

Introduction

A new method for efficient modelling of image sequences has been proposed by Kappei and Liedtke *IM* using an analysis by synthesis algorithm. It is based on extracting the three dimensional information from a sequence of TV images in order to construct a 3D model of the scene. The method can be used for the efficient transmission of TV images to a receiver, where the image sequence has to be reconstructed from the model. The surface of three dimensional objects is modeled with the help of a mesh of triangles containing small triangular parts of the image. An image of the sequence can be reconstructed from the 3D model object by projecting all the triangular images onto a hypothetical camera target.

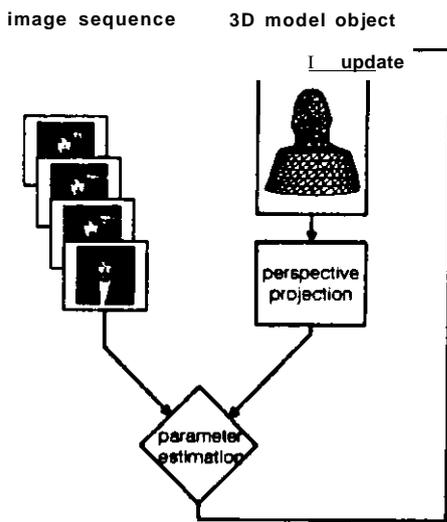


Figure 1, The analysis by synthesis algorithm

A basic algorithm developed by Kappei is the motion estimation algorithm. It estimates the 6 motion parameters for 3D motion of an object by comparing an image from the sequence with the reconstruction from the model. Similar algorithms permit to control the vertices which constitute the shape of the model.

Since the model is composed out of a mesh of triangles, it is even possible to model non rigid objects by moving the vertices which determine the shape of the triangular net. However modeling the motion of a person who is shaking its head strongly by this method is quite difficult and time consuming, since a lot of calculation has to be done in order to extract the motion equations for all the vertices.

Here an approach is put forward to automatically subdivide the model of a flexible object into two or more quasi rigid components automatically and to estimate the motion parameters for each submodel. The algorithm described below does not change the topology of the triangular net, but assigns the triangles of an existing model to different submodels. Triangles of different submodels remain connected to each other so that regions between different submodels are connected flexible to each other.

The subdivision algorithm

The goal of subdividing the model of a flexible object into submodels is to find the largest possible quasi rigid submodels of which the corresponding part of the object moves in a similar direction.

The direct approach to subdivision is to estimate the motion for each triangle separately and unite all triangles with similar motion parameters. However, due to the small area of a triangle, full 3D motion estimation will not lead to accurate motion parameters and hence the direct approach will not work.

Here a two stage approach is applied (see figure 2). Because the area of a triangle is large enough to estimate 2D translative motion, in the first step motion estimation within the image plane is used to search for clusters of triangles with similar motion parameters. In the second step full 3D motion compensation is used to verify the homogeneity of the submodel. If all triangles belong to part of an object moving together, the quality of the reconstruction will increase substantially. Otherwise there will exist no motion parameters which are able to improve the quality of the reconstruction.

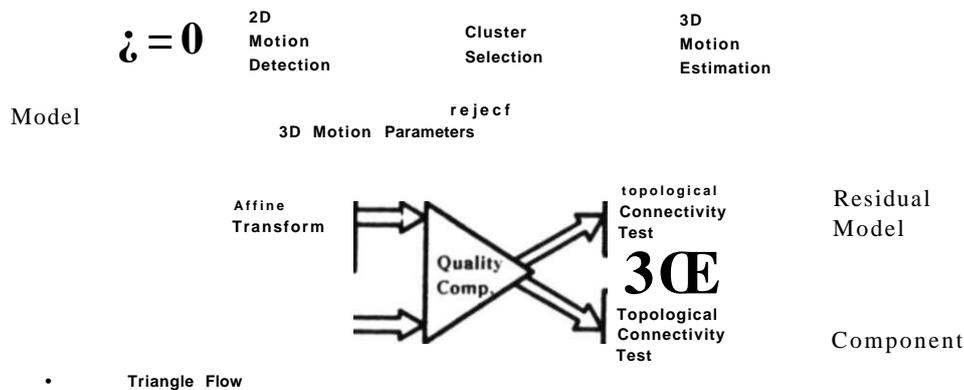


Figure 1 The decomposition algorithm

**The Motion Estimation Algorithm**

Full 3D motion estimation is carried out by estimating the three translation and the three rotation parameters from the model object and an image of the scene. The motion estimation is carried out with the help of so called observation points. Each of these points is assigned its initial 3D position (P) in space, its luminance value (I) and the corresponding gradient values (Gx, Gy). The luminance and gradients are obtained by projecting them out of a reference image onto the 3D model.

Kappei [31] has shown that a linearized set of equations can be setup, which can be solved by the least squares method.

The linearized 3D motion equation for one surface element can be denoted as:

$$\begin{aligned}
 P'_x &= P_x + P_z \cdot \alpha + T_x \\
 P'_y &= P_y + P_z \cdot \beta + T_y \\
 P'_z &= P_z + T_z
 \end{aligned}
 \tag{1}$$

with  $P, P'$  = 3D position before and after movement  
 $R$  = 3D rotation  
 $T$  = 3D translation

The equations for the perspective projection are

$$\begin{aligned}
 B_x &= P_x + F / P_z \\
 B_y &= P_y + F / P_z
 \end{aligned}
 \tag{2}$$

with  $B$  = position in the image plane  
 $F$  = focal length of the camera

The linearized luminance function can be written as

$$I' = I + G_x \cdot (B'_x - B_x) + G_y \cdot (B'_y - B_y)
 \tag{3}$$

with  $I$  = luminance at the location  $B$   
 $I'$  = approximated luminance at the location  $B$

Inserting equations 1 and 2 in 3 leads by linearization to:

$$\begin{aligned}
 I' - I &= F \cdot G_x / P_z + T_x \\
 &+ F \cdot G_y / P_z + T_y \\
 &+ F \cdot (P_x \cdot G_x + P_y \cdot G_y - P_z \cdot G_z) / P_z^2 \cdot T_z \\
 &+ F \cdot (P_x \cdot G_x + P_y \cdot G_y - P_z \cdot G_z) / P_z^2 \cdot R_x \\
 &+ F \cdot (P_x \cdot G_x + P_y \cdot G_y - P_z \cdot G_z) / P_z^2 \cdot R_y \\
 &+ F \cdot (M_x \cdot P_x - Q_y \cdot P_y) / P_z^2 \cdot R_z
 \end{aligned}
 \tag{4}$$

Estimating the 2D translative motion only reduces equation 4 to:

$$\begin{aligned}
 I' - I &= F \cdot G_x / P_z + T_x \\
 &+ F \cdot G_y / P_z + T_y
 \end{aligned}
 \tag{5}$$

**Cluster Formation**

The cluster formation step searches for the largest number of triangles which exhibit a similar 2D motion.

For each triangle the 2D translation vector is estimated separately using equation 5. To reduce the number of motion vectors, all those are selected, which exhibit a substantially greater length than the uncertainty threshold. The uncertainty threshold is calculated from the inverse matrix obtained by the least squares method.

Now all translation vectors are mapped into a three dimensional histogram, the 2D motion probability cube (see figure 3). The three dimensions of the cube represent the rotational compensated translation ( $d_x, d_y$ ) and the rotation within the image plane.

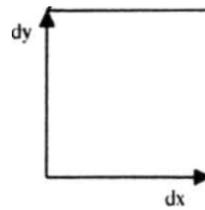


Figure 3. The Probability Density Cube for 2D motion

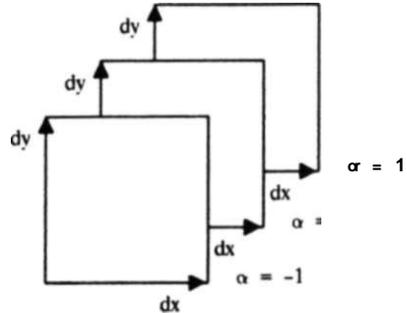


Figure 4. The planes with constant rotation angle alpha

The mapping equations are

$$dy(a) = T_y - P_z \cdot \frac{\sin a}{\sin \alpha}
 \tag{6}$$

Figure 4 and equation 6 show that the 3D histogram can be seen as a set of 2d histograms, each for a constant rotation angle. For one angle each vector supports exactly one translation hypothesis as shown in the example of figure 5 to 7. Figure 6 and 7 show how ten vectors belonging to two components with different motion vectors result in two maxima in the motion distribution cube.

The local maxima in the probability cube constitute a first guess of the 2d motion of objects in the scene. To each triangle, of which the translation vector corresponds to a maximum, the label of the maximum is assigned and the largest cluster of these is selected as an initial component for the 3D step. A cluster is a group of topological connected triangles which belong to the same maximum.

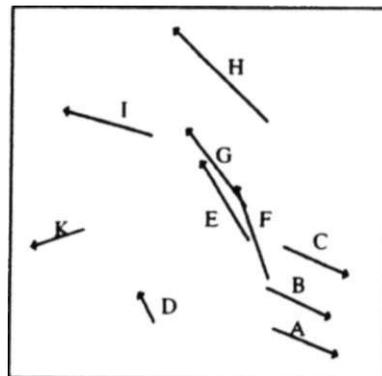


Figure 5. An example with ten 2d translation vectors

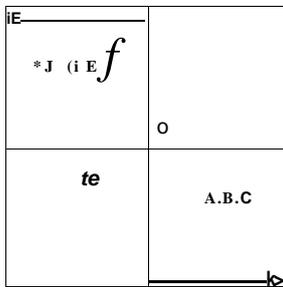


Figure 6. The motion vector distribution of the example of figure 4 for translative motion ( $\alpha = 0^\circ$ )

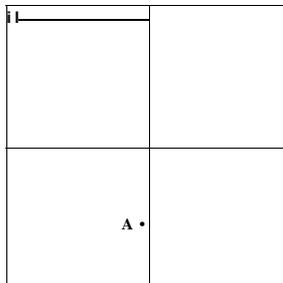


Figure 7. The motion vector distribution of the example of figure 4 for translative motion ( $\alpha = 30^\circ$ )

### 3D Motion Verification

Full 3D motion compensation using six parameters is carried out on the largest cluster found by the 2d step. This motion compensation step must reduce the displaced frame difference (DFD) substantially.

otherwise the cluster is rejected and the 2d step is called to perform a search for another cluster. Is the motion compensation successful on the cluster, the estimated 3D motion is applied to all triangles of the object. For each triangle the DFD after motion compensation is compared with the DFD calculated before. All those triangles are assigned to the new subobject, for which the DFD has decreased.

In a last step all components of the object are checked for topological connectivity of their triangles, to avoid creating components consisting out of separate parts.

### Results

Figure 8 shows an image of the sequence "Miss America" reconstructed from the model by projecting all its triangular images back into the image plane. Figure 9 shows the triangular net of the model of "Miss America". The bright triangles are selected by the motion detection algorithm. Triangles belonging to the same cluster are marked with equal grey levels. The relation between DFD before and after applying 3D motion is displayed in figure 10. Assigning the dark triangles of figure 10 to the residual model and the light ones to the new component results in a decomposition shown in figure 11. At last the compactivity test checks all components of a model for topological connectivity, which leads to the final decomposition of figure 12.

The two DFD images of figure 13 and 14 show the substantial reduction of the frame difference between a test image from the sequence and depiction of the model after subdividing the model into two parts.

It has to be remarked, that the subdivision technique does not use any knowledge about the context but only uses the measured motion. This is the reason, why part of the hair is not correctly assigned to the submodel. It is not our goal to subdivide a person into "head" and "shoulder" but to find the largest possible rigid submodels which model the visible motion of an object.



Figure 8, Image reconstructed out of the model

$$A \quad 4$$

$$\beta t \quad k \gg S$$

$$WW'*$$

Figure 9, The clustered triangles (each graylevel corresponds to a cluster)

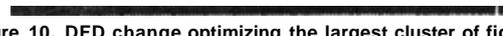


Figure 10. DFD change optimizing the largest cluster of figure 9 (light = better, dark = worse)

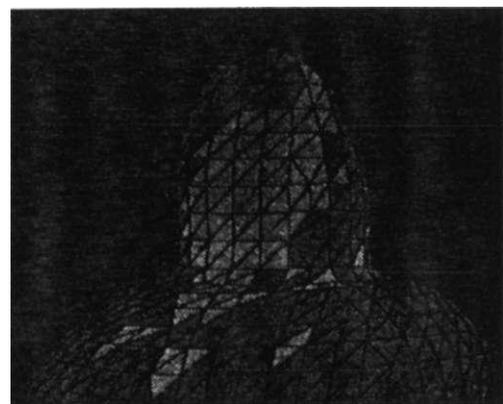


Figure 11. The DFD of the highlighted triangles was decreased by compensating 3D motion for the largest cluster of figure 9.

- m



Figure 12. Final decomposition after connectivity test



Figure 14. DFD after motion compensation of the decomposed model

#### References

- 11/ F. Kappeler, C.-E. Liedtke. 1987. Modelling of a natural 3D scene consisting of moving objects from a sequence of monocular TV images. Proc. SPIE Volume 860, pp. 126.
- 12/ F. Kappeler, C.-E. Liedtke. 1987. Ein Verfahren zur Modellierung von 3D-Objekten aus Fernsehbildfolgen. Proc. of the 9. DAGM-Symposium, Oktober 1987.
- 13/ F. Kappeler. 1988. Modellierung und Rekonstruktion bewegter dreidimensionaler Objekte in einer Fernsehbildfolge. PhD Thesis. University of Hannover. 1988.

Figure 13, DFD of motion compensated model before decomposition