Modelling of 3D moving objects for an analysis−synthesis coder

Jörn Ostermann

Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung
Appelstr. 9a
D−3000 Hannover 1

ABSTRACT

For object−oriented analysis−synthesis coding an image analysis algorithm is required which automatically generates the parameter sets which describe moving 3D objects in an image sequence.

Areas changed between two consecutive images are detected by means of change detection. Special processing is carried out to compute areas which coincide with object boundaries and to eliminate areas which represent illumination changes. These areas are segmented into silhouettes of moving objects and uncovered background. The border of an object silhouette is interpreted as the outermost contour of an object. These contours in combination with a simple function giving the z−distance between them provide a first estimiate for 3D shape of a model object. In order to improve the efficiency of motion analysis a concept for combining model objects with new parts of moving objects is proposed.

Results of an automatic image analysis based on moving 3D objects are shown using video telephone test sequences.

1. INTRODUCTION

For encoding moving video signals at data rates as low as 64 kbit/s, conventional hybrid coders are normally used. A hybrid coder subdivides each image of an image sequence into blocks of NxN picture elements (pels) [1]. The luminance and chrominance signals of each block are encoded by motion compensated predictive and transform coding algorithms. Recently a coding method called object−oriented analysis−synthesis coding has been proposed [2]. Object−oriented analysis−synthesis coding subdivides each image into moving objects and encodes each object by three sets of parameters defining motion, shape, and color information of the object.

This paper presents the image analysis part of an analysis−synthesis coder based on the model of moving 3D objects. The input to the image analyser is an image sequence from a monocular TV camera. The image analysis computes a 3D **model world** [2,7]. A **model image** is synthesized by projecting the model world onto the target of an artificial **model camera** by means of perspective projection. The goal of image analysis is to generate a model world with a model image identical to the **real image**. Therefore the moving 3D **real objects** are modelled by 3D **model objects** in the model world. The surface of a model object is represented by a mesh of triangles. In order to define the surface colors

Reprint from Jörn Ostermann, "Modelling of 3D−moving objects for an analysis−synthesis coder", Sensing and Reconstruction of Three−Dimensional Objects,Bernd Girod, Editor, Proc. SPIE 1260, pp 240 − 249 (1990)

of the model objects an image of the TV sequence is projected onto the model world. By this, the surface colors of each visible triangle of the objects are defined.

Analysis of images based on the model of moving 3D objects consists of four major problems: detection of an initial object silhouette, modelling 3D shape from the silhouette, motion analysis, and adaptation of 3D shape. Detection of moving objects can be improved by motion analysis [3]. Hence the results of motion analysis are used to adapt object shape.

For getting started a change detection is carried out on the first two images of the sequence. Areas where objects move and background which is uncovered due to object motion are declared as changed. However, in a first step changed areas are interpreted as silhouettes of moving objects. In the next processing step a 3D model object having a silhouette as calculated by change detection is generated and placed in the model world. Motion estimation is carried out for this model object. The motion parameters are used to improve the silhouette.
After this initialization of image analysis, for each new image a change detection is carried out between the current model image and the new real image. After motion compensating the model objects, the silhouettes of the model objects are adapted such that they are consistent with the change detection mask.

In Section 2 change detection as needed for an automatic modelling of 3D moving objects will be described. Section 3 presents an algorithm for modelling 3D shape from an object silhouette. Motion estimation is explained in Section 4. Section 5 outlines the algorithms for updating model object shape. Section 6 gives experimental results.

2. CHANGE DETECTION

The aim of change detection is to distinguish between temporally changed and unchanged regions of two successive images. In the processing steps following change detection each changed region represents one object together with background uncovered due to motion of the object. The performance of this change detector determines the efficiency of the analysis system and therefore the efficiency of the object−oriented analysis−synthesis coder. Hence, the change detection mask has to be precise at object boundaries and only moving objects should be declared as changed. Problems arise when images are noisy or when shadows move due to object motion. Therefore the output of a basic change detection algorithm (Section 2.1) is postprocessed such that the resulting masks meet the requirements for image analysis based on moving 3D objects. In a first step the changed regions get aligned with contours in the original images (Section 2.2), in a second step areas changed due to shadows and illumination changes are detected (Section 2.3).

2.1 The basic change detection algorithm

Fig. 1 shows a block diagram of the change detector [3]. Change detection evaluates the absolute frame difference between image k and image k+1. The differences are summed up over a measurement window of size 3x3 pels. If this sum is greater than a certain change threshold T_{ch}, the pel considered is declared as changed, otherwise as static. A median filter of size 5x5 pels smooths the region boundaries and eliminates single pels. Finally small areas are eliminated in order to get homogeneous regions. Output is the initial binary change detection mask B_c. For adapting the algorithm to noise, the threshold T_{ch} is updated according to the noise of the static image parts. The updated threshold will be applied to the next pair of images.

2.2 Correction of change detection masks by means of local contour information

It has been found that change detection masks generated by the presented algorithm are generally some pels too large due to non−linear filter operations. Therefore in a first postprocessing step, boundaries of changed areas are corrected such that they coincide with occluding contours in the original image [4].

In the first part of this correction a stripe on the border of the mask is defined as the area of interest in which contours in the original image are searched. This area has to be defined such that at object boundaries with low contrast the change detection mask is not shrunk to coincide with strong edges inside moving objects or blown up to match edges in the static background. Since the change detection masks are too large, the stripe lies completely inside the mask (Fig. 2). In this area of interest, contours

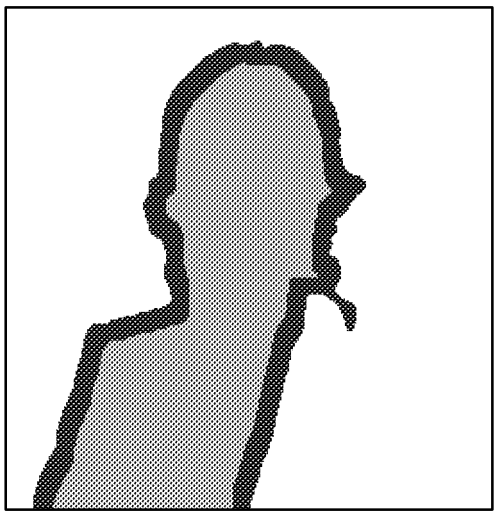

Fig. 2: Area of interest for change detection mask correction

are searched by evaluating local gradients. Usually it is not possible to find a continuous contour line. Therefore borders of changed areas are shifted to contours in those instances where relevant contour sections have been

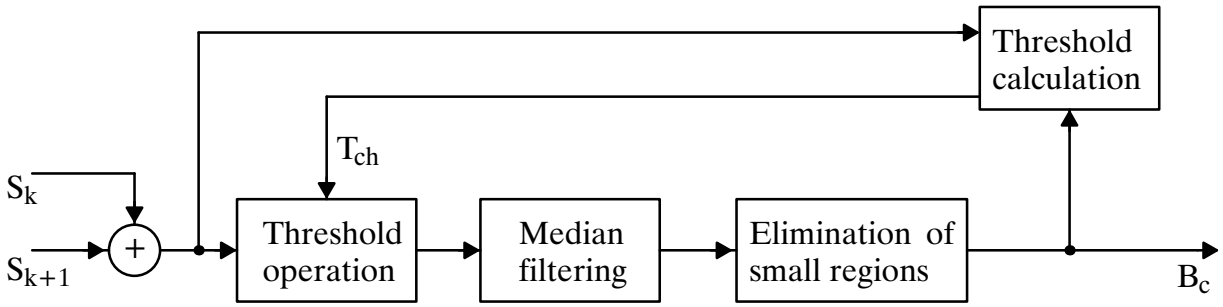

Fig. 1: Block diagram of the change detector

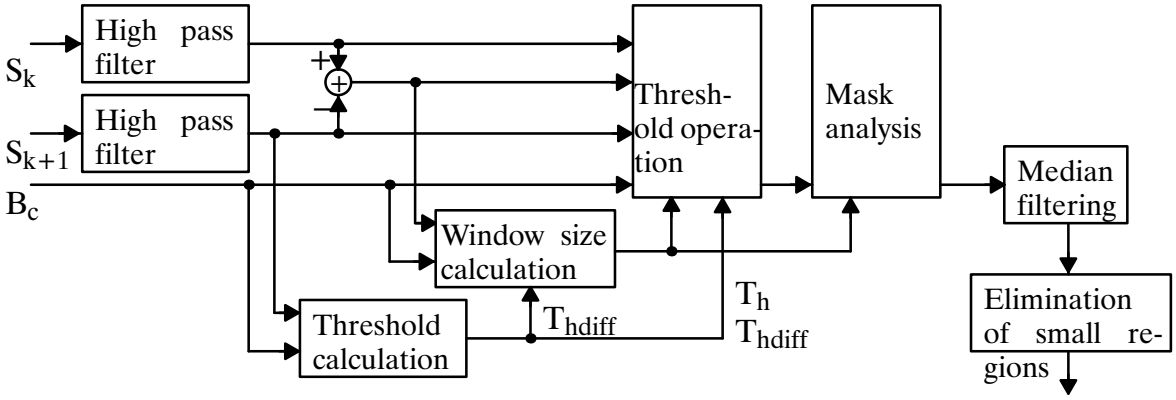

Fig. 3: Correction of change detection masks due to illumination changes

found. After this correction of the change detection mask, the boundaries of mask B'_c are very close to the contours of the detected objects.

2.3 Correction of change detection mask errors due to illumination changes and shadows

Natural scenes are usually illuminated by diffuse and direct light sources which have slight temporal variations of intensity. The effect of the illumination is that objects have shadows and moving objects have moving shadows. In this context shadows are interpreted as illumination changes and the basic change detector declares such areas as changed. However, these areas must not be classified as moving objects, because motion compensation only works on opaque objects and illumination changes are not opaque objects as long as background texture is visible. This is where the following algorithm comes in. Image regions classified as changed are reclassified by comparing the high frequencies of the two original images, assuming that illumination changes mainly affect low frequencies of a region.

Input to the algorithm are the change detection mask B_c and the two images on which change detection was carried out (Fig. 3). The original images are highpass filtered and a local adaptive scaling is carried out for normalization. From these two highpass images an absolute frame difference image is calculated. Two parameters are determined for an automatic adaptation to different input sequences. A new noise threshold T_h is computed for discriminating changed and unchanged parts of the highpass filtered images. T_h is proportional to the noise variance of the highpass filtered image. From this noise threshold T_h a noise threshold T_{hdiff} can be calculated from the noise variance in the highpassed filtered images for distinguishing noise from signal in the difference image [3]. The size of a measurement window is determined such that the mean absolute frame difference of the highpassed images inside the measurement window is greater than noise threshold T_{hdiff} in approximately 90% off all window positions where the pel in the window center is declared as changed.

In a threshold operation each pel declared as changed in the change detection mask is classified again if in the measurement window of both highpass filtered images the signal exceeds T_h. This means that a reclassification of a changed pel is only taken into consideration if structure is found in the measurement windows. If the threshold T_{hdiff} is not exceeded in the window of the difference image the pel under consideration is declared as unchanged because the difference signal mainly contains noise. Otherwise, if the threshold T_{hdiff} is exceeded, the measurement window contains background and parts of moving objects and the

pel is marked for further processing in the block mask analysis.

Mask analysis classifies pels marked for processing by the threshold operation. Marked pels are located near object boundaries and the distribution of changed pels in the measurement window determines whether the point belongs to object or background. The distribution of changed pels inside the measurement window of each marked pel is analyzed. Therefore the centre of gravity of changed pels of the current window is computed. From the distribution of changed pels and its center of gravity it can be decided whether the pel belongs to an object or to an area of illumination changes.

The resulting modified change detection mask is median filtered and small regions are eliminated.

3. MODELLING 3D SHAPE FROM AN OBJECT SILHOUETTE

When the silhouette of an object is known the problem arises as to how to model a 3D object from its silhouette. An algorithm is presented here which generates a natural looking shape with a silhouette as required. Fig. 4 illustrates

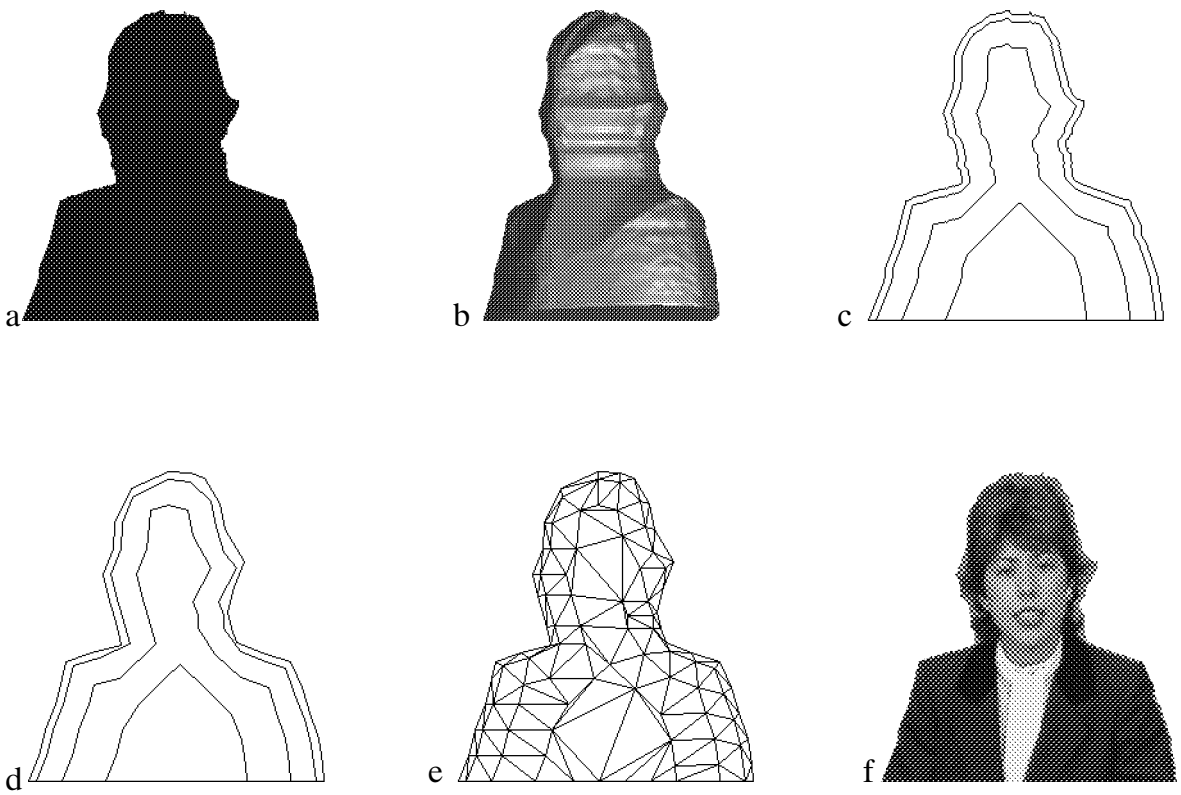

Fig. 4. Processing steps from object silhouette to model object. (a) Object silhouette; (b) 3D object shape with required silhouette; (c) Contour lines approximating object shape; (d) Polygons approximating contour lines; (e) Mesh of triangles using polygon points as vertices; (f) Model object: mesh of triangles with luminances projected onto it.

five steps from silhouette (Fig. 4a) towards a model object as required for this modelling system.

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 dimension (Fig. 4b). The parameters giving the z dimension are the object width and the ratio of object height to object width (Fig. 5). 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 (Fig. 4c, 5). This enables the algorithm to compute a homogeneous mesh of triangles. In a subsequent step contour lines are approximated by polygons (Fig. 4d). Special care is taken to generate polygons which preserve points of maximum curvature of contour lines. The resulting polygon points are the vertices of a mesh of triangles (Fig. 4e). By projecting luminances onto this mesh a model object is generated.

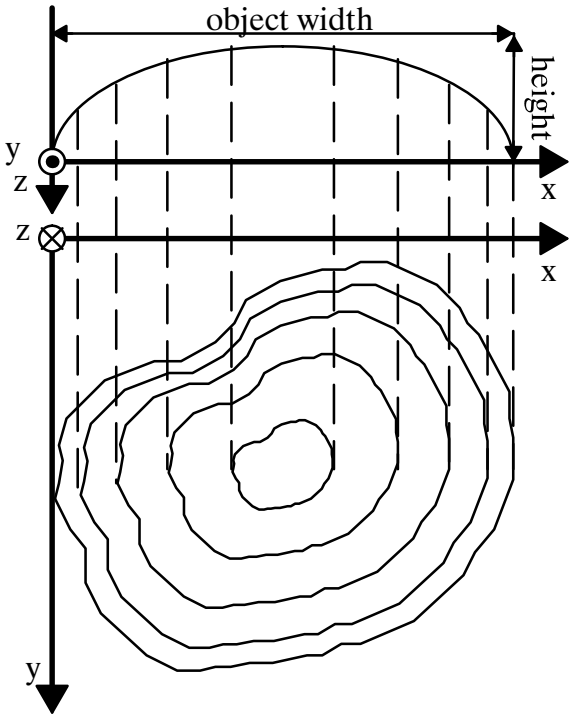

Fig. 5: Function giving z−Dimension

4. ESTIMATING MOTION

3D motion estimation is carried out. For motion analysis of an image the algorithm [5] computes three translation and three rotation parameters for each model object. If necessary a model object is subdivided into flexible connected components [6]. For each component a separate set of motion parameters is computed.

Motion estimation is carried out with the help of observation points. A point is defined as an observation point if it has a large local gradient. To each observation point initial 3D coordinates, a luminance value in a reference image and local gradients are assigned. A linearized set of equations can be set up for the observation points. Motion estimation is carried out by solving iteratively this set of equations using linear regression. As indicated in [3] it is important to check the relevance of an observation point during iteration.

5. UPDATING MODEL OBJECT SHAPE

Model object shape has to be updated in two cases. First, the initial model object is generated from the first change detection mask. Results of motion analysis must be used to improve object silhouette because a change detection mask contains both the moving object and the background which is uncovered by object motion. Secondly, there are objects which do not move in their entirety, e.g. the head of person might move and the body might not move. The change detection will detect the moving head at once but not the body. As soon

as the body moves as well, there is the problem as to how to modify the existing model object such that new object parts are incoorporated into the model object.

First the problem of uncovered background is tackled. It arises initially when changed areas of the first change detection mask are interpreted as objects although they contain uncovered background. After initialization the problem always arises that parts of an object which have not moved before start moving and are discovered to have done so by means of change detection. After motion estimation six motion parameters are known for each object. They are the key for distinguishing objects and background. From motion parameters a displacement vector can be calculated for each point of the changed area. If the end point of a vector lies inside the changed area, the corresponding starting point belongs to the object before motion, otherwise it belongs to uncovered background [8]. Applying this test, an object mask is computed. The vertices of the model object get shifted such that the object silhouette matches the object mask.

Now we consider the problem of modifying a model object shape such that new object parts detected by change detection are incoorporated into the shape. There are several solutions to this problem.

Assuming that knowledge about object shape acquired from previous images is correct, the net of a model object gets enlarged **after** motion compensation such that the new object silhouette includes new object parts indicated by the change detection mask. The advantage of this procedure is that knowledge does not get lost. However, if change detection masks include more than moving objects and background uncovered by object motion, background will be incorporated into the object and there is no way of correcting this error. Therefore this approach of adapting model object shape after motion compensation is more useful for image analysis systems where depth sensors discriminate foreground and background. The image analysis scheme proposed here takes care that the movement of a model object silhouette is consistent with its change detection mask. It combines model object and change detection mask **before** motion compensation. Motion parameters are calculated for each object. Then they are applied to the objects and the change detection mask. Uncovered background is eliminated as described above. By this means, shape errors introduced previously due to erroneous change detection masks can be corrected, if the object moves such that the erroneous parts of the model object shape after motion compensation are outside the initial change detection mask.

6. EXPERIMENTAL RESULTS

Fig. 6: Frame 10 from test sequence "MICHAEL"

Fig. 7: Frame 10 from test sequence "CLAIRE"

The proposed algorithms were tested on videophone test sequences. Figs. 6 and 7 show images of the test sequences "Michael" having structured background and "Claire" with a uniform background. Figs. 8 and 9 show model objects as they are modelled automatically by the presented image analysis after analyzing 10 images using basic change detection, modelling object shape from object silhouettes, motion estimation and updating shape. As can be seen from Fig. 8, areas in the background are declared as model objects. They represent moving shadows due to moving real objects and direct light sources. Fig. 9 shows that background is added to the head of the model object. Figs. 10 and 11 show model objects as they are modelled automatically by image analysis again after analyzing 10 images using in addition to basic change detection the algorithm for correcting the change detection mask by means of local contour information. As can be seen this algorithm improves image segmentation only if image background is uniform. Figs. 12 and 13 show model objects as they are modelled using basic change detection and additionally the algorithm for eliminating areas changed due to illumination changes. These figures show that illumination effects are only removed when image background is structured. In the case of uniform image background analysis results do not change. Comparing Figs. 8 to 13 it can be seen that best results are achievable using basic change detection in combination with elimination of illumination changes and local contour information. Then the results look like Fig. 12 for the test sequence "MICHAEL" and like Fig. 11 for the test sequence "CLAIRE". It was found that an automatic adaptation of image analysis to different image sequences is possible.

7. CONCLUSION

An automatic image analysis system for an object−oriented analysis−synthesis coder has been presented. From a natural TV image sequence, image analysis is able to model moving 3D objects. Objects are detected automatically by change detection. Motion analysis and shape updates are used to improve object shape. Special care is taken that pseudo−objects are not caused by illumination changes and that model object boundaries match real object bound-

Fig. 8: Model world ”MICHAEL” using basic change detection

Fig. 9: Model world ”CLAIRE” using basic change detection

Fig. 10: Model world ”MICHAEL” using basic change detection and local contour information

Fig. 11: Model world ”CLAIRE” using basic change detection and local contour information

Fig. 12: Model world ”MICHAEL” using basic change detection and elimination of illumination changes

Fig. 13: Model world ”CLAIRE” using basic change detection and elimination of illumination changes

aries. For an analysis of scenes of low complexity there seems to be no need to use additional stereo or depth sensors in addition to a single TV camera. Integration of this implementation into a complete object−oriented analysis−synthesis coder still has to be investigated.

8. REFERENCES

[1] R. Plomben, Y. Hatori, W. Geuen, J. Guichard, M. Guglielmo, H. Brusewitz, "Motion Video Coding in CCITT SG XV − The Video Source Coding", Proc. IEEE GLOBECOM, vol. II, pp. 31.2.1 − 31.2.8, Dec. 1988.

[2] H.G. Musmann, M. Hötter, J. Ostermann, "Object−oriented analysis−synthesis coding of moving images", SIGNAL PROCESSING, Image Communication, no. 2, Nov. 1989.

[3] M. Hötter, R. Thoma, "Image segmentation based on object oriented mapping parameter estimation", SIGNAL PROCESSING, vol. 15, no. 3, Oct. 1988.

[4] T. Kassbaum, "Korrektur von Änderungsmasken durch Auswertung lokaler Konturinformation", thesis, University of Hannover, 1989.

[5] F. Kappei, "Modellierung und Rekonstruktion bewegter dreidimensionaler Objekte in einer Fernsehbildfolge", PhD thesis, University of Hannover, 1988.

[6] H. Busch, "Subdividing non rigid 3D objects into quasi rigid parts", Proceedings of the 3rd IEE International Conference on Image Processing and its Applications, University of Warwick, 1989.

[7] C.−E. Liedtke, H. Busch, R. Koch, "Automatic modelling of 3D moving objects from a TV image sequence", to be published at SPIE/SPSE Symposium on Electronic Imaging, California, USA, 1990.

[8] M. Bierling, R. Thoma, "Motion compensating field interpolation using a hierarchically structured displacement estimator", Signal Processing, vol. 11, no. 4, 1986.