal welcomes contributions from every country in the world.

Scope. SIGNAL PROCESSING: IMAGE COMMUNICATION publishes articles relating to aspects of the design, implementation and use of image communication systems.
SIGNAL PROCESSING: IMAGE COMMUNICATION features original research work, tutorial and review articles, and accounts of practical developments.

Subjects. Subject areas covered by the journal include:

TV, HDTV and 3D-TV systems
Visual Science
Image Coding
Image Transmission
Interactive Image Communication
TV and Advanced TV Broadcasting
Image Storage and Retrieval
Graphic Arts
Electronic Printing
Imaging Technology
Display Technology
VLSI Processors for Image Communications

Membership and Subscription Information.
SIGNAL PROCESSING: IMAGE COMMUNICATION (ISBN 0923-5965) is published in one volume (six issues) a year. For 1994 Volume 6 is scheduled for publication. Subscription prices are available upon request from the publisher. Subscriptions are accepted on a prepaid basis only and are entered on a calendar year basis. Issues are sent by surface mail except to the following countries, where air delivery (S.A.L. – Surface Air Lifted) is ensured: Argentina, Australia, Brazil, Canada, Hong Kong, India, Israel, Japan, Malaysia, Mexico, New Zealand, Pakistan, People's Republic of China, Singapore, South Africa, South Korea, Taiwan, Thailand, USA. For the rest of the world, airmail charges are available upon request. Claims for missing issues will be honoured free of charge if made within six months after the publication date of the issues. Mail orders and inquiries to: Elsevier Science B.V., Journals Department, P.O. Box 211, 1000 AE Amsterdam, The Netherlands. For full membership information of the Association, possibly combined with a subscription at a reduced rate, please contact: EURASIP, P.O. Box 134, CH-1000 Lausanne 13, Switzerland.

© **1994 Elsevier Science B.V. All right reserved**

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publisher, Elsevier Science B.V., Copyright and Permissions Department, P.O. Box 521, 1000 AM Amsterdam, The Netherlands.
Special regulations for authors. Upon acceptance of an article by the journal, the author(s) will be asked to transfer copyright of the article to the publisher. This transfer will ensure the widest possible dissemination of information.
Special regulations for readers in the USA. This journal has been registered with the Copyright Clearance Center, Inc. Consent is given for copying of articles for personal or internal use, or for the personal use of specific clients. This consent is given on the condition that the copier pays through the Center the per-copy fee stated in the code on the first page of each article for copying beyond that permitted by Sections 107 or 108 of the US Copyright Law. The appropriate fee should be forwarded with a copy of the first page of the article to the Copyright Clearance Center, Inc., 27 Congress Street, Salem, MA 01970, USA. If no code appears in an article, the author has not given broad consent to copy and permission to copy must be obtained directly from the author. The fees indicated on the first page of an article in this issue will apply retroactively to all articles published in the journal, regardless of the year of publication. This consent does not extend to other kinds of copying, such as for general distribution, resale, advertising and promotion purposes, or for creating new collective works. Special written permission must be obtained from the publisher for such copying.
No responsibility is assumed by the Publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein.
Although all advertising material is expected to conform to ethical standards, inclusion in this publication does not constitute a guarantee or endorsement of the quality or value of such product or of the claims made of it by its manufacturer.

This volume is printed on acid-free paper.

Published 6 times a year 0923-5965/94/$07.00 Printed in The Netherlands

ELSEVIER

Signal Processing: *Image Communication* 6 (1994) 143–161

SIGNAL PROCESSING:
IMAGE
COMMUNICATION

Object-based analysis–synthesis coding based on the source model of moving rigid 3D objects

Jörn Ostermann

Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, Universität Hannover, Appelstraße 9A, D-30167 Hannover 1, Germany

Received 1 March 1993

Abstract

The topic of investigation was object-based analysis–synthesis coding (OBASC) using the source model of 'moving rigid 3D objects' for the encoding of moving images at very low data rates. According to the coding concept, each moving object of an image is described and encoded by three parameter sets defining its motion, shape and surface color. The parameter sets of each object are obtained by image analysis. They are coded using an object-dependent parameter coding. Using the coded parameter sets, an image can be synthesized by model-based image synthesis. In comparison to block-based hybrid coding, OBASC requires the additional transmission of shape parameters. The transmission of shape information avoids the mosquito and block artifacts of a block-based coder. Furthermore, important areas such as facial areas can be reconstructed with a significant image quality improvement. OBASC based on the source models of 'moving flexible 2D objects' and of 'moving rigid 3D objects' gives almost identical image quality for the same data rate. Therefore the use of more advanced source models like flexible 3D objects or 3D face models is expected to further improve image quality.

Key words: Object-based image coding; Analysis–synthesis coding; Videophone; 3D motion estimation; Image analysis; Parameter coding; Shape coding; MPEG-4

1. Introduction

For the coding of moving images with low data rates between 64 kbit/s and 2 Mbit/s, a block-based hybrid coder has been standardized by the CCITT [5] where each image of a sequence is subdivided into independently moving blocks of size 16×16 picture elements (pels). Each block is coded by 2D motion compensated prediction and transform coding [37]. This corresponds to a source model of '2D square blocks moving translationally in the image plane', which fails at boundaries of naturally moving objects and causes coding artifacts known as blocking and mosquito effects at low data rates.

In order to avoid these coding distortions, the new concept of object-based analysis–synthesis coding (OBASC) aiming at a data rate of 64 kbit/s and below was proposed [32]. A coder based on this concept divides an image sequence into moving

* Corresponding author: Email: osterman@tnt.uni-hannover.de

0923-5965/94/$7.00 © 1994 Elsevier Science B.V. All rights reserved
SSDI 0923-5965(94)00003-2

objects. An object is defined by its uniform motion and described by motion, shape and color parameters, where color parameters denote luminance and chrominance reflectance of the object surface. The coding efficiency of OBASC mainly depends on the selection of an appropriate source model and the availability of an automatic image analysis which estimates the model parameters.

Several approaches to analysis–synthesis coding have been published. Analysis–synthesis coding of moving known objects like faces [2, 9, 45] is referred to as knowledge-based coding. Semantic coding [11, 10, 15, 16, 45] which uses a high-level description of image contents is investigated for human faces and describes the facial mimic by parameters. Knowledge-based coding and semantic coding require a recognition algorithm for the known objects. The main interest of these approaches is focussed on the human face. Whereas some systems require only the adaptation of a facial mask to the mouth and eyes of a person, other approaches assume 37 feature points of a face to be reliably detected [8]. For the human face several recognition algorithms have been presented [41], but so far just one algorithm for the adaptation of a facial mask to the mouth and eyes has been demonstrated together with an automatic image sequence analysis system [27].

OBASC in its basic form does not require any a priori knowledge of the moving objects. A first complete implementation of an object-based analysis–synthesis coder based on the model of 'flexible 2D objects with 2D motion' using a displacement vector field to describe object motion was developed by Hötter [21, 22]. This implementation shows better picture quality than conventional hybrid coding [20]. Block and mosquito artifacts are avoided. The next step towards improving the coding efficiency is to apply a more realistic source model. Therefore the concept and implementation of an OBASC based on the source model 'moving rigid 3D objects' with 3D motion is introduced in this paper. An explicit 3D description of motion and shape is used to adapt the model objects in the model world to the real objects in the real world [25, 28, 33, 34]. Hence new algorithms for modelling 3D shapes, for 3D motion estimation and for motion parameter coding had to be implemented. The coding efficiencies obtained with the source models of moving flexible 2D and moving rigid 3D objects are compared in terms of data rate required for the same picture quality. The coding schemes will be evaluated using videophone test sequences. As a well-known reference for picture quality, the block-based hybrid coder H.261 [5, 6] is used.

In Section 2, the principle of object-based analysis–synthesis coding is described. Furthermore, the source model of 'moving rigid 3D objects' is defined. Image synthesis as required for this source model is briefly presented in Section 3. Section 4 explains image analysis and gives the details of 3D motion estimation and the detection of those image areas which cannot be described by the source model. The goal of image analysis is to extract the parameter sets for an efficient description of the image sequence. Section 5 explains how these parameters are coded. In Section 6, experimental results for real image sequences are given. The final discussion in Section 7 reviews the achievements of OBASC based on the source model of 'moving rigid 3D objects' and presents possibilities to improve the efficiency of OBASC.

2. The principle of object-based analysis–synthesis coding based on the source model of moving 3D objects

2.1. *Object-based analysis–synthesis coding*

OBASC [32] subdivides each image of a sequence into uniformly moving objects and describes each object m in terms of three sets of parameters $A^{(m)}$, $M^{(m)}$ and $S^{(m)}$, defining its motion, shape and color, respectively. Motion parameters define the position and motion of the object and shape parameters define its shape. Color parameters denote the luminance as well as the chrominance reflectance on the surface of the object. In computer graphics, they are sometimes called texture. Fig. 1 is used to explain the concept and structure of OBASC. Instead of a frame memory used in block-based hybrid coding, OBASC requires a memory for parameters in order to store the coded and transmitted object parameters $A'^{(m)}$,

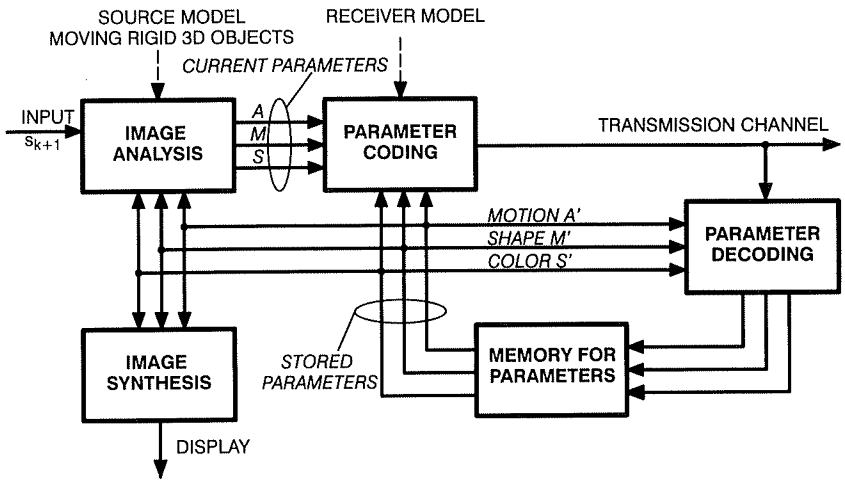


Fig. 1. Block diagram of an object-based analysis–synthesis coder.

$M'^{(m)}$ and $S'^{(m)}$. The parameter memories in the coder and decoder contain the same information. Evaluating these parameter sets, image synthesis synthesizes a model image s'_k which is displayed at the decoder. The parameter sets of the memory and the current image s_{k+1} are used as input for the image analysis. This feedback of the coded parameter sets to the analysis system is necessary in order to avoid the accumulation of coding and analysis errors.

The task of image analysis is to analyze the current image s_{k+1} to be coded and to estimate the parameter sets $A_{k+1}^{(m)}$, $M_{k+1}^{(m)}$ and $S_{k+1}^{(m)}$ of each object m by use of the decoded parameters $A_k'^{(m)}$, $M_k'^{(m)}$ and $S_k'^{(m)}$ of the preceding image. In the current image, moving and static objects are detected first. For moving objects, new motion and shape parameters are estimated in order to reuse most of the already transmitted color parameters $S_k'^{(m)}$. Objects for which motion and shape parameters can be estimated successfully are referred to as *MC-objects* (*model compliance*). In the final step of image analysis, image areas which cannot be described by MC-objects using the transmitted color parameters $S_k'^{(m)}$ and the new motion and shape parameters $A_{k+1}^{(m)}$ and $M_{k+1}^{(m)}$, respectively, are detected. These areas of *model failures* (MF) [34] are defined by 2D shape and color parameters only and are referred to as *MF-objects*. In [32] they are labelled special objects. The detection of MF-objects takes into account that small position and shape errors of the MC-objects – referred to as *geometrical distortions* – do not disturb subjective image quality. Thus MF-objects are reduced to those image regions with significant differences between the motion and shape compensated prediction image and the current image s_{k+1}. They tend to be small in size. This allows one to code color parameters of MF-objects with high quality, thus avoiding subjectively annoying quantization errors. Since the transmission of color parameters is expensive in terms of data rate, the total area of MF-objects should not be bigger than 4% of the image area assuming 64 kbit/s, CIF and 10 Hz.

In order to incorporate different properties of the source models, different types of model failures are to be distinguished. The source model 'moving rigid 3D object' will generate *rigid model failures* denoted as *MF_{3Dr}-objects* and the source model 'moving flexible 2D object' will generate *flexible model failures* denoted as *MF_{2Df}-objects*.

Depending on the *object class* MC/MF, the parameter sets of each object are coded using predictive coding techniques. Motion and shape parameters are encoded and transmitted for MC-objects and shape and color parameters for MF-objects. Since the coding of color parameters is most expensive in terms of bit-rate, parameter coding and image analysis have to be designed jointly. By minimization of the overall data rate $R = R_A + R_M + R_S$ for coding all parameter sets, a higher coding gain than in block-based hybrid coding techniques can be achieved.

Parameter decoding decodes the two parameter sets transmitted for each object class. In the memory for parameters, the position and shape of MC-objects are updated. Furthermore, in areas of model failure, color parameters of MC-objects are substituted by the color parameters of the transmitted MF-objects.

In object-based analysis–synthesis coding the suitability of source models can be judged by comparing the data rates required for coding the same image sequence with the same image quality. Image quality is influenced mainly by the algorithm for detecting model failures and by the bit-rate available for coding the color parameters of model failures.

2.2. The source model of moving rigid 3D objects

The source model used here assumes a 3D *real world* which has to be modelled by a 3D *model world*. While the *real image* is taken by a *real camera* looking into the real world, a *model image* is synthesized using a *model camera* looking into the model world. A world is described by a *scene*, its *illumination* and its camera. A scene consists of objects, their motion and their relative position. The image area representing the projection of an object m is referred to as *projection m*.

The goal of the modelling is to generate a model world W_k with a model image identical to the real image s_k at a time instance k. This implies that the model objects may differ from the real objects (Fig. 2). However, a similarity between real objects and model objects generally helps performing proper image analysis.

The real illumination is modelled by constant diffuse illumination and the real camera by a static pinhole camera whose target is the model image. This camera projects the point $\boldsymbol{P}^{(i)} = (P_x^{(i)}, P_y^{(i)}, P_z^{(i)})^{\mathrm{T}}$ of the scene onto the point $\boldsymbol{p}^{(i)} = (p_X^{(i)}, p_Y^{(i)})^{\mathrm{T}}$ of the image plane according to

$$p_X^{(i)} = F \cdot \frac{P_x^{(i)}}{P_z^{(i)}}, \qquad p_Y^{(i)} = F \cdot \frac{P_y^{(i)}}{P_z^{(i)}}, \tag{2.1}$$

where F is the focal length of the camera, (X, Y) are image coordinates and (x, y, z) are model world coordinates (Fig. 3).

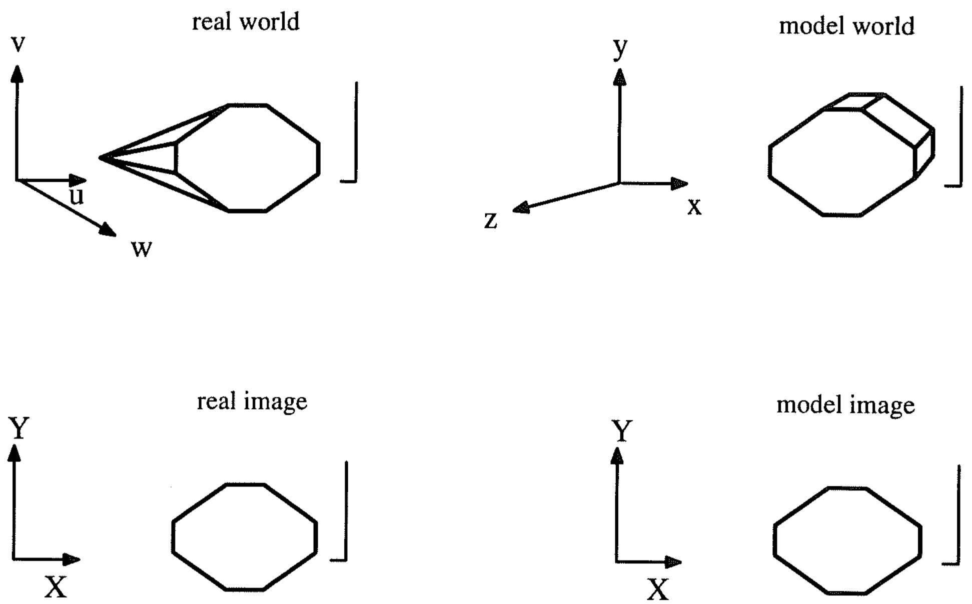


Fig. 2. A real world and a model world where real and model images are identical.

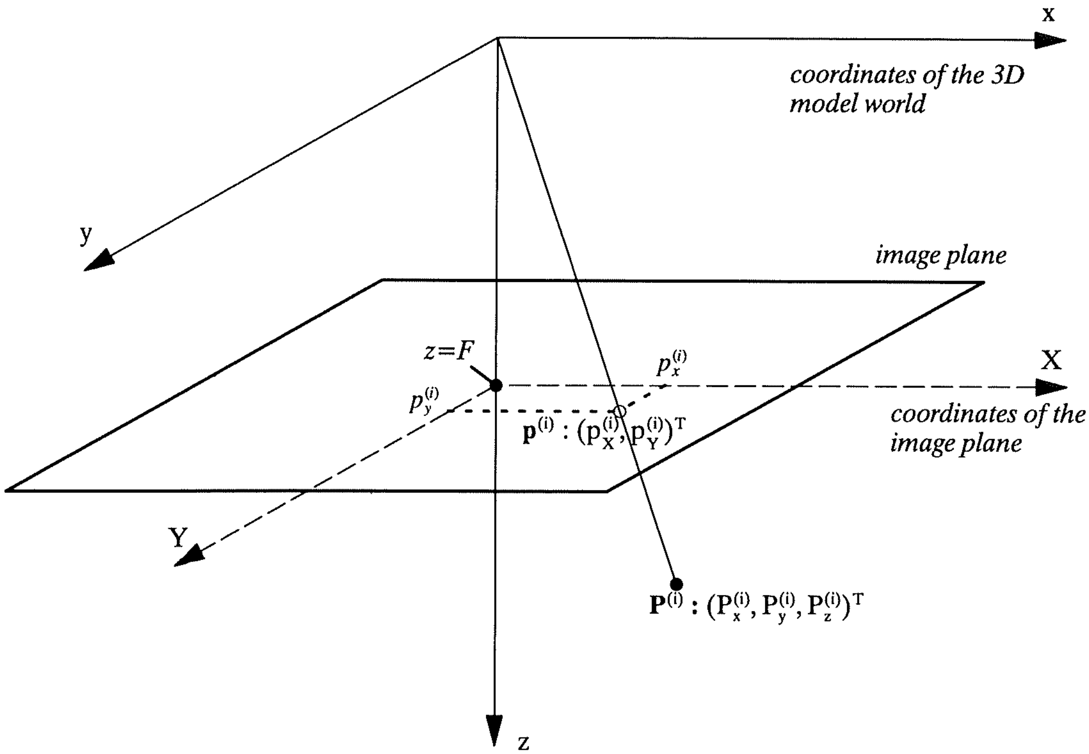
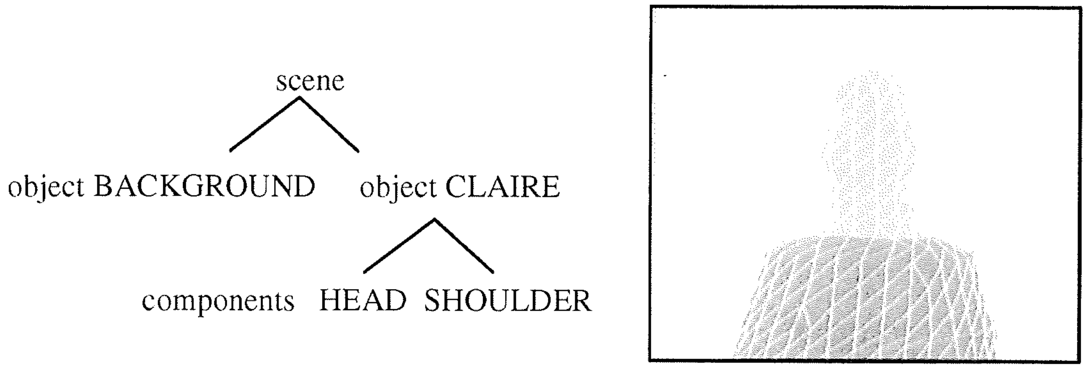


Fig. 3. Camera model.

In the model used here, the shape parameters $M^{(m)}$ of an object m represent a 2D binary mask which defines the *silhouette* of the model object in the model image. During initialization, the 3D shape of an object is completely described by its 2D silhouette, i.e. there is an algorithm which computes a 3D shape from a 2D silhouette. After initialization, the shape parameters $M^{(m)}$ are used as update parameters to the model-object shape. The 3D shape is represented by a mesh of triangles which is put up by vertices referred to as *control points* $\boldsymbol{P}_C^{(i)}$. The appearance of the model-object surface is described by the color parameters $S^{(m)}$, which define the luminance and chrominance reflectance. An object may consist of two or more rigid *components* [4]. Each component has its own set of motion parameters. Since each component is defined by its control points, the components are linked by those triangles of the object having control points belonging to different components. Due to these triangles, components are flexibly connected. Fig. 4 shows a scene with the objects BACKGROUND and CLAIRE. The model object CLAIRE consists of the two components HEAD and SHOULDER. One special model object BACKGROUND, which is a non-moving rigid 2D plane, is assigned to each scene. For BACKGROUND, only color parameters have to be coded because motion and shape parameters are fixed. When referring to model objects in this paper, BACKGROUND is not considered.

The model objects of the scene are opaque. Their 3D motion is described by the parameters $A^{(m)} = (T_x^{(m)}, T_y^{(m)}, T_z^{(m)}, R_x^{(m)}, R_y^{(m)}, R_z^{(m)})$ defining translation and rotation. A point $\boldsymbol{P}^{(i)}$ on the surface of object m with N control points $\boldsymbol{P}_C^{(i)}$ is moved to its new position $\boldsymbol{P}'^{(i)}$ according to

$$\boldsymbol{P}'^{(i)} = [\boldsymbol{R}_C^{(m)}] \cdot (\boldsymbol{P}^{(i)} - \boldsymbol{C}^{(m)}) + \boldsymbol{C}^{(m)} + \boldsymbol{T}^{(m)}, \qquad (2.2)$$



(a) (b)

Fig. 4. Model scene and model object CLAIRE subdivided into two flexibly connected components: (a) scene consisting of two objects; (b) components of model object CLAIRE.

with the translation vector $\boldsymbol{T}^{(m)} = (T_x^{(m)}, T_y^{(m)}, T_z^{(m)})^{\mathrm{T}}$, the object center $\boldsymbol{C} = (C_x, C_y, C_z) = (1/N)\sum_{i=1}^{N} \boldsymbol{P}_C^{(i)}$, the rotation angles $\boldsymbol{R}_C = (R_x^{(C)}, R_y^{(C)}, R_z^{(C)})^{\mathrm{T}}$ and the rotation matrix $[\boldsymbol{R}_C]$ defining the rotation in the mathematically positive direction around the x-, y- and z-axis with the rotation center $\boldsymbol{C}$:

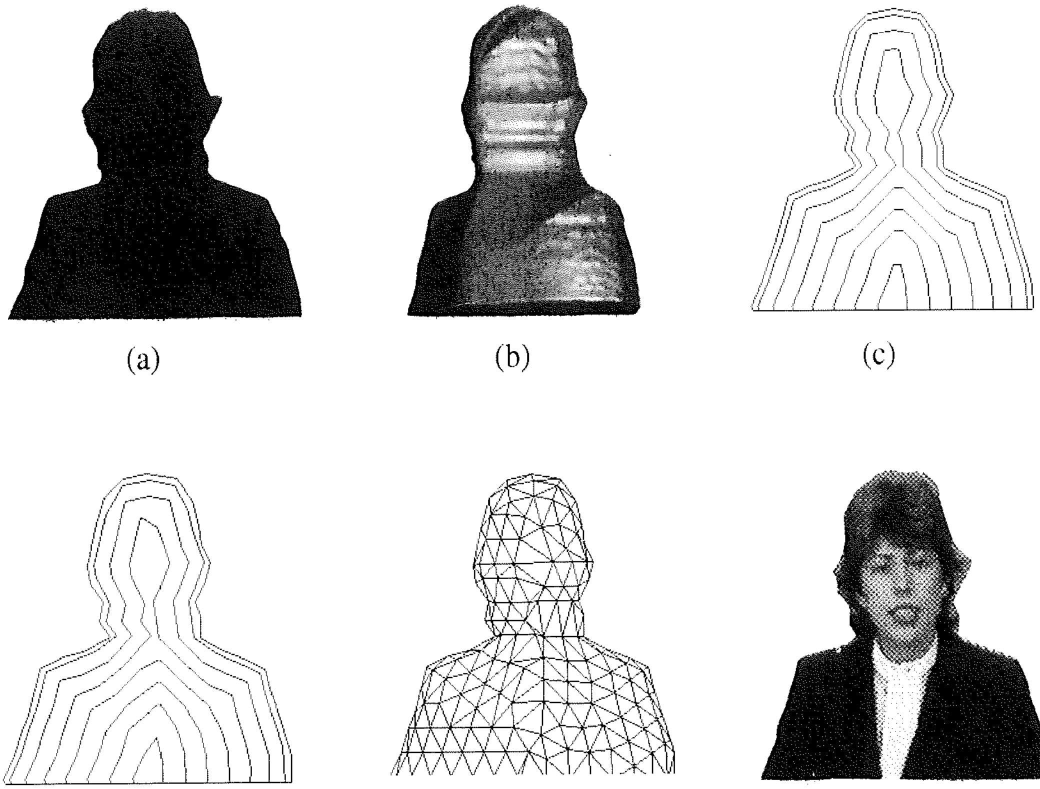
$$[\boldsymbol{R}_C] = \begin{bmatrix} \cos R_y \cos R_z, & \sin R_x \sin R_y \cos R_z - \cos R_x \sin R_z, & \cos R_x \sin R_y \cos R_z + \sin R_x \sin R_z \\ \cos R_y \sin R_z, & \sin R_x \sin R_y \sin R_z + \cos R_x \cos R_z, & \cos R_x \sin R_y \sin R_z - \sin R_x \cos R_z \\ -\sin R_y, & \sin R_x \cos R_y, & \cos R_x \cos R_y \end{bmatrix}. \tag{2.3}$$

In contrast to the motion equation $\boldsymbol{P}'^{(i)} = [\boldsymbol{R}] \cdot \boldsymbol{P}^{(i)} + \boldsymbol{T}$, Eq. (2.2) rotates an object around its center $\boldsymbol{C}$ and then moves the object by the translation $\boldsymbol{T}$. Hence $\boldsymbol{R}$ and $\boldsymbol{T}$ are independent of the camera position.

2.2.1. *Generation of a model object*

When the silhouette of an object is known, a 3D object has to be modelled from its silhouette. The algorithm presented in the following generates a natural-looking 3D shape from a silhouette. Fig. 5 illustrates the five steps from the silhouette (Fig. 5(a)) towards a model object (Fig. 5(f)) with the given silhouette.

In order to generate a 3D shape from an object silhouette, a distance transformation is applied to the silhouette. An ellipse is used as a generating function giving the z-coordinate of the shape for each point of the silhouette as a function of its distance to the boundary of the silhouette (Fig. 6). This generating function may be different for specific scenes. However, an ellipse is a good initial guess for most applications. Fig. 5(b) shows a view of the 3D shape which is generated from the silhouette of Fig. 5(a) using the distance transformation. In Fig. 5(b) the 3D shape is rotated and illuminated in order to give a better impression of its shape. In contrast to algorithms which rotate the silhouette [25], this use of a generating function for computing the z-dimension allows one to generate a model object from an arbitrarily shaped silhouette. The parameters defining the major axis of the ellipse are the object width in the image plane and the maximum object depth perpendicular to the image plane. For videophone test sequences, subjectively good results are achieved with the ratio of object



(a) (b) (c)

(d) (e) (f)

Fig. 5. Processing steps from object silhouette to model object: (a) object silhouette; (b) 3D object shape with required silhouette rotated by 30° and illuminated; (c) contour lines approximating the object shape; (d) polygons approximating the contour lines; (e) mesh of triangles using polygon points as vertices; (f) model object with color parameters projected onto.

width to object depth set to 1.5 (Fig. 6). The resulting 3D shape is approximated by contour lines. Contour lines are situated on the 3D shape. The distance along the surface between two contour lines is constant (Figs. 5(c) and 6). This enables the algorithm to compute a homogeneous mesh of triangles. In a subsequent step, contour lines are approximated by polygons (Fig. 5(d)). For approximation of the contour lines by polygons, the same error criterion d^*_{max} which will later be used for 2D shape coding is applied. d^*_{max} defines the maximum allowable distance between a contour line and its approximating polygon [18]. Experiments on videophone sequences with the spatial resolution according to CIF have shown that $d^*_{max} = 1.4$ pel measured in the image plane is a subjectively sufficient quality [21]. The resulting polygon points are the control points of a mesh of triangles (Fig. 5(e)). The model object defined by this mesh of triangles is then placed in the model world. The original image s_1 is now projected into the model scene using the geometry of the model camera. This defines the color parameters of the new model object and of the BACKGROUND.

3. Image synthesis

In OBASC, image synthesis is required for displaying the decoded picture at the receiver, for

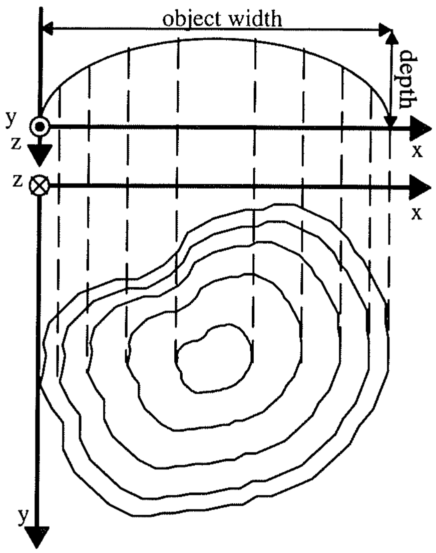


Fig. 6. Function giving z-coordinates of the 3D shape and contour lines.

image analysis and for the memory of object parameters in order to replace color parameters of MC-objects if MF-objects are transmitted.

Image synthesis uses the z-buffer algorithm [38] for selecting visible object parts. The texture of each visible triangle of a model object is mapped into the model image using an affine transformation [36, 45]. Due to the discrete sampling grid of the model image, aliasing occurs at object boundaries. It is suppressed by the Fourier window method [43] using the mathematical description of the edges of the triangles.

4. Image analysis

The goal of image analysis is to gain a compact description of the current real image s_{k+1} taking the transmitted parameter sets $A_k'^{(m)}, M_k'^{(m)}, S_k'^{(m)}$ and subjective image quality requirements into account. The image analysis consists of the following parts: image synthesis, change detection, 3D motion estimation, detection of object silhouettes, shape adaptation and model failure detection. Whereas Section 4.1 gives a short overview of image analysis, the following sections describe 3D motion estimation and detection of model failures in more detail.

4.1. Overview of image analysis

Fig. 7 shows the structure of image analysis. Input to image analysis are the current real image s_{k+1} and the model world W_k' described by its parameters $A_k'^{(m)}$, $M_k'^{(m)}$ and $S_k'^{(m)}$ for each object m. First, a model image s_k' of the current model world is computed by means of image synthesis.

In order to compute the change detection mask B_{k+1}, the change detection evaluates the images s_k' and s_{k+1} on the hypothesis that moving real objects generate significant temporal changes in the images [44, 17], that they have occluding contours [33] and that they are opaque [33]. This mask B_{k+1} marks the projections of moving objects and the background uncovered due to object motion as changed. However, areas of moving shadows or illumination changes are not marked as changed because illumination changes can be modelled by semitransparent objects. This is in contradiction to the assumption that moving real objects are opaque. Since change detection accounts for the silhouettes of the model objects, the changed areas in mask B_{k+1} will be at least as large as these silhouettes. Fig. 8 gives the change detection mask B_{k+1} for the second and 18th frame of the test sequence *Claire* [7]. Due to little motion of Claire at the beginning of the sequence, only parts of the projection of the real object are detected during the first frames. However, this does not disturb the quality of the decoded pictures.

In order to compensate for real object motion and in order to separate the moving objects from the uncovered background, 3D motion parameters are estimated for the model objects [1, 26, 27, 29]. The applied motion estimation algorithm requires motion, shape and color parameters of the model objects and the current real image s_{k+1} as input. Motion parameters A_{k+1} are estimated in such a way that the mean square error between the projections of the model objects and the projections of the real objects in the real image s_{k+1} is minimized.

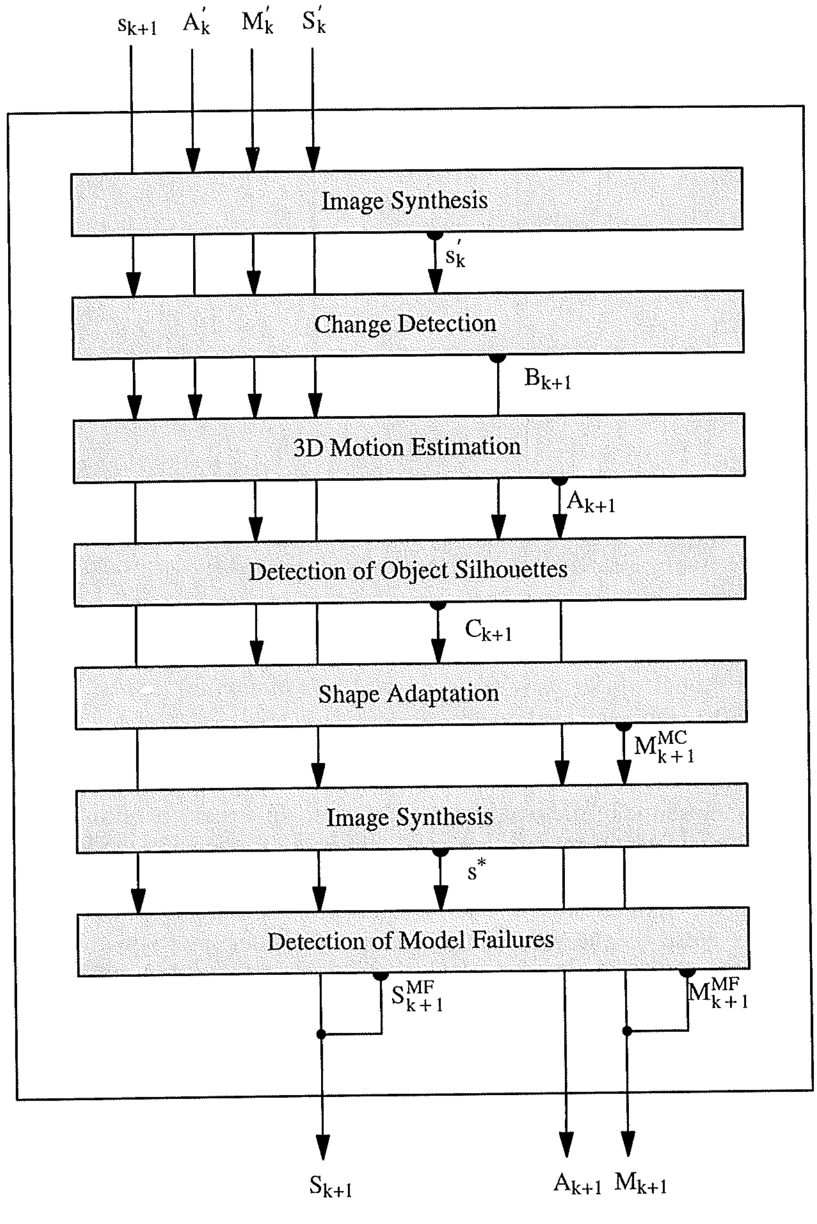


Fig. 7. Block diagram of image analysis: A'_k, M'_k, S'_k stored motion, shape and color parameters; s_{k+1} real image to be analyzed; s'_k, s^* model images; B_{k+1} change detection mask; C_{k+1} object silhouettes; M_{k+1} shape parameters for MC- and MF-objects; S_{k+1} color parameters for MC- and MF-objects. Arrows indicate the information used in various parts of image analysis. Semicircles indicate the output of processing steps.

The resulting motion parameters are used to detect the uncovered background, which is included in mask B_{k+1}. The basic idea for the detection of uncovered background is that the projection of the moving object before and after motion has to lie completely in the changed area [17]. Subtracting the uncovered background from mask B_{k+1} gives the new silhouette C_{k+1} of all model objects (Fig. 8).

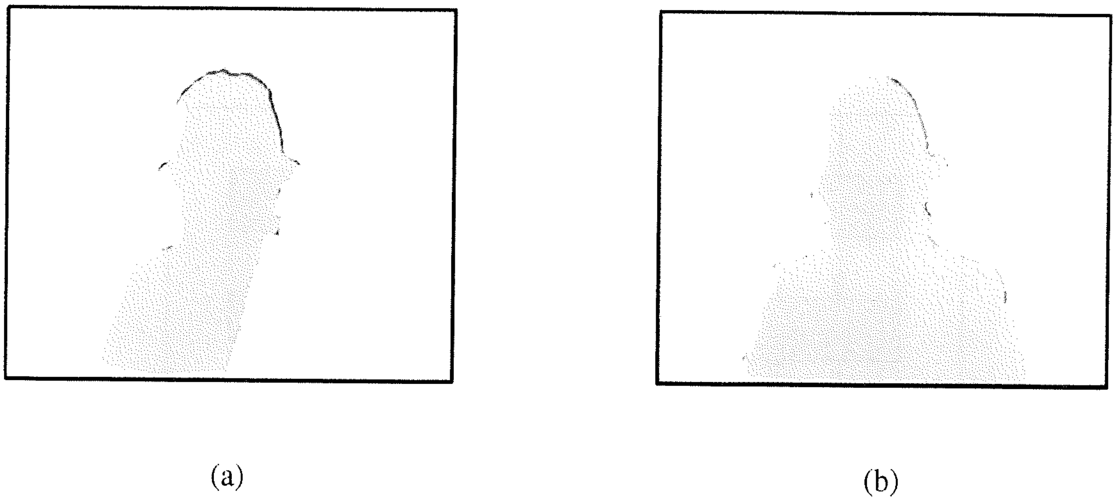


Fig. 8. Object silhouettes of CLAIRE: (a) frame 2; (b) frame 18. Gray and black areas are marked changed by change detection. Detection of object silhouettes gives the object silhouette (gray) and the uncovered background (black).

The silhouette of each model object m is then compared and adapted to the real silhouette $C_{k+1}^{(m)}$ [33]. Differences occur either when parts of the real object start moving for the first time or when differences between the shape of the real and the model object become visible during rotation. In order to compensate for the differences between the silhouettes of the model objects and C_{k+1}, the control points close to the silhouette boundary are shifted perpendicular to the model-object surface so that the model object gets the required silhouette. This gives the new shape parameters M_{k+1}^{MC}, where MC denotes model compliance.

For the detection of model failures, a model image s^* is synthesized using the previous color parameters S_k' and the current motion and shape parameters A_{k+1} and M_{k+1}^{MC}, respectively. The differences between the images s^* and s_{k+1} are evaluated for determining the areas of model failure. The areas of model failure cannot be compensated using the source model of 'moving rigid 3D objects'. Therefore, they are named *rigid model failures* $\mathrm{MF}_{3\mathrm{Dr}}$ and are represented by $\mathrm{MF}_{3\mathrm{Dr}}$-objects. These MF-objects are described by 2D shape parameters M_{k+1}^{MF} and color parameters S_{k+1}^{MF} only.

4.2. 3D motion estimation

It is assumed that differences between two consecutive images s_k and s_{k+1} are due to object motion. In order to estimate these motion parameters, a gradient method is applied here. A first version of this algorithm has been successfully applied to test sequences like *Claire* [7], *Michael* [12] and *Miss America* [3] with a manual initialization of the model world in 1987 [25] and 1988 [26].

From each model object, motion estimation uses one set of observation points. Each observation point $\boldsymbol{O}^{(j)} = (\boldsymbol{P}^{(j)}, \boldsymbol{g}^{(j)}, I^{(j)})$ is located on the model-object surface at position $\boldsymbol{P}^{(j)}$ and holds its luminance value $I^{(j)}$ and its linear gradients $\boldsymbol{g}^{(j)} = (g_x^{(j)}, g_y^{(j)})^{\mathrm{T}}$. $\boldsymbol{g}$ are the horizontal and vertical luminance gradients from the image which provided the color parameters for the object. The measure for selecting observation points is a high spatial gradient. This adds robustness against noise to the estimation algorithm. Fig. 9 shows the location of all observation points belonging to model object CLAIRE.

Motion estimation minimizes the mean square difference between the model image and the real image. It is assumed that objects are rigid and have diffuse reflecting surfaces. Furthermore, diffuse

Fig. 9. The position of all observation points of model object CLAIRE.

illumination of the scene is assumed. Hence, color parameters are constant. With an observation point $O_k^{(j)} = (P_k^{(j)}, g^{(j)}, I^{(j)})$ at time instant k projected onto the image plane at $p_k^{(j)}$ and the same observation point after motion $O_{k+1}^{(j)} = (P_{k+1}^{(j)}, g^{(j)}, I^{(j)})$ projected onto $p_{k+1}^{(j)}$, the luminance difference between image s_k and image s_{k+1} at position $p_k^{(j)}$ is then related to motion by

$$\begin{aligned}\Delta I^{(j)} &= s_{k+1}(p_k^{(j)}) - s_k(p_k^{(j)}) \\ &= (g_x^{(j)}, g_y^{(j)})^{\mathrm{T}} \cdot (p_{k+1}^{(j)} - p_k^{(j)}).\end{aligned} \tag{4.1}$$

Substituting image coordinates by model world coordinates with Eq. (2.1) yields

$$\begin{aligned}\Delta I^{(j)} = {} & F \cdot g_x^{(j)} \left[\frac{P_{x,k+1}^{(j)}}{P_{z,k+1}^{(j)}} - \frac{P_{x,k}^{(j)}}{P_{z,k}^{(j)}} \right] \\ & + F \cdot g_y^{(j)} \left[\frac{P_{y,k+1}^{(j)}}{P_{z,k+1}^{(j)}} - \frac{P_{y,k}^{(j)}}{P_{z,k}^{(j)}} \right].\end{aligned} \tag{4.2}$$

The position $P_k^{(j)}$ of the observation point $O^{(j)}$ is known. By relating P_k to P_{k+1} by means of the motion equation (2.2), a non-linear equation with the known parameters ΔI, g and F and the six unknown motion parameters results. This equation is linearized by linearizing the rotation matrix R_C (2.3) assuming small rotation angles

$$[R'_C] = \begin{bmatrix} 1 & -R_z & R_y \\ R_z & 1 & -R_x \\ -R_y & R_x & 1 \end{bmatrix} \tag{4.3}$$

giving

$$P_{k+1} = [R'_C] \cdot (P_k - C) + C + T. \tag{4.4}$$

Substituting Eq. (4.4) in (4.2) the linearized equation for one observation point is

$$\begin{aligned}\Delta I = {} & F \cdot g_x / P_z \cdot T_x \\ & + F \cdot g_y / P_z \cdot T_y \\ & - [(P_x g_x + P_y g_y) F / P_z^2 + \Delta I / P_z] \cdot T_z \\ & - [[P_x g_x (P_y - C_y) + P_y g_y (P_y - C_y) \\ & + P_z g_y (P_z - C_z)] F / P_z^2 \\ & + \Delta I / P_z (P_y - C_y)] \cdot R_x \\ & + [[P_y g_y (P_x - C_x) + P_x g_x (P_x - C_x) \\ & + P_z g_x (P_z - C_z)] F / P_z^2 \\ & + \Delta I / P_z (P_x - C_x)] \cdot R_y \\ & - [g_x (P_y - C_y) - g_y (P_x - C_x)] F / P_z \cdot R_z,\end{aligned} \tag{4.5}$$

with the unknown motion parameters $T = (T_x, T_y, T_z)^{\mathrm{T}}$ and $R_C = (R_x, R_y, R_z)^{\mathrm{T}}$ and the observation point $O_k = (P_k, g, I)$ at position $P_k = (P_x, P_y, P_z)^{\mathrm{T}}$.

In order to get reliable estimates for the six motion parameters, Eq. (4.5) has to be established for several hundred observation points. The residuum of this equation system is minimized by linear regression:

$$\sum_{O^{(j)}} (\Delta I^{(j)})^2 \to \mathrm{MIN}. \tag{4.6}$$

In order to make the estimation more robust, the variance σ_I of all the residuals ΔI according to Eq. (4.5) is computed. Equations with $\Delta I > \sigma_I$ are discarded [17]. Instead of linear regression, other methods can be used [30, 40, 46].

Due to linearization, motion parameters have to be estimated iteratively for each model object. After every iteration, the model object is moved according to Eq. (2.2) using the estimated motion parameters. After this, a new set of motion equations is established, giving new motion parameter updates. Since the motion parameter updates approach zero during the iterations, the introduced linearizations do not harm motion estimation.

4.3. Detection of model failures

As a final step of image analysis, the estimated shape and motion parameters are verified. The goal of this verification test is to detect those regions of the current image s_{k+1} which cannot be described by the previously transmitted color parameters S_k' and the current motion and shape parameters A_{k+1} and M_{k+1}^{MC}, respectively. Therefore, model objects are motion and shape compensated, giving the synthesized prediction image s^*. The difference image between s^* and the current image s_{k+1} is evaluated by binarizing it using an adaptive threshold T_e so that the error variance of the areas which are not

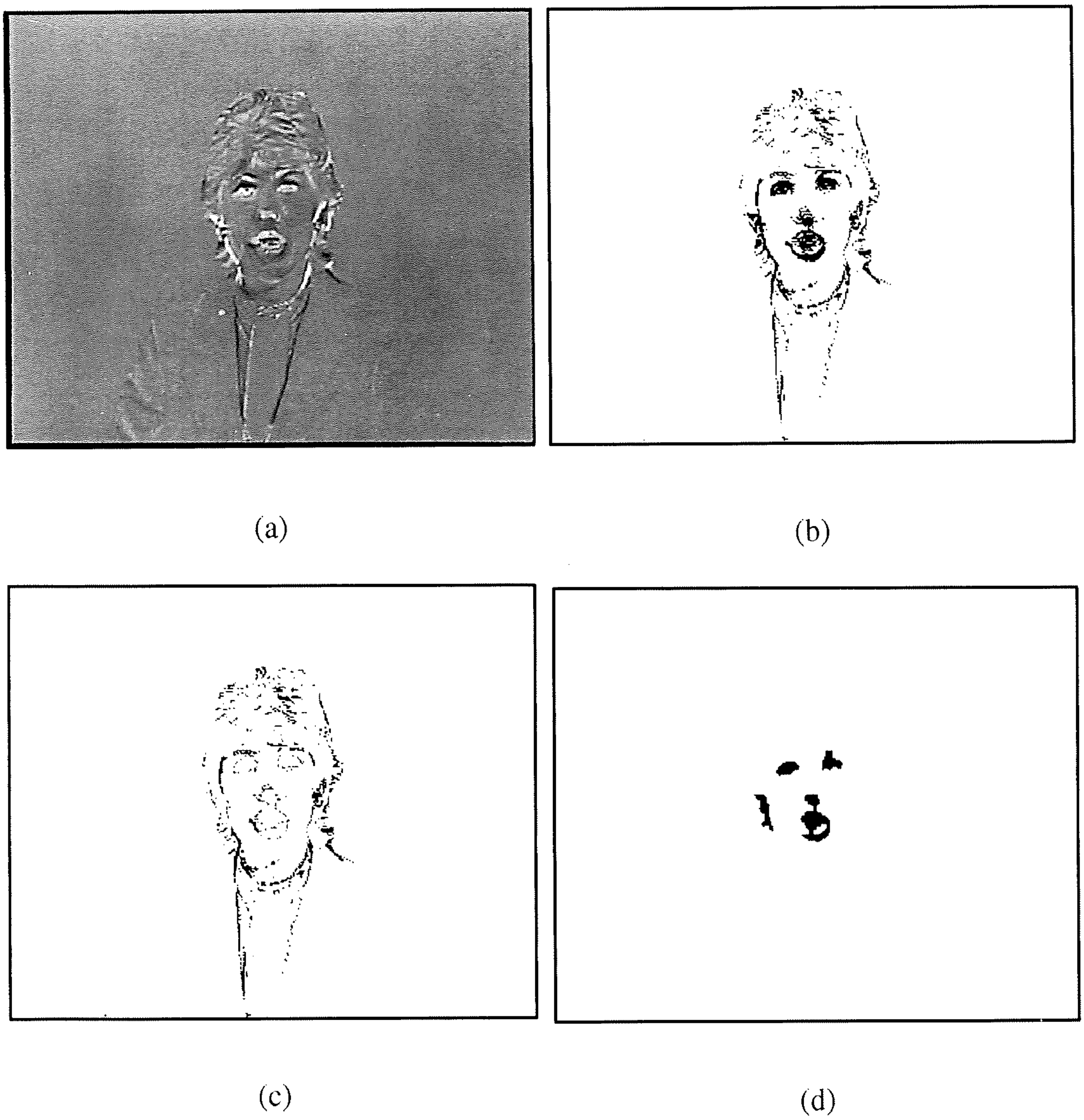


Fig. 10. Detection of model failures: (a) scaled difference image between real and model image after motion and shape compensation; (b) synthesis error mask; (c) geometrical distortions and perceptually irrelevant regions; (d) mask $\mathrm{MF}_{3\mathrm{Dr}}$ with model failures of the source model rigid 3D object.

declared as synthesis errors is below a given allowed noise level N_e. The resulting mask is named synthesis error mask. Figs. 10(a) and 10(b) shows a scaled difference image and the resulting synthesis error mask, respectively.

The synthesis error mask marks those pels of image s^* which differ significantly from the corresponding pels of s_{k+1}. Since the areas of synthesis errors are frequently larger than 4% of the image area, it is not possible to transmit color parameters for these areas with a sufficiently high image quality, i.e. visible quantization errors would occur. However, from a subjective point of view it is not necessary to transmit color parameters for all areas of synthesis errors. Due to the object-based image description, the prediction image s^* is subjectively pleasant. There are no block artifacts and object boundaries are synthesized properly.

There are two major reasons for synthesis errors. First of all, synthesis errors are due to position and shape differences between a moving real object and its corresponding model object. These errors are caused by motion and shape estimation errors. They displace contours in the image signal and will produce line structures in the synthesis error mask. Due to the feedback of the estimated and coded motion and shape parameters into image analysis (Fig. 1), these estimation errors tend to be small, unbiased and they do not accumulate. Therefore it is reasonable to assume that these errors do not disturb subjective image quality. They are classified as geometrical distortions. As a simple detector of geometrical distortions, a median filter of size 5×5 pel is applied to the mask of synthesis errors (Fig. 10(c)).

Secondly, events in the real world which cannot be modelled by the source model will contribute to synthesis errors. Using the source model of 'moving rigid 3D objects', it is not possible to model changing human facial expressions or specular highlights. Especially facial expressions are subjectively important. In order to be of subjective importance, it is assumed that an erroneous image region has to be larger than 0.05% of the image area (Fig. 10(c)). Model failures are those image areas where the model image s^* is subjectively wrong (Fig. 10(d)). Each area of model failure is modelled by an MF_{3Dr}-object defined by color and 2D shape parameters.

5. Parameter coding

The task of parameter coding is the efficient coding of the parameter sets defining motion, shape and color provided by image analysis (Fig. 11). Parameter coding uses a coder-mode control to select the appropriate parameter sets to be transmitted for each object class. The priority of the parameter sets is arranged by a priority control.

5.1. Motion parameter coding

The unit of the estimated object translation $\boldsymbol{T} = (T_x, T_y, T_z)^{\mathrm{T}}$ is pel. The unit of the estimated object rotation $\boldsymbol{R}_C = (R_x^{(C)}, R_y^{(C)}, R_z^{(C)})^{\mathrm{T}}$ is degree. These motion parameters are PCM coded by quantizing each component with 8 bits within an interval of ± 10 pels and degrees, respectively. This ensures a subjectively lossless coding of motion parameters.

5.2. 2D shape parameter coding

Since the model-object shape is computed from its silhouette, shape parameters are essentially 2D. The principles for coding shape parameters of MF- and MC-objects are identical. Shape parameters are coded using a polygon/spline approximation developed by Hötter [18]. A measure $d_{\max}$ describes the maximum distance between original and approximated shape. First, an initial polygon approximation of the shape is generated using four points

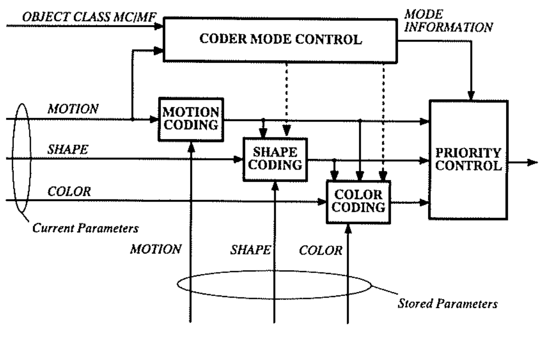


Fig. 11. Block diagram of parameter coding.

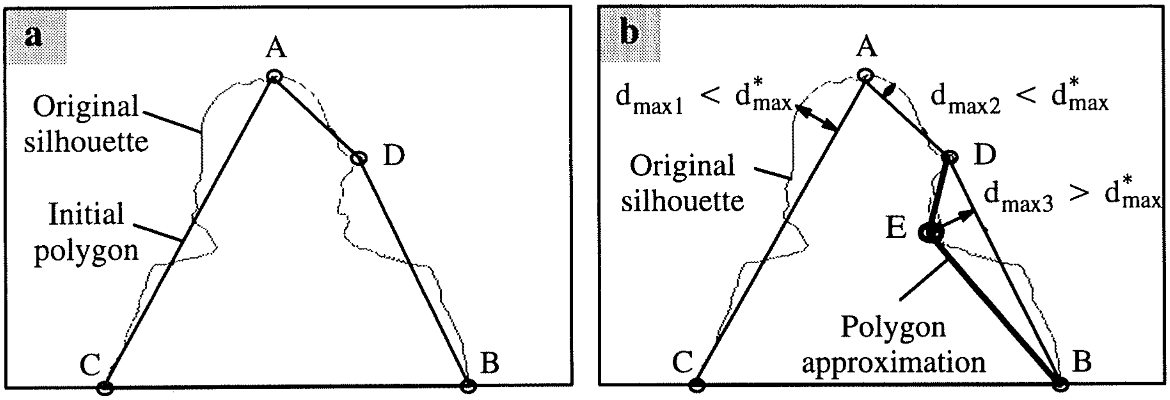


Fig. 12. Polygon approximation: (a) initial polygon; (b) insertion of a new polygon point (from [22]).

(Fig. 12(a)). Where the quality measure d^*_{max} is not satisfied, the approximation is iteratively refined through insertion of additional polygon points until the measure fulfills $d_{max} \leqslant d^*_{max}$ (Fig. 12(b)). In a final step for each line of the polygon, it is checked whether an approximation of the corresponding contour piece by a spline approximation also satisfies d^*_{max}. If so, the spline approximation is used giving a natural shape approximation for curved shapes (Fig. 13). In order to avoid visible distortions at object boundaries, MC-objects are coded with $d^*_{max} = 1.4$ pels. Experimental results showed that MF-objects should be coded with d^*_{max} about 2.1 pels [34]. The coordinates of the polygon points are coded relative to their perspective predecessor. Then the curve type line/spline is coded for each line of the polygon.

The data rate for coding shape parameters of MC-objects is cut to half by using the motion compensated coded silhouette of the last image as a prediction of the current silhouette. Starting with this approximation, only shape update parameters have to be transmitted.

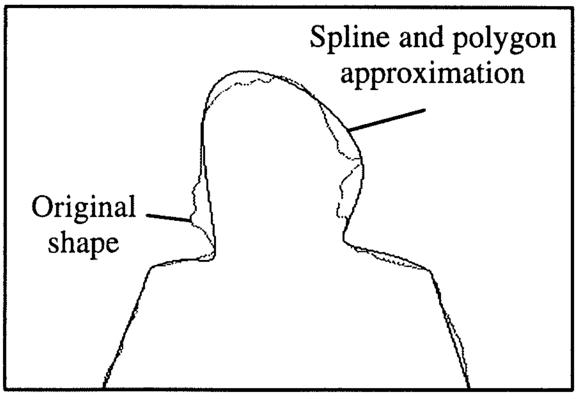


Fig. 13. Combination of polygon and spline approximation for a quality measure $d^*_{max} = 15$ (from [22]).

5.3. Color parameters

Conventional DCT is not suitable for the coding of color parameters of arbitrarily shaped regions. New algorithms have been developed for this application [14, 42]. Here the special type of DCT for arbitrarily shaped regions developed by Gilge [14] is improved by applying a segmentation of the color parameters into homogeneous regions prior to transform coding [35]. The segmentation is based on the minimum spanning tree [31] using the signal variance as criterion. The boundaries of the regions are coded using a chain code [13]. The DCT coefficients are quantized with a linear quantizer of signal-dependent step size. The advantage of this scheme using segmentation prior to transform coding is that errors due to coarse quantization are mainly concentrated at the boundaries of the segmented regions, where they are less visible due to masking of the human visual system in areas of high local activity.

5.4. Control of parameter coding

Due to limited data rate, a transmission of all parameter sets cannot be guaranteed. Coder control is used to overcome this difficulty. It consists of coder-mode control and priority control (Fig. 11). Coder-mode control selects the relevant parameter sets and coder adjustments for each object, and priority control arranges these parameter sets for transmission.

Depending on model-object class MF or MC, coder-mode control selects two parameter sets for transmission. For MC-objects, only motion $A_{k+1}^{(m)}$ and shape update parameters $M_{k+1}^{(m)}$ are coded. Coding of color parameters is not necessary because the existing color parameters $S_k^{(m)}$ of the model objects are sufficient to synthesize the image properly. Two-dimensional shape parameters, defining the location of the model failures in the image plane, and color parameters are coded for MF-objects.

Priority control guarantees that motion parameters of all MC-objects are transmitted first. In a second step, shape parameters of MC-objects are transmitted. Finally, the shape and color parameters of MF-objects are transmitted until the available data rate is exhausted.

6. Experimental results

In the sequel, the object-based analysis–synthesis coder based on the source model of 'moving rigid 3D objects' is applied to the test sequences *Claire* [7] and *Miss America* [3] with a spatial resolution corresponding to CIF and a frame rate of 10 Hz. The results are compared to those of an H.261 coder [5, 6] and OBASC based on the source model of 'moving flexible 2D objects' presented by Hötter [20–22]. The source model of 'moving flexible 2D objects' uses a displacement vector field for a joint description of object motion and local flexible deformation. As far as detection of model failures and coding of shape parameters are concerned, the same algorithms and coder adjustments are applied. Parameter coding is aiming at a data rate of approximately 64 kbit/s. However, the bit-rate of the coder is not controlled and no buffer is implemented. In the experiment, the allowed noise level N_e for detection of model failures was set to 6/255. Color parameters of model failures are coded according to Section 5.3 with PSNR = 36 dB. In all experiments the coders are initialized with the first original image of the sequence, i.e. for the block-based coder H.261 the frame memory is initialized with the first original image. For the two object-based analysis–synthesis coders, the model object BACKGROUND in the memory for parameters is initialized with the first original image.

For head and shoulder scenes the 3D model object is usually divided into two to three components. Applying the estimated motion parameter sets to the model object gives a natural impression of object motion. This indicates that the estimated motion parameters are close to the real motion parameters and that the distance transform applied to the object silhouette for generating the 3D model-object shape is suitable for the analysis of head and shoulder scenes.

The area of rigid model failures MF_{3Dr} is on average below 4% of the image area (Fig. 14). Generalizing this, for head and shoulder scenes the rigid model failures can perhaps be expected to be below 15% of the moving area. The exact figures of model failure area as related to moving area are 12% for *Claire* and 7% for *Miss America*. The test sequence *Claire* seems to be more demanding, due to the fast rotation of her head, whereas *Miss America*'s motion is almost 2D.

Table 1 compares the average bit-rate for the different parameter sets defining motion, shape and color and the two source models of 'moving rigid 3D objects' and 'moving flexible 2D objects'. Both coders need approximately the same data rate. Due to the displacement vector field, OBASC based on the source model of 'moving flexible 2D objects' requires a relatively high amount of motion information. Shape parameters include the shape of MC- and MF-objects. Shape parameters of MC-objects require similar data rates for both source models. However, the source model of 'moving flexible 2D objects' causes only a few large MF-objects whereas the source model of 'moving rigid 3D objects' causes smaller but more MF-objects. This larger number of MF_{3Dr}-objects is due to the applied source model assuming rigid shapes. Shape

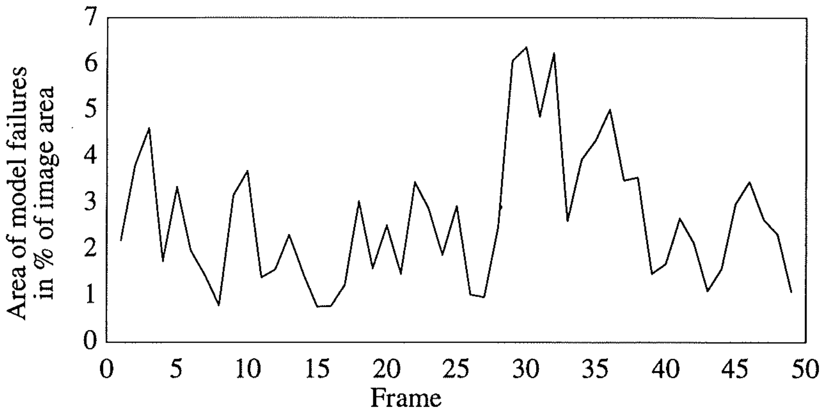


Fig. 14. Area of rigid model failures MF_{3D_r} in pels for the test sequence *Claire*. The total area is 101 376 pels. The average area of model failures is 3.5% of the image area.

Table 1
Average bit-rate of parameter sets for the source models of 3D rigid objects and 2D flexible objects (all figures in bit/frame). The coders use the same algorithm for detection of model failures. The bit-rate for coding of color parameters is 1.2 bit/pel

	Motion	Flexible shape	MC shape	MF shape	$\sum$ shape	MF-color	$\sum$
Rigid 3D object	200	—	550	1100	1650	4400	6250
Flexible 2D object	1100		450	850	1300	4000	6400

differences between real and model objects as well as small flexible motion on the surface of real objects cannot be compensated for. These effects cause small local position errors of the model objects. If texture with high local activity is displaced for more than 0.5 pels, model failures are detected. Since these small position errors can be compensated for when using the source model of 'moving flexible 2D objects', the overall data rate for shape parameters of MF-objects is 250 bits higher for the source model of 'moving rigid 3D objects'.

Fig. 15 shows the 33rd decoded frame of the test sequence *Claire* using the source model of 'moving rigid 3D objects' and of 'moving flexible 2D objects'. Subjectively, there is no difference between the two decoded image sequences. When compared to decoded images of an H.261 coder [6], picture quality is improved twofold (Fig. 16): At boundaries of moving objects, no block or mosquito artefacts are visible, due to the introduction of shape parameters. Image quality in the face is improved, because coding of color parameters is limited to model failures which are mainly located in the face. Since the average area of model failures, i.e. the area where color parameters have to be coded, covers 4% of the image area, color parameters can be coded at a data rate higher than 1.0 bit/pel. This compares to 0.1–0.4 bit/pel available for coding of color parameters with an H.261 (RM8) coder.

It can be expected from the experience with the source models of 'moving rigid 2D objects' and of 'moving flexible 2D objects' [19] that the introduction of the source model of 'moving flexible 3D objects' will reduce the area of model failures significantly. Although applying the source model of 'moving flexible 3D objects' will require the additional coding of flexible shape parameters, this will be overcompensated by the reduction of the bit-rate for coding model failures.

(a)

(b)

Fig. 15. 33rd decoded frame of test sequence *Claire* at a data rate of 64 kbit/s using the source model of (a) 'moving rigid 3D objects' and of (b) 'moving flexible 2D objects'.

Fig. 16. 33rd decoded frame of test sequence *Claire* using H.261 at a data rate of 64 kbit/s.

7. Conclusions

In this paper the concept and implementation of an object-based analysis–synthesis coder based on the source model of 'moving rigid 3D objects' aiming at a data rate of 64 kbit/s has been presented. The coder consists of five parts: image analysis, parameter coding and decoding, image synthesis and memory for parameters. Each object is defined by its uniform 3D motion and described by motion, shape and color parameters. Moving objects are modelled by 3D model objects.

The goal of image analysis is to arrive at a compact parametric description of the current image of a sequence, taking already transmitted parameter sets into account. Image analysis includes shape and motion estimation as well as detection of model failures. Moving objects are segmented using temporal change detection and motion parameters. The algorithm for estimating these motion parameters is based on previous work on gradient-based motion estimation. A new set of equations relating the difference signal between two images to 3D motion parameters has been established. The 3D shape of a model object is computed by applying a distance transformation, giving object depth, to the object silhouette. Since the estimated motion parameters applied to these 3D model objects give a natural impression of motion, this distance transform is very suitable for analysis of head and shoulder scenes.

Those areas of an image which cannot be modelled by the applied source model are referred to as model failures and modelled by MF-objects. They are described by color and 2D shape parameters only. Model failures are detected taking subjective criteria into account. It is assumed that geometrical distortions like small position and shape errors of the moving objects do not disturb subjective image quality. Due to these subjective criteria, the average area of model failures is below 4% of the image area for typical videophone test sequences.

With respect to coding, shape and color parameters are coded for MF-objects, while motion and shape update parameters have to be coded for MC-objects (model compliance). Motion parameters are PCM coded and shape parameters are coded using a polygon/spline approximation. Prior to coding, color parameters are segmented into homogeneous regions. Then a DCT for arbitrarily shaped regions is applied.

The presented coder has been compared to OBASC based on the model of 'moving flexible 2D objects'. With regard to typical videophone test sequences it is shown that the picture quality at the average bit-rate of 64 kbit/s is the same regardless of whether the source model of 'moving rigid 3D objects' or that of 'moving flexible 2D objects' is applied. When compared to images coded according to H.261, there are no mosquito and no block artefacts because the average area for which color parameters are transmitted is 10% of the image area for H.261 and 4% for OBASC. Therefore OBASC allows coding of color parameters for MF-objects with a data rate higher than 1.0 bit/pel. At the same time, MC-objects are displayed without subjectively annoying artifacts.

The experience with the source models of 'moving rigid 2D objects' and 'moving flexible 2D objects' leads to the expectation that the introduction of the source model of 'moving *flexible* 3D objects' will significantly reduce the required data rate. In the future, the source model will be extended to incorporate a priori knowledge such as face and mimic models for coding of head and shoulder scenes.

8. Acknowledgments

I wish to thank Professor Musmann for encouraging this project and for many discussions about OBASC. Furthermore, I wish to thank Dr. Hötter for fruitful discussions on image analysis. Finally, two of my colleagues took over an important part in software development. Dipl.-Ing. H. Li adapted the algorithms for coding 2D shapes to the requirements of 3D objects. Dipl.-Ing. J. Stauder implemented a background memory for the OBASC coder. Martina Siebke, M.A., improved English grammar and style. This research was supported by the German Federal Government and the Deutsche Bundespost Telekom.

9. References

[1] J.K. Aggarwal and N. Nandhakumar, "On the computation of motion from sequences of images – A review", *Proc. IEEE*, Vol. 76, No. 8, August 1988, pp. 917–935.

[2] K. Aizawa, H. Harashima and T. Saito, "Model-based analysis–synthesis image coding (MBASIC) system for a person's face", *Signal Processing: Image Communication*, Vol. 1, No. 2, October 1989, pp. 139–152.

[3] British Telecom Research Lab. (BTRL), Test sequence *Miss America*, CIF, 10 Hz, 50 frames, Martlesham, Great Britain.

[4] H. Busch, "Subdividing non rigid 3D objects into quasi rigid parts", *IEE 3rd Internat. Conf. Image Processing and Applications*, IEE Publ. 307, Warwick, UK, July 1989, pp. 1–4.

[5] CCITT, "Draft revision of recommendation H.261: Video codec for audio visual services at $p \times 64$ kbit/s", Study Group XV, WP/1/Q4, Specialist group on coding for visual telephony, Doc. No. 584, November 1989.

[6] CCITT, SG XV, Doc. #525, Description of Ref. Model 8 (RM8), 1989.

[7] Centre National d'Etudes des Telecommunication (CNET), Test sequence *Claire*, CIF, 10 Hz, 156 frames, Paris, France.

[8] C.S. Choi, T. Takebe and H. Harashima, "Three-dimensional (3-D) facial model-based description and synthesis of facial expressions", *IEICE Trans. Elect. and Commun. Japan*, Part 3, Vol. 74, No. 7, 1991, pp. 12–23.

[9] C.S. Choi, M. Sato, H. Harashima and T. Takebe, "Parameterization of facial images in knowledge-based coding and waveform coding", *Picture Coding Symp. (PCS '91)*, Tokyo, Japan, No. 9.5, September 1991.

[10] P. Eckman and V.W. Friesen, *Facial Coding System*, Consulting Psychologists Press Inc., Palo Alto, USA, 1977.

[11] R. Forchheimer, O. Fahlander and T. Kronander, "A semantic approach to the transmission of face images", *Picture Coding Symp. (PCS '84)*, Cesson-Sevigne, France, No. 10.5, July 1984.

[12] Forschungsinstitut der Deutschen Bundespost Telekom, Test sequence *Michael*, CIF, 25 Hz, 90 frames, Darmstadt, Germany.

[13] M. Eden and M. Kocher, "On the performance of a contour coding algorithm in the context of image coding, Part 1: Contour segment coding", *Signal Processing*, Vol. 8, No. 4, July 1985, pp. 381–386.

[14] M. Gilge, T. Engelhardt and R. Mehlan, "Coding of arbitrarily shaped image segments based on a generalized

orthogonal transform", *Signal Processing: Image Communication*, Vol. 1, No. 2, October 1989, pp. 153–180.

[15] H. Harashima, K. Aizawa and T. Saito, "Model-based analysis synthesis coding of videotelephone images – Conception and basic study of intelligent image coding", *Trans. IEICE*, Vol. E 72, No. 5, May 1989, pp. 452–459.

[16] H. Harashima and F. Kishino, "Intelligent image coding and communications with realistic sensations – Recent trends", *Trans. IEICE*, Vol. E 74, No. 6, June 1991, pp. 1582–1592.

[17] M. Hötter and R. Thoma, "Image segmentation based on object oriented mapping parameter estimation", *Signal Processing*, Vol. 15, No. 3, October 1988, pp. 315–334.

[18] M. Hötter, "Predictive contour coding for an object-oriented analysis–synthesis coder", *IEEE Internat. Symp. Inform. Theory*, San Diego, California, USA, January 1990, p. 75.

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

[20] M. Hötter, "Optimization of an object-oriented analysis–synthesis coder based on the model of flexible 2D-objects", *Picture Coding Symp. (PCS '91)*, Tokyo, Japan, No. 10.4, September 1991.

[21] M. Hötter, "Optimization and efficiency of an object-oriented analysis–synthesis coder", *IEEE Trans. Circuit Systems Video Technology*, submitted for publication.

[22] M. Hötter, "Objektorientierte Analyse–Synthese-Codierung basierend auf dem Modell bewegter, zweidimensionaler Objekte", Fortschr.-Ber. VDI Reihe 10 Nr. 217, Düsseldorf, VDI-Verlag, 1992.

[23] M. Hötter, Private communication, January 1992.

[24] Commitee Draft of Standard ISO 11172, Coding of moving pictures and associated audio, ISO/MPEG 90/176, December 1990.

[25] F. Kappei and C.-E. Liedtke, "Modelling of a natural 3-D scene consisting of moving objects from a sequence of monocular TV images", *SPIE*, Cannes, Vol. 860, 1987.

[26] F. Kappei and G. Heipel, "3D model based image coding", *Picture Coding Symp. (PCS '88)*, Torino, Italy, September 1988, p. 4.2.

[27] R. Koch, "Adaptation of a 3D facial mask to human faces in videophone sequences using model based image analysis", *Picture Coding Symp. (PCS '91)*, Tokyo, Japan, September 1991, pp. 285–288.

[28] R. Koch, "Dynamic 3-D scene analysis through synthesis feedback control", *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 15, No. 6, June 1993, pp. 556–568.

[29] H. Li, P. Roivainen and R. Forchheimer, "3-D motion estimation in model-based facial image coding", *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 15, No. 6, June 1993, pp. 545–555.

[30] P. Meer, D. Mintz and D.Y. Kim, "Robust regression methods in computer vision: A review", *Internat. J. Computer Vision*, No. 6, 1991, pp. 59–70.

[31] O.J. Morris, M. de J. Lee and A.G. Constantinides, "Graph theory for image analysis: An approach based on the shortest spanning tree", *IEE Proc.*, No. 2, April 1986.

[32] H.G. Musmann, M. Hötter and J. Ostermann, "Object-oriented analysis–synthesis coding of moving images", *Signal Processing: Image Communication*, Vol. 1, No. 2, October 1989, pp. 117–138.

[33] J. Ostermann, "Modelling of 3D-moving objects for an analysis–synthesis coder", in: B. Girod, ed., *SPIE/SPSE Symposium on Sensing and Reconstruction of 3D Objects and Scenes '90, Proc. SPIE* 1260, Santa Clara, California, February 1990, pp. 240–250.

[34] J. Ostermann and H. Li, "Detection and coding of model failures in an analysis–synthesis coder based on moving 3D-objects", *3rd Internat. Workshop on 64 kbit/s Coding of Moving Video*, Rotterdam, The Netherlands, No. 4.4, September 1990.

[35] J. Ostermann, "Coding of color parameters in an analysis–synthesis coder based on moving 3D-objects", *Picture Coding Symp. (PCS '91)*, Tokyo, Japan, No. 10.5, September 1991.

[36] D. Pearson, "Texture mapping in model-based image coding", *Signal Processing: Image Communication*, Vol. 2, No. 4, December 1990, pp. 377–395.

[37] R. Plomben, Y. Hatori, W. Geuen, J. Guichard, M. Guglielmo and H. Brusewitz, "Motion video coding in CCITT SG XV – The video source coding", *Proc. IEEE GLOBECOM*, Vol. II, December 1988, pp. 31.2.1–31.2.8.

[38] D.F. Rogers, *Procedural Elements for Computer Graphics*, McGraw-Hill, Singapore, 1985.

[39] P. Roivainen and R. Forchheimer, "Model-based coding: Estimation of face motion", *Picture Coding Symp. (PCS '91)*, Tokyo, Japan, No. 10.5, September 1991.

[40] P.J. Rousseeuw and A.M. Leroy, *Robust Regression and Outlier Detection*, Wiley, New York, 1987.

[41] T. Sasaki, S. Akamatsu, H. Fukamachi and Y. Suenaga, "A color segmentation method for image registration in automatic face recognition", *Advances in Color Vision*, Optical Society of America, *Proc. of 1992 Technical Digest Series*, Vol. 4, SaB17, January 1992, pp. 179–181.

[42] H. Schiller and M. Hötter, "Investigations on colour coding in an object-oriented analysis–synthesis coder", *Signal Processing: Image Communication*, Vol. 5, No. 4, October 1993, pp. 319–326.

[43] D. Thalman, *Image Synthesis: Theory and Practice*, Springer, Tokyo, 1987.

[44] R. Thoma and M. Bierling, "Motion compensating interpolation considering covered and uncovered background", *Signal Processing: Image Communication*, Vol. 1, No. 2, October 1989, pp. 191–212.

[45] B. Welsh, Model-based coding of images, PhD Thesis, Essex University, Great Britain, January 1991.

[46] X. Zhuang, T. Wang and P. Zhang, "A highly robust estimator through partially likelihood function modelling and its application in computer vision", *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 14, No. 1, January 1992, pp. 19–35.

Instructions to authors

General. Prospective authors are encouraged to submit manuscripts within the scope of the Journal. To quality for publication, papers must be previously unpublished and not be under consideration for publication elsewhere. All material should be sent in quadruplicate (original plus three copies) to the Editor-in-Chief. Contributors are reminded that once their contribution has been accepted for publication, all further correspondence should not be sent to the Editor, but directly to the publishers (Editorial Department, Elsevier Science B.V., P.O. Box 1991, 1000 BZ Amsterdam, The Netherlands).
All manuscripts will be assessed by at least two (anonymous) referees.
Upon acceptance of an article, the author(s) will be asked to transfer copyright of the article to the publisher. This transfer will ensure the widest possible dissemination of information.
Accepted languages are English (preferred), French and German. Each contribution should be accompanied by abstracts (see *Abstracts*).
Page proofs will be sent to the principal author with an offprint order form. Fifty offprints of each article can be ordered free of charge. Costs arising from alterations in proof, other than of printer's errors, will be charged to the authors.
Short papers not exceeding five typewritten pages will be reviewed immediately. As their proofreading will be done by the publishers and no page proofs will be sent to the authors, their presentation should be very clear.
Typed material (including abstracts, list of references) should be double spaced on one side of A4 size (297 210 mm) paper with a wide margin. All pages should be numbered. The first page should include the article title and the author's name and affiliation, as well as a name and mailing address to be used for correspondence and transmission of proofs. The second page should include a list of unusual symbols used in the article and the number of pages, tables and figures. It should also contain a proposed running headline of less than thirty-five characters and the keywords in English.

Abstracts. The text of the paper should be preceded by abstracts of no more than 200 words in English. Abstracts should contain the substance of the methods and results of the paper. The abstracts of short papers should be less than 50 words.

Figures. All illustrations are to be considered as figures, and each should be numbered in sequence with Arabic numerals. The drawings of the figures must be originals, drawn in black india ink and carefully lettered, or printed on a high-quality laser printer. Each figure should have a caption and these should be listed on a separate sheet. Care should be taken that lettering on the original is large enough to be legible after reduction. Each figure should be identified. The approximate place of a figure in the text should be indicated in the margin.

Reproduction in colour. In case the author wishes one or more figures to be printed in colour, the *extra* costs arising from such printing will be charged to the author. In this case 200 offprints may be ordered free of charge. More details are available from the Publisher.

Equations. Equations to which reference is made in the text should be numbered; equations should be referred to by citing the equation number enclosed in parentheses. Special care should be taken of those symbols which might easily cause confusion.

Tables. Tables should be typed on separate sheets. Each table should have a number and a title. The approximate places for their insertion in the text should be indicated in the margin.

References. References must be *in alphabetical order* in the style shown below:

Book	[1]	A.V. Oppenheim et al., *Digital Signal Processing*, Prentice Hall, Englewood Cliffs, NJ, 1975, Chapter 10, pp. 491–499.
Journal	[2]	F.J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform", *Proc. IEEE*, Vol. 66, No. 1, January 1978, pp. 53–83.
Conference Proceedings	[3]	D. Coulon and D. Kayser, "A supervised-learning technique to identify short natural language sentence", *Proc. 3rd Internat. Joint Conf. on Pattern Recognition*, Coronado, CA, 8–11 November 1976, pp. 85–89.
Contributed Volume	[4]	E.F. Moore, "The firing squad synchronization problem", in: E.F. Moore, ed., *Sequential Machines, Selected Papers*, Addison-Wesley, Reading, MA, 1964, pp. 213–214.

Footnotes in text. Footnotes in the text should be identified by superscript numbers and listed consecutively on a separate page.

SUBMISSION OF ELECTRONIC TEXT

In order to publish the paper as quickly as possible after acceptance, authors are encouraged to submit the final text also on a 3.5" or 5.25" diskette. Both double density (DD) and high density (HD) diskettes are acceptable. The diskette may be formatted with either MS-DOS/PC-DOS or with Macintosh OR. See the Notes for Electronic Text Preparation at the back of this issue for further information. The final manuscript may contain parts (e.g. formulae, complex tables) or last minute corrections which are not included in the electronic text on the diskette; however, this should be clearly marked in an additional hardcopy of the manuscript. Authors are encouraged to ensure that apart from any such small last-minute corrections, **the disk version and the hardcopy must be identical.** Discrepancies can lead to proofs of the wrong version being made.

Detailed guidelines for authors are available either from the Publisher or from the Editor-in-Chief.

EUROPEAN ASSOCIATION FOR SIGNAL PROCESSING

Administrative Committee
President: M. Bellanger, Conservatoire des Arts et Métiers, 292, Rue St. Martin, F-75141 Paris Cedex 03, France
Secretary-treasurer: U. Heute, LNS/Techn. Fakultät, Universität Kiel, Kaiserstraße 2, D-24143 Kiel, Germany
Conference and Workshop Coordinator: J. Biemond, Department of Electrical Engineering, Delft University of Technology, Mekelweg 4, 2626 CJ Delft, The Netherlands
Member-at-large: W. Mecklenbräuker, Institüt für Nachrichtentechnik, Technische Universität Wien, Gußhausstraße 29/389, 1040 Wien, Austria
Member-at-large: J. Vandewalle, Katholieke Universiteit Leuven, Kardinaal Mercierlaan 94, B-3030 Leuven (Heverlee), Belgium

ELSEVIER SCIENCE

prefers the submission of electronic manuscripts

Electronic manuscripts have the advantage that there is no need for the rekeying of text, thereby avoiding the possibility of introducing errors and resulting in reliable and fast delivery of proofs.

The preferred storage medium is a 5.25 or 3.5 inch disk in MS-DOS format, although other systems are welcome, e.g. Macintosh.

After **final acceptance,** your disk plus one final, printed and exactly matching version (as a printout) should be submitted together to the accepting editor. **It is important that the file on disk and the printout are identical.** Both will then be forwarded by the editor to Elsevier.

Please follow the general instructions on style/arrangement and, in particular, the reference style of this journal as given in "Instructions to Authors."

Please label the disk with your name, the software & hardware used and the name of the file to be processed.

SIGNAL PROCESSING: IMAGE COMMUNICATION

- ☐ Please send me a free sample copy
- ☐ Please send me subscription information
- ☐ Please send me Instructions to Authors

Name ____________________

Address ____________________

____

____

ELSEVIER SCIENCE B.V.

Send this coupon or a photocopy to:

ELSEVIER SCIENCE B.V
Attn: Engineering and Technology Department
P.O. Box 1991, 100 BZ Amsterdam, The Netherlands

c:\sheila\repcovs.94