Automatic analysis of flexibly connected rigid 3D objects for an OBASC

Geovanni Martínez
Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung
Universität Hannover, Appelstraße 9A, 3000 Hannover 1, F.R. Germany
Phone: ++49–511–762–5311, email: martinez@tnt.uni–hannover.de

Abstract

An algorithm for automatic analysis of flexibly connected rigid 3D object components is described. The developed algorithm is used for image analysis in an object–based analysis–synthesis coder (OBASC)[2] based on the source model of flexibly connected rigid 3D object components. To find the flexibly connected 3D object components, a local 3D motion estimation technique, using a motion compensated Kalman Filter and a local luminance error model, is applied. In order to improve the accuracy and the reliability of the estimates, results obtained in previous frames are considered for the actual analysis. The efficiency of the developed algorithm is measured by the amount of model failures.

1. Introduction

To model 3D scenes from 2D image sequences for purposes of reconstruction, recognition, or compression[2], a model–based image analysis subdivides each image of a sequence into moving objects and describes each object O_m by three sets of parameters defining its motion A_m, shape M_m and color S_m. The parameter sets depend on the type of source model being applied. Image regions where the assumptions of the applied source model do not hold are called model failures and are represented by MF–objects. In an object–based analysis–synthesis coder (OBASC)[2] the size of the MF–objects define its efficiency because the color parameters must be coded and transmitted for MF–objects only. In this contribution, the source model of flexibly connected rigid 3D object components is used. This source model was first defined in [2][4][5] for an OBASC. Moving real objects are modelled by moving model objects. According to this source model, the 3D shape M_m of each real object O_m is represented by a mesh of triangles which is put up by vertices denoted as control points, its color S_m is defined by the luminance and chrominance reflectance of the model object surface, and its motion A_m is defined by a set of 6 parameters which describe the translation and rotation of the object in the 3D space. In the case of flexibly connected rigid 3D object components each component is described by its own set of parameters. Since the shape of 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.

In order to find 3D rigid objects and their flexibly connected 3D object components, the image analysis described in [4][5] carries out a change detection to distinguish between temporally changed and unchanged regions of the two first images of the sequence, and assumes that each changed region represents one moving object together with background uncovered due to object motion. Then, a rigid 3D model object represented by a mesh of triangles is generated for each changed region assuming a 3D ellipsoidal shape (Fig. 1). In order to find the rigid 3D object components, global 3D motion estimation and compensation for each rigid 3D model object is carried out. Since connected 3D real objects are described by only one rigid 3D object, residual synthesis errors will remain because they cannot be compensated by the current model object. Then, the 3D object components are found by the algorithm described in [1] which estimates the residual local 2D motion of each triangle of the 3D model object after 3D global motion compensation. It unites all triangles with large residual local 2D motion into one region. The largest region of triangles which can reduce the residual synthesis errors is defined to be a 3D object component. However, this component normally does not represent a ”true” 3D model object component, although it can improve the image synthesis and reduce the MF–objects. This false subdivision of the rigid 3D model objects causes that MF–objects to be larger than those resulting from a correct subdivision.

In this contribution a new algorithm is described which automatically generates a more reliable subdivision of each rigid 3D model object into rigid 3D model object components, for example a subdivision of a person into ”head” and ”shoulders” without apriory knowledge of the scene. To find the flexibly connected 3D object components, a 3D motion estimation technique of the residual local 3D motion of each rigid 3D object is applied, after global 3D motion compensation, using a motion compensated Kalman Filter[3] and a local luminance error model. The approach is to unite all neighboring triangles with similar residual local 3D motion parameters. Additionally, in order to improve the accuracy and the reliability of the subdivision, results obtained in previous frames are considered for the current analysis. In section 2, this new algorithm is described in detail. In section 3, an OBASC, based on the source model of flexibly connected rigid 3D object components, using this new algorithm is explained. In section 4, first results are discussed.

2. Search of flexibly connected rigid 3D object components.

In order to find the 3D object components, the following algorithm is used in this contribution. The algorithm applies four steps to each frame of the image sequence until each 3D object component is found. By the first step, a global 3D motion estimation and compensation for each 3D model object is carried out. Because connected 3D real objects are treated only as one rigid 3D object, residual synthesis errors may remain after motion compensation because they cannot be explained by the current model object. By the second step, the residual local 3D motion for each triangle of the 3D model object after its global 3D motion compensation is estimated. Then, the local 3D motion parameters of neighboring triangles are compared and triangles with similar local 3D motion are combined into regions. Two neighboring triangles are combined into the same region if the following criterion applies:

$$C = \begin{cases} |(P_{1c} - P_{2c})/(P_{1c} + P_{2c})| \le th \;\; \forall \;\; c = x,y,z \;\; \text{and} \;\; P = T,R & \rightarrow \textit{combined} \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \textit{else} & \rightarrow \textit{not combined} \end{cases} \tag{1}$$

Here, $\boldsymbol{T}_i = (T_{ix}, T_{iy}, T_{iz})$ and $\boldsymbol{R}_i = (R_{ix}, R_{iy}, R_{iz})$ are the estimated local 3D translation and rotation parameters of the triangle *i*. The threshold *th* is considered constant during the sequence and its typical value is 0.4.

In an ideal case, the regions which are found by the second step will represent the 3D object components. However, due to errors of the local 3D motion estimation or insufficient motion of the real object, sometimes only parts of the components can be found. Therefore, in order to find the complete 3D object components, the third and fourth steps are used. The third step tries to unite neighboring regions obtained by the second step into a larger region. For that purpose, because motion estimation for regions is more reliable, the local 3D motion parameter of each region is estimated and again the local 3D motion in this case of neighboring regions is compared to determine if they can be joined to form a new larger region. In order to compare neighboring regions, the same criterion Eq. (1) with $th = 0.3$ is used. By the fourth step, results obtained in previous frames are considered for the current analysis. For this purpose, because previously estimated regions and the current regions have a high similarity in term of shape and position, the large regions found by the third step are used to improve regions stored in a region–memory attached to each triangle of the wire frame. These stored regions were obtained by the analysis of the previous frames and represent regions which are demonstrated to be probable candidates for 3D model object components. Each stored region is improved during analysis of the sequence until it can be recognized, if possible, as a complete rigid 3D object component. As a first approach, an improved stored region is recognized as a 3D model object component if, after *n* frames of the sequence, it can not be further increased and the corresponding new regions, which were found during these *n* frames, were least 90% inside of it. *n* was set to 2. An improvement of a region means that a region in memory can be increased or decreased depending on the position, size and texture of the new regions, which are found by the analysis of the current frame. As a first approach, the stored regions are only enlarged and compared to the new region with respect to position and size.

When a region is identified as a 3D model object component, its subdivision results a 3D model object with two components. The subdivision does not change the topology of the triangle net of the 3D model object, but assigns some of the control points of the 3D model object to the new 3D model object component. The components remain connected to each other so that regions between these two 3D model object components are flexibly connected to each other. In the next frames, the 3D model object components are considered as a new 3D model object to look for more 3D model object components.

For estimation of the global 3D motion, the shape of the objects is assumed to be known. A gradient method [4][5] is applied. It uses only a set of observation points from each 3D model object. Each observation point is located on the model object surface and is described by its luminance value and its spatial linear gradients. The luminance and gradients are taken from the same image from which the color parameters were derived. The measure for selecting observation points is a high spatial gradient. For each observation point, a linearized equation is set up for motion estimation. The set of equations is solved by an iterative least mean square method.

For estimation of the local 3D motion of a triangle or a region, neighboring triangles are also considered, because the small number of observation points belonging to a triangle or region does not lead to an accurate estimation. For each observation point the same linearized equation as for global 3D motion estimation is set up for local 3D motion estimation. To solve the set of equations a motion compensated Kalman filter using a local luminance error model is applied, because the iterative least square method does not produce a reliable local 3D motion estimation to ensure a reliable subdivision of a rigid 3D model object into rigid 3D object components. In a motion compensated Kalman filter[3] a local 3D motion estimation of the triangle or region and its neighboring triangles is carried out before the Kalman filter is applied. This is achieved by solving the set of equations by the iterative least square mean method. Then, to improve the accuracy of these motion estimates, a Kalman filter is applied using a local luminance error model for each observation point. Only those observation points are used by the Kalman Filter where the local luminance error model was found to be valid. The luminance error model represents both the projection into the image plane of the 3D position error (shape error) and image noise of each observation point. As a first approach, the shape error is modelled by a gaussian stationary random process describing the position error in the x, y and z directions of each observation point. These position errors are considered to be uncorrelated.

3. An OBASC based on the source model of flexibly connected rigid 3D object components using the new algorithm.

Fig. 3 shows the structure of an OBASC. An OBASC subdivides the current real image s_{k+1} into moving objects and encodes each object by three sets of parameters defining the motion A_{k+1}, shape M_{k+1} and color information S_{k+1} of the object. The objects and their parameters are obtained by image analysis. The inputs to the image analysis are the current real image s_{k+1}, a synthesized image S_k', which is generated by the transmitted parameters A_k', M_k', S_k', and the transmitted parameters A_k', M_k', S_k'. Then, the calculated parameter sets are coded by parameter coding. The parameter coding depends on a receiver model which includes assumptions about the visibility of coding errors. Using the encoded and transmitted parameters, an image S_k' can be reconstructed by image synthesis at the decoder as well as at the coder. This OBASC differs from the OBASC described in [2][3] by its image analysis which uses the new algorithm proposed in this contribution to find the rigid 3D objects and their flexibly connected rigid 3D object components.

4. Experimental results

Experiments using the videophone test sequence ”Claire” (CIF, 10Hz) show that the average area of model failures can be reduced from 4% of the image area, obtained by the OBASC described in [4], to 3% obtained by the OBASC described in this approach (Fig. 4). At transmission rate of 64 kbit/s this allows to increase the bit rate available for coding for color parameter of model failures from 1.3 to 1.5 bit/pel and to improve the picture quality. Fig. 2 shows the flexibly connected rigid 3D object components which were found by the algorithm described in this approach for the same test sequence. The 3D object component ”head” was recognized and subdivided from image 11 onwards.

5. Conclusions

A method for automatic analysis of flexibly connected rigid 3D model object components is proposed. It uses a 3D motion estimation technique of the residual local 3D motion of each rigid 3D object and results obtained in previous frames to make a realistic subdivision of 3D model objects into connected rigid 3D model object components, for example a person into ”head” and ”shoulders”. This realistic subdivision reduces the model failures from the image sequence ”Claire” from 4% to 3% of the image area without applaing any knowledge of scene content.

6. References

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

[2] H. G. Musmann, M. Hötter, J. Ostermann, ”Object–oriented analysis–synthesis coding of moving images”, Signal Processing, Image Communication, Vol. 1, No 2, Nov. 1989.

[3] I. Sezan, R. Lagendijk, ”Motion analysis and image sequence processing”, Kluwer Academic Publishers, Massachusetts, USA, 1993.

[4] J. Ostermann, ”An analysis–synthesis coder based on moving rigid 3D–objects”, accepted in publication to Signal Processing: Image Communication.

[5] J. Ostermann, ”Modelling of 3D–moving objects for an analysis–synthesis coder,” SPIE/SPSE Symposium on Sensing and Reconstruction of 3D Objects and Scenes ’90, B. Girod, Editor, Proc. SPIE 1260, pp.240–250, Santa Clara, California, February 1990.

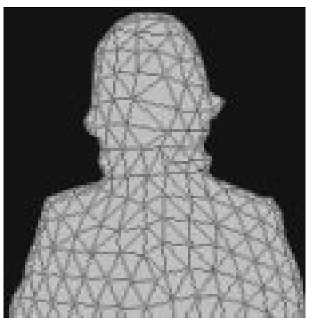

Fig. 1
Initial rigid 3D Object for the image analysis of the testsequence CLAIRE

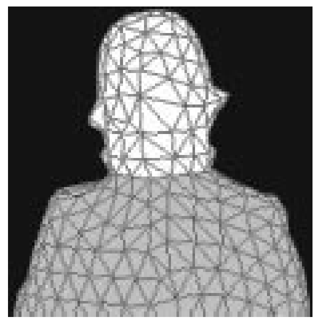

Fig. 2
Flexibly connected rigid 3D object components of the test sequence CLAIRE obtained by the algorithm described in this approach
The subdivision was possible from image 11 onwards

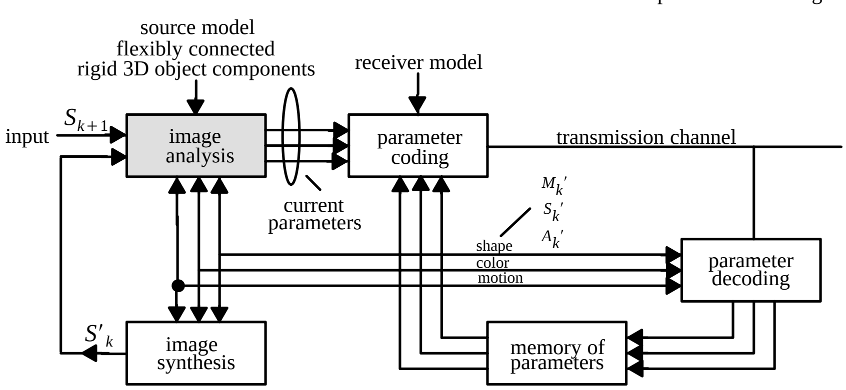

Fig. 3 Structure of an OBASC based on the source model of flexibly connected rigid 3D object components

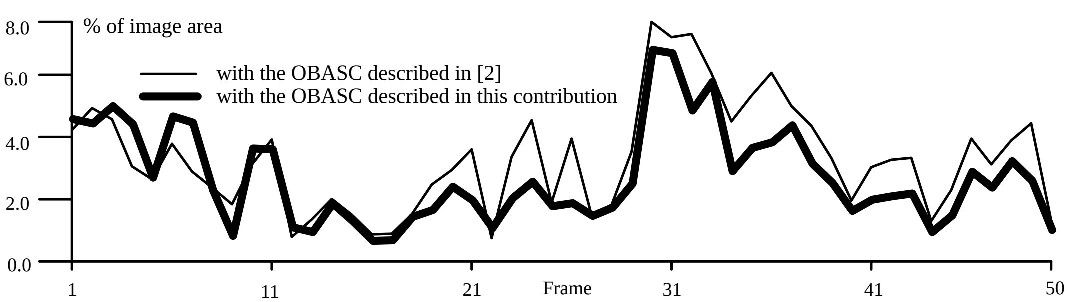

Fig. 4 Area of model failures in pel for the test sequence CLAIRE. The average area of model failures is reduced form 4% to 3% of the image area by the algorithm described in this approach