AUTOMATIC MODELLING OF 3D MOVING OBJECTS FROM A TV IMAGE SEQUENCE

C.−E. Liedtke, H. Busch, R. Koch
Institut für Theoretische Nachrichtentechnik
und Informationsverarbeitung
Universität Hannover
Appelstr. 9A, D−3000 Hannover 1, FRG
Tel.49 511 762 5318, Fax. 49 511 762 5333

ABSTRACT

Luminance changes in a TV image sequence can be interpreted as being due to a scene consisting of 3D−objects moving with 6 degrees of freedom in 3D−space. The 3D−space is illuminated by a light source and is observed by a camera. A parametric model is presented which employs an explicit representation of the illumination source, the camera, and the 3D−objects. A method is suggested, in which an Analysis−by−Synthesis−process automatically extracts the model parameters from an incoming sequence of monocular TV images and thereby allows the modelling of the 3D scene. Since a parametric description of a complex scene resulting in a physically satisfactory modelling cannot be achieved in all generality, we have concentrated on the modelling of quasi−rigid, natural objects with homogeneous surfaces like the head−shoulder parts of human beings.

1. INTRODUCTION

Various approaches for the assessment of a 3D object description from image sequences have been published. A detailed overview over the field has recently been compiled by Aggarval et.al. [1]. We use a model−based image analysis method, which is based on the Analysis−by−Synthesis principle. During the analysis a model world is build up, which explains the content of the scene and can be used to predict future images of an incoming image sequence. Some fundamental ideas about this approach have been published earlier by Kappei and Liedtke [2, 3].

The method has primarily been developed for future image transmission systems, which work at very low bit−rates, where the 3D model world is obtained in the transmitter and used in the receiver to synthesize 2D image sequences. These transmission systems will be used for picture phones or video−conferencing. Other applications comprise rate conversion, extraction of stereo sequences from monocular TV sequences, and general applications in Computer Graphics.

2. PARAMETRIC MODEL

The 3D−model world represents a possible interpretation of the real world. It consists of three parameterized parts, which are the **illumination source**, described by the illumintion intensity, **the camera**, described by its orientation and length of focus, and the **3D model objects** in the scene, described by their motion, shape, and surface color.

Of special interest are the model objects and their parameters. Each model object consists of several flexibly interconnected components, which are described by their position and orientation in the 3D−space. The components themselves are quasi rigid, i.e. only small flexible distortions are allowed. The motion parameters of an object model consist of a 3D−translation vector and a 3D−rotation vector. As shown in Fig. 1, the object shape, consisting of the connected component shapes, is constructed as a mesh of planar triangles. The shape parameters constitute the relative positions of the vertices of the triangles. The surface color is described by 2D luminance values, distributed over the faces of the triangles.

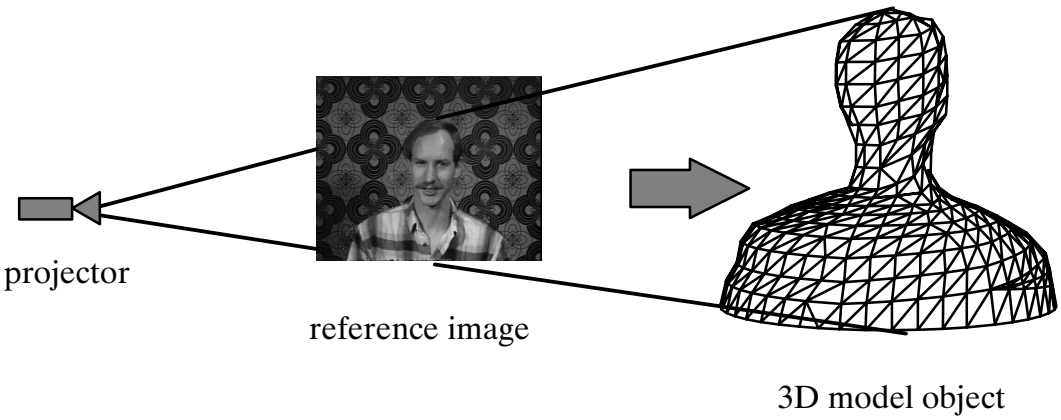

Fig. 1. Model object consisting of a 3D surface shaped by triangular planar patches and a surface color obtained by projection of a reference image onto the surface.

For an automated analysis of the parameters of the model world, certain restrictions have to be made concerning the 3D modelling ability. These restriction imply a) diffuse illumination of the model, b) opaque model objects with diffuse reflecting surfaces, and c) quasi rigid model objects.

A model−based approach for image analysis requires the existence of an initial model. The method is illustrated by Fig. 2. The initial 3D model object is derived from the image sequence by employing the object silhouette. The object silhouette is obtained by image segmentation techniques from the input images. Two methods have been tested and turned out to be successful. One approach, which uses stereoscopic information to separate objects from each other and from a background, has been presented by Waldowski [5]. Another approach, which identifies objects by change of photometric information in an image sequence has been described by Musmann et.al. [6].

After the silhouette has been obtained, one way to arrive at a 3D−body is by rotating the 2D silhouette. Finally by automatic triangulation techniques a mesh of triangles is spread over the 3D body. The coarseness of initial shape modelling can be predetermined by the number of triangles used in the mesh. An example for an initial mesh of triangles is shown in Fig. 2c. Another approach at arriving at an initial estimate of the object shape on the basis of contour lines has been described by Ostermann [7].

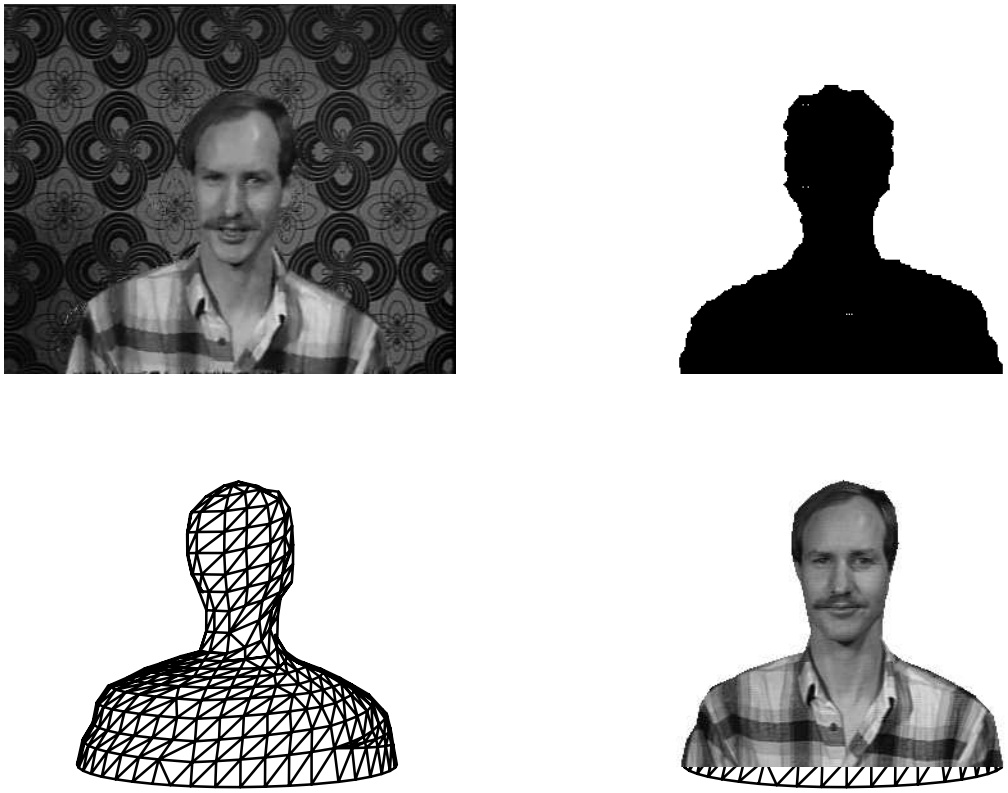

Fig. 2. a|b c|d Generation of the initial 3D model object: (a) Initial input frames, (b) segmentation mask, (c) 3D model shape obtained by rotation of mask, (d) projection of color onto object surface.

The color information within each triangle is obtained by projection of a reference image from the real image plane onto the triangular mesh as is shown in Fig. 1. and results in the initial 3D−model object of Fig. 2d. The color information is stored as a list of triangular segments.

3. ADAPTATION PROCESS

A model image can be obtained from the model object by perspective projection, as is indicated in Fig. 3. The projection process is efficiently realized by perspective projection of just the vertices of the triangles onto the model image plane and by assembly of the affine transformed triangular segments containing the color information within each triangle.

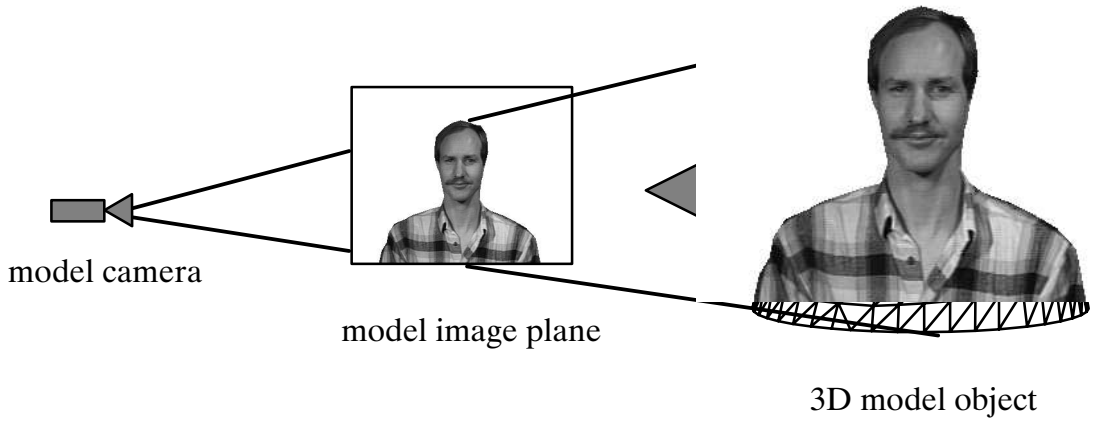

Fig. 3. Model image obtained by perspective projection from a model object.

The parameters of the parametric model world are determined fully automatically by an Analysis–by–Synthesis technique, which is illustrated in Fig. 4.

The incoming sequence of 2D images is compared image by image with the model images, which have been obtained by perspective projection from the model object as has been described above. The comparison is made on the basis of gray level differences. The model parameters are adapted iteratively until differences between the input image under consideration and the model image vanish. Two types of adaptation are used, a parametric adaptation for shape and motion and a non–parametric adaptation for object color. In the following the process for the parametric adaptation is described in more detail.

During parametric adaptation the variation of the mesh of surface triangles is assessed and explained as variation of object shape and object motion. The adaptation is achieved by using the vertices of the triangles as control points and adjusting these on a local basis for shape adaptation and on a global basis for the compensation of object movement.

For practical reasons and for reasons of increased efficiency, the parametric adaptation is based on the use of observation points. An observation point is a location on the object surface containing

color intensity and spatial image gradient value of the object surface. Observation points are selected surface points with a gradient exceeding a noise threshold.

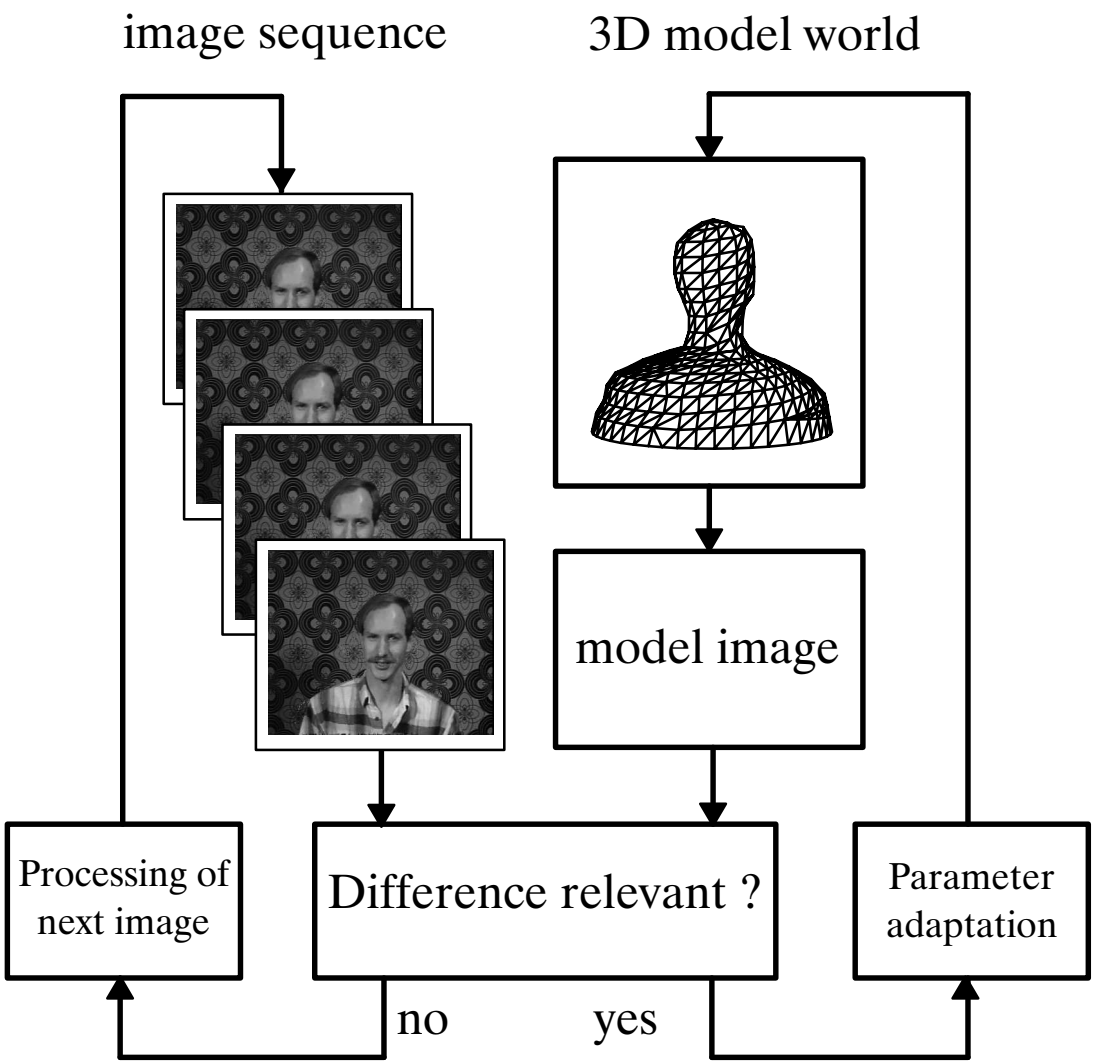

Fig. 4. Analysis−by−Synthesis strategy for adaptation of model parameters.

During parametric adaptation the variation of the mesh of surface triangles is assessed and explained as variation of object shape and object motion. The adaptation is achieved by using the vertices of the triangles as control points and adjusting these on a local basis for shape adaptation and on a global basis for the compensation of object movement.

For practical reasons and for reasons of increased efficiency, the parametric adaptation is based on the use of observation points. An observation point is a location on the object surface containing color intensity and spatial image gradient value of the object surface. Observation points are selected surface points with a gradient exceeding a noise threshold.

The adaptation of shape and motion parameters is based on the assumption of a linear signal model, which assumes that there exists a linear relation between the spatial displacements of the observation points in 3D space and the temporal and spatial gradient in the images. The spatial gradient is obtained from a Sobel operator. The temporal gradient is obtained from the difference of a reference image and the following image.

During parametric adaptation within a limited area of the object surface a linear equation is set up for each observation point. This equation relates the intensity differences between model image and input image under consideration with the unknown parameters, employing the linear signal model mentioned above. The over−determined set of linear equations is solved using a least squares fit by standard techniques. The linear equations may contain 1, 2, 3 or 6 parameters dependent on the particular adaptation goal. A 1− and 2−parametric approach is used for shape adaptation, a 3−parametric approach during object subdivision into components, and a 6−parametric approach for the estimation of the 6 motion parameters in 3D space.

As an example, the estimation of the motion parameters in 3D−space is demonstrated below. for the luminance I of a point on the model object surface and the gray level I(B'') of the corresponding pixel in the observed input image, the following approximate linearized equation holds:

$$
\begin{aligned}
& I(B'') - I = \\
& F * G_x/P_z'' * T_x' \\
+\ & F * G_y/P_z'' * T_y' \\
-\ & F * (P_x'' * G_x + P_y'' * G_y)/P_z''^2 * T_z' \\
-\ & F * (P_x'' * G_x * (P_y'' - C_y) + P_y'' * G_y * (P_y'' - C_y) + P_z'' * G_y * (P_z'' - C_z))/P_z''^2 * R_x' \\
-\ & F * (P_y'' * G_y * (P_x'' - C_x) + P_x'' * G_x * (P_x'' - C_x) + P_z'' * G_x * (P_z'' - C_z))/P_z''^2 * R_y' \\
-\ & F * (G_x * (P_y'' - C_y) - G_y * (P_x'' - C_x)) / P_z'' * R_z'
\end{aligned}
$$

where	$P'' = (P_x'', P_y'', P_z'')^T$	Surface point
	F	Focal length
	$G = (G_x, G_y)^T$	Gradient of the observation point
	$C = (C_x, C_y, C_z)^T$	Center of gravity of the object
	$T' = (T_x', T_y', T_z')^T$	Translation vector
	$R' = (R_x', R_y', R_z')^T$	Rotation vector

Since the relation between the luminance differences and the 6 motion parameters is nonlinear, in the special case of motion parameter estimation the system of equations has to be solved iteratively. The method has been described in detail by Kappei and Liedtke [2, 3].

4. ADAPTATION STRATEGY

The following adaptation strategy for single objects has been developed and is iteratively applied: In the beginning it is assumed that an object consists of one rigid moving 3D object, which can be described by 6 motion parameters (3 translatory and 3 rotatory parameters). Differences between the model image and the image sequence are compensated by adaptation of the 6 motion parameters.

Remaining differences between the model image and the image sequence are explained by the assumption that the object is not really rigid, but rather consists of flexibly interconnected rigid components. For this purpose the object is subdivided into components. The components are identified by observing the individual movements of each triangle of the shape mesh. All simply connected triangles exhibiting similar motion parameters form a component. Since the areas of the individual triangles are frequently too small to reliably estimate the 6 motion parameters for a spatial movement, a two stage approach has been applied. In the first step clusters of triangles with similar motion are obtained on the basis of 2D motion estimates. In a second step full 3D motion compensation for the clusters is used to verify the homogeneity of the model component. An example for the automated subdivision of an object in components is shown in Fig. 5. The method has been described in detail by Busch[4].

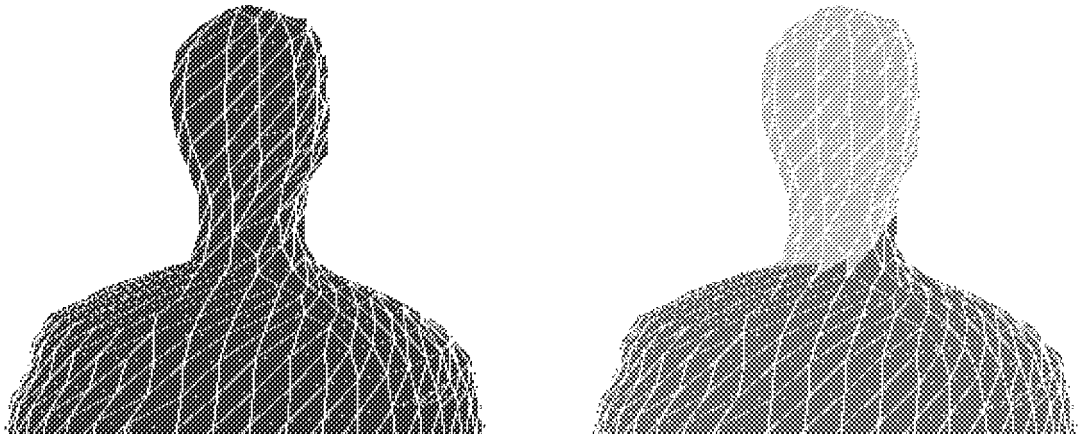

Fig. 5. a|b Subdivision of a model objects into components: (a) model object before subdivision, (b) model object after subdivision into 2 components.

Having indentified the object components, the motion equations for all observation points of each component are solved together in order to obtain the component position and orientation relative to the complete object and thereby adapt the individual component motion.

Any remaining differences between the model image and the image sequence are accounted for by a slowly varying flexible shape of the components of the object. For each observation point the linea-

rized equation for local shape adaptation is formulated. The shape equations of all observation points grouped around each control point are solved to gain local control point displacement as is indicated in Fig. 6.

The displacement of the control points is estimated in two steps. In the first step, during a 1−parameter optimization, the control points are moved along the line of perspective of the reference image. The line of perspective is the line which connects the point on the surface of the 3D model object under investigation with the point of view of the model camera. In a subsequent 2−parameter optimization the control points are moved on a 2D plane perpendicular to the line of perspective. While the first step adapts the gross shape, the second step compensates for movements along the surface, such as skin movements along the cheeks during talking, or the shift of the shirt along the body surface during movements of the body.

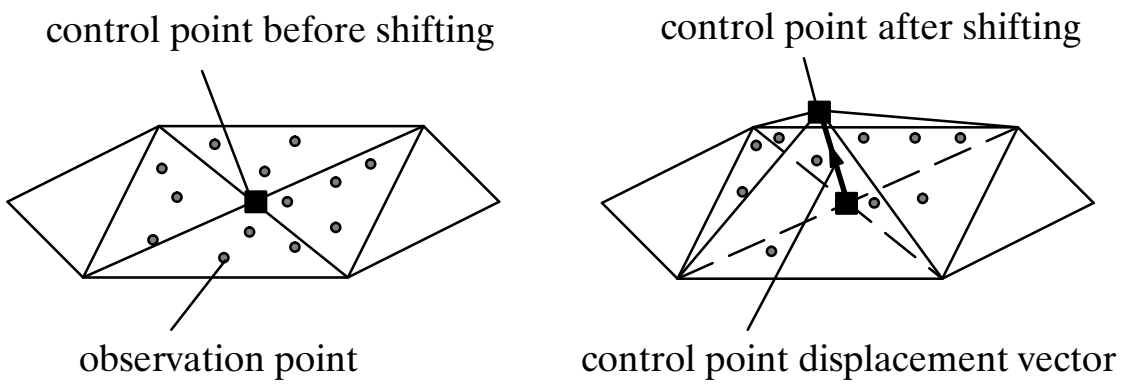

Fig. 6. Local shape adaptation through control point shift.

The position of each control point is corrected using the calculated displacement in order to adapt the component shape.

Residual errors between real and model objects are accounted for as changes in the object color. First a mask is obtained from areas which show large differences. The mask is then postprocessed by elimination of holes and smoothing of the edge. Finally, color data from the real image within the mask area are mapped onto the object surface.

5. RESULTS

The modelling method has been applied successfully to various image sequences. In Fig. 7 some qualitative results are presented describing the modelling process at the beginning and after having processed 50 input images. The performance in modelling the shape is illustrated by using front and side views of an object, and the ability to produce subjectively satisfactory images is demonstrated by the corresponding synthesized model images in the front view.

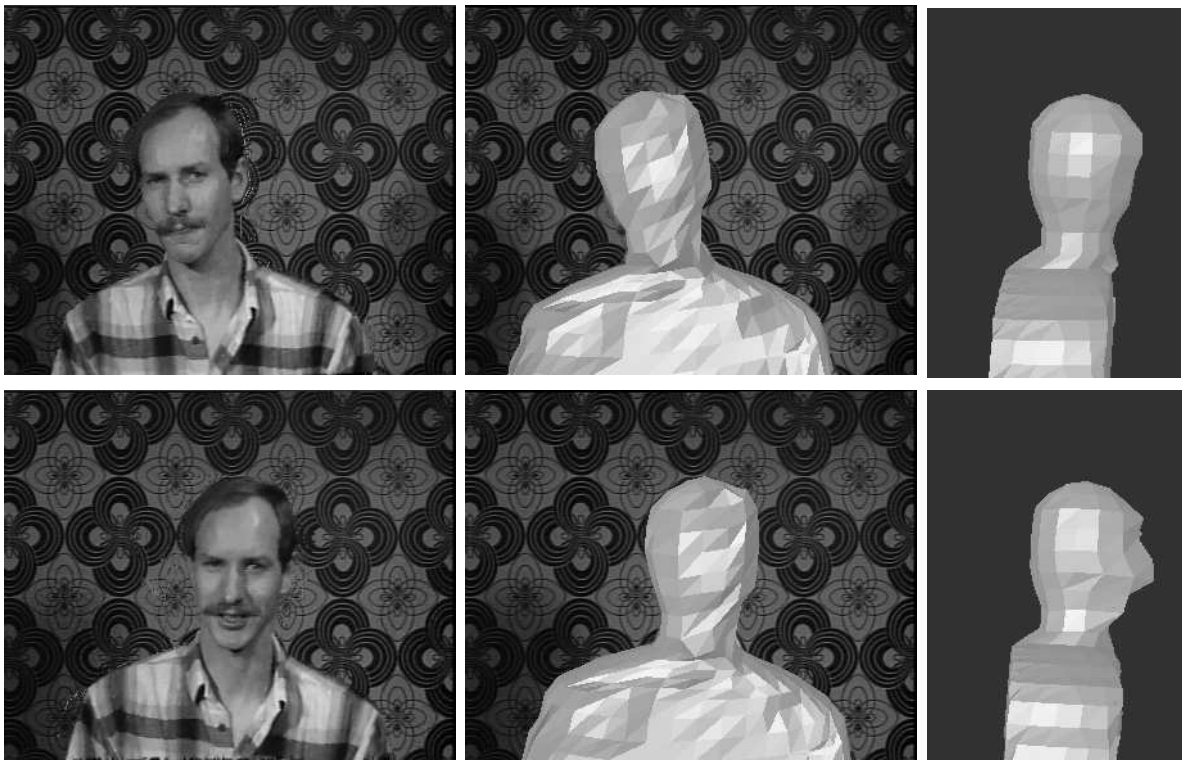

Fig. 7. Modelling results: (a) Synthesized model image and (b) front view and (c) side view of a|b|c object shape after processing of 10 input images; (d) synthesized model image and (e) d|e|f front view and (f) side view of object shape after processing of 50 input images.

6. ACKNOWLEDGEMENTS

The work has been supported by the Forschungsinstitut of the Deutsche Bundespost.

7. REFERENCES

1. J.K. Aggarval, N. Nandhakumar, "On the Computation of Motion from Sequences of Images – A Review", Proceedings of the IEEE, Vol. 76, No. 8, 1988.
2. F. Kappei, C.–Liedtke, "Modelling of a Natural 3–D Scene Consisting of Moving Objects from a Sequence of Monocular TV Images", SPIE, Cannes, 1987.
3. F. Kappei, "Modellierung und Rekonstruktion bewegter dreidimensionaler Objekte in einer Fernsehbildfolge", Ph.D.–Thesis, Universität Hannover, 1989.

4. H. Busch, ”Subdividing Non Rigid 3D Objects into Quasi Rigid Parts”, Proceedings of the 3rd IEE International Conference on Image Processing, University of Warwick, 1989.

5. Waldowski, M., ”Segmentation of a typical video phone scene based on stereo image pairs”, Rank Prize Funds Mini−symposium on Model Based Image Coding. Broadway, England, Dec. 1988.

6. Musmann, H.G., Hötter, M., Ostermann, J., ”Object−oriented analysis−synthesis coding of moving images”, SIGNAL PROCESSING, image communication, No. 2, Nov. 1989.

7. Ostermann, J., ”Modelling of 3D moving objects for an analysis−synthesis coder”, to be published at SPIE/SPSE Symposium on Electronic Imaging, Santa Clara, California, USA, 1990.