Analysis–Synthesis Coding Based on the Source Model of Articulated Three–Dimensional Objects

Geovanni Martinez
Escuela de Ingeniería Eléctrica, Universidad de Costa Rica
2060, San José, Costa Rica
email: gmartin@pacuare.eie.ucr.ac.cr

ABSTRACT

An object–based analysis–synthesis coder for coding moving images at low data rates is proposed. The coder is based on the source model of articulated three–dimensional objects. This model describes the real objects by means of model objects defined by shape, motion and color parameters. The model objects may be articulated, i.e., they consist of several rigid object components linked to each other by joints. Since in many video applications real objects are articulated, the new source model allows a better description of shape and motion of real objects particularly by the different motion of object components. Algorithms for estimation and coding of model parameters are presented. The coder was applied to the test sequences Claire (CIF, 10 Hz) and Miss America (CIF, 10 Hz). Experimental results show that, at a fixed image quality figure measured by using SNR, the new source model achieves a data rate reduction of up to 20% in comparison to the source model of rigid three–dimensional objects and of up to 10% in comparison to the source model of flexible three–dimensional objects.

1. INTRODUCTION

For coding of moving images at low data rates object–based analysis–synthesis coding is investigated[5]. An object based analysis–synthesis coder subdivides each image of a sequence into arbitrarily shaped moving objects and describes each object by a model object. A model object is defined by three sets of parameters: motion, shape and color parameters. Color parameters denote the luminance as well as the chrominance reflectance on the object surface. For each object only shape and motion parameters are coded and transmitted. An image is reconstructed by image synthesis using the current shape and motion parameters and the already transmitted color parameters. Image regions which can not be reconstructed with sufficient image quality are called Model Failure objects or MF objects. For each MF object its two–dimensional (2D) shape and its color parameters must be coded and transmitted. Since the transmission of color parameters is expensive in terms of data rate, the total size of all MF objects should be kept as small as possible.

For object based analysis–synthesis coding two different three–dimensional (3D) source models have been proposed: the source model of rigid 3D objects (R3D)[6] and the source model of flexible 3D objects (F3D)[7]. By the source model of rigid 3D objects the shape of an object is represented by a rigid wire–frame (see Fig. 1.a.iii). The vertices represent the shape parameters. The wire–frame is completely described by the 2D object silhouette, i.e., there is an algorithm which computes a generalized cylinder from the silhouette[6]. The object silhouette is estimated by a change detection between the current image and the last transmitted image. The 3D motion of an object is described by one set of six motion parameters: three rotation angles and one 3D translation vector (see Fig. 1.a.ii). The color parameters are defined by projecting a real image onto the wire–frame (see Fig. 1.a.i). While the source model of rigid objects describes only the shape of rigid objects, the source model of moving flexible objects allows also a local deformation of the objects tangential to the object surface. This deformation is carried out by shifting vertices of the wire–frame and is described by shift vectors (see Fig. 1.b.iii). The rest of the model parameters is defined as by the source model of rigid 3D objects (see Fig. 1.b.i and 1.b.ii).

In many video applications real objects are articulated, i.e. they consist of several connected object components. The source model of rigid objects fails when the object components of an articulated object move differently. In this paper, an object based analysis–synthesis coder based on the source model of moving articulated 3D objects (A3D) is developed. With a more realistic description of shape and motion of real objects particularly by different motion of object components, a reduction of data rate for the transmission of color parameters is expected.

According to the source model of articulated 3D objects, an articulated object consists of several connected object components (see Fig. 1.c). Each object component is described by its own shape, motion and color parameters. The shape of an object component is supposed to be rigid and described by a rigid wire–frame whose vertices represent the shape parameters of the object component (see Fig. 1.c.iii). In order to reduce the date rate for the transmission of shape parameters, the shape of the object components will be estimated by subdividing a first rigid model object into model object components. The object components are connected to each other by spherical joints (see Fig. 1.c.iii). A spherical joint is considered here to be a part of the shape of an articulated object. A spherical joint position is described by three coordinates. The 3D motion of an object component is also described by six motion parameters (see Fig. 1.c.ii). Since for each object component one set of motion parameters is transmitted, the data rate for the transmission of motion parameters will also increase. The color parameters of an object component are also defined by projecting a real image onto its wire–frame (see Fig. 1.c.i). Even though in the new source model the data rate for transmission of shape and motion parameters increases, a gain can still be achieved because a larger reduction is expected in the data rate for the transmission of color parameters.

In order to implement an object–based coder based on the new source model of articulated 3D objects, some algorithms for the estimation and coding of model parameters have to be developed.

A comparison with previously known 3D source models will be carried out. Therefore, the image quality of coded image sequences will be first measured and then compared. In order to measure the image quality, the objective criterion PSNR (Peak Signal to Noise Ratio) will be used.

This paper is organized as follows: in Section 2 the algorithms for estimating the model parameters are described. In Section 3 the algorithms for coding the model parameters are presented. In Section 4 experimental results for real image sequences are given. The conclusions are presented in Section 5.

2. ESTIMATION OF MODEL PARAMETERS

For estimation of the model parameters, the luminance component of the current frame and a rigid model object estimated by analysis of previous frames are input. The shape of the rigid model object is represented by a rigid wire–frame. The wire–frame is generated from the object silhouette as by the source model of rigid objects. The silhouette is estimated by a change detection of the first two images of the image sequence.
In the first step, the parameters of the 3D motion of each triangle of the rigid model object are estimated. For the motion estimation, a maximum likelihood estimator is applied which models the measured values statistically and also detects outliers in them[1][2].

In the next step, called object articulation, the rigid model object is subdivided into object components. Therefore, neighboring triangles which exhibit similar 3D motion parameters are clustered into surface patches[1][2][3]. In an ideal case, the patches will represent the complete object components. However, due to the unreliability of motion estimation of single triangles or the insufficient motion of the object components, only parts can be found. In order to find all object components completely, clustering results obtained in previous frames are taken into account for the current clustering. Each triangle has a memory which indicates its membership into different surface patches. A surface patch found by the current analysis gets a new register number if it does not overlap any of the already registered surface patches. In case of overlapping, all triangles of the new patch get the register number of the already registered surface patch. This increases the surface of the already registered patch. For stability reasons, if the surface increases more than 30% of the object surface, the fusion will not be carried out. A registered patch will be detected as an object component if its surface does not increase after 2 consecutive frames and if it is larger than the 10% of the object surface. If an object component is detected, the corresponding vertices of the rigid model object are assigned to the new object component. The wire–frames of two connected object components remain flexibly connected to each other by those triangles having vertices belonging to different object components.
After object articulation, the position of each spherical joint is estimated from motion. Two consecutive sets of motion parameters for each object component are then estimated. The first set describes its motion from the last but one transmitted frame to the last transmitted frame. The second set describes its motion from the last transmitted frame to the current frame. For the motion estimation, a Maximum Likelihood estimator is applied. A spherical joint position is then determined by evaluating the current position and the estimated sets of motion parameters of the object components linked to each other by the joint in the equations representing the constraints imposed by the joint on their relative motion[1]. Motion estimation errors decrease the reliability of joint position estimation. In order to improve reliability, more than two sets of motion parameters for each object component are estimated. The position of the joint is then determined by linear regression[1].
After the position of all spherical joints are known, the definitive estimation of the motion parameters of each object component is carried out but now considering the spherical joints[1][4]. This improves the accuracy of motion estimation[1]. The articulated object is then considered as a graph with a tree structure whose root

is the largest object component. For each object component only the independent motion parameters are estimated. The rotation parameters are the independent motion parameters of an object component. The translation and rotation parameters of the root object component are first estimated using a maximum likelihood estimator. Only the rotation parameters for the rest of the object components are then estimated beginning from the root object component one after the other. For the estimation of the rotation parameters of an object component two steps are applied. Firstly, the translation parameters are calculated by evaluating the position and motion parameters of the previous object component in the equations representing the constraints imposed by the spherical joint on the relative motion of both object components. The object component is then moved using the calculated translation parameters. Secondly, the rotation parameters are estimated using a maximum likelihood estimator.

3. CODING OF MODEL PARAMETERS

A polygon/spline approximation is applied for the object silhouette coding[6]. For the coding of the subdivision into object components, a number for each vertice of the wire–frame indicating the membership to an object component is run–length–coded and transmitted[1]. For the coding of a joint position, its three coordinates are uniformly quantized and then transmitted in PCM representation[1]. For motion parameter coding, both the rotation and translation parameters for the largest object component and only the rotation parameters for the rest of the object components are uniformly quantized and transmitted in PCM representation.

The 2D shape of the MF objects is also coded by polygon/spline approximation. The texture parameters are coded using a Discrete Cosine Transformation (DCT) for arbitrarily shaped regions[6].

4. EXPERIMENTAL RESULTS

The object based coder described in this contribution was applied to the test sequences Claire and Miss America with spatial resolution CIF and a reduced frame rate of 10 Hz. Each sequence shows an articulated object (human person) consisting of the head and shoulders (object components), which move differently from frame to frame.

At an image quality of PSNR=38 dB, a coder based on the A3D model achieves a data rate of 5000 bit/image approximately. At the same image quality figure, a coder based on the F3D model and a coder based on the R3D model achieves a data rate of 5600 bit/image and 6400 bit/image, respectively[1][6][7]. Even though more information is needed for the description of the motion and shape of real objects in the A3D model, a gain could be achieved. This is due to a more precise description of the shape and motion of real objects particularly by the different motion of the object components. The new A3D model is more efficient than the F3D model because in this last model the shift vectors must also be transmitted and this is costly in terms of data rate[7].

5. CONCLUSIONS

An object–based analysis–synthesis coder based on the source model of articulated 3D objects is presented for coding moving images at low data rates.

For estimating the shape of an articulated object a three–dimensional rigid shape from the object silhouette is first calculated. This is then articulated in object components by clustering small surface elements with similar motion parameters into an object component. For estimating a joint position, a linear regression–based method is applied, which makes use of relative motion constraints of the object components connected to the joints. For a more precise estimation of the motion parameters of an object component, joints are also taken into account in motion estimation.

Experimental results show that, at the same image quality figure, this new source model achieves a data rate reduction of up to 20% in comparison to the source model of rigid three–dimensional objects and of up to 10% in comparison to the source model of flexible three–dimensional objects.

ACKNOWLEDGEMENT

The author is grateful to Prof. Musmann for his invaluable advice.

6. REFERENCES

[1] G. Martínez, ”Analyse–Synthese–Codierung basierend auf dem Modell bewegter, dreidimensionaler gegliederter Objekte”, Ph.D Thesis, University of Hannover, Germany, February, 1998.

[2] G. Martínez, ”Shape Estimation of Articulated 3D Objects for Object–Based Analysis–Synthesis Coding (OBASC)”, Signal Processing: Image Communications, Vol. 9, No. 3, pp. 175–199, March1997.

[3] G. Martínez, ”Shape Estimation of Articulated 3D Objects considering mutual occlusions for object–based analysis–synthesis coding (OBASC)”, Picture Coding Symposium 96 (PCS’96), Melbourne, Australia, pp. 141–146, March 1996.

[4] G. Martínez, ”3D Motion Estimation of Articulated 3D Objects for Object–Based Analysis–Synthesis Coding (OBASC)”, International Workshop on Coding Techniques for Very Low Bit–rate Video, Tokyo, Japan, No. G–1, November 1995.

[5] H. G. Musmann, M. Hötter, J. Ostermann, ”Object–Oriented Analysis–Synthesis Coding of Moving Images”, Signal Processing: Image Communication, Vol. 1, No. 2, pp. 117–138, October 1989.

[6] J. Ostermann, ”Object–Based Analysis–Synthesis Coding Based on the Source Model of Moving Rigid 3D–Objects”, Signal Processing: Image Communication, No. 6, pp. 143–161, 1994.

[7] J. Ostermann, ”Object–Based Analysis–Synthesis Coding (OBASC) Based on the Source Model of Moving Flexible 3D–Objects”, IEEE Trans. on Image Processing, Vol. 3, No. 5, September 1994.

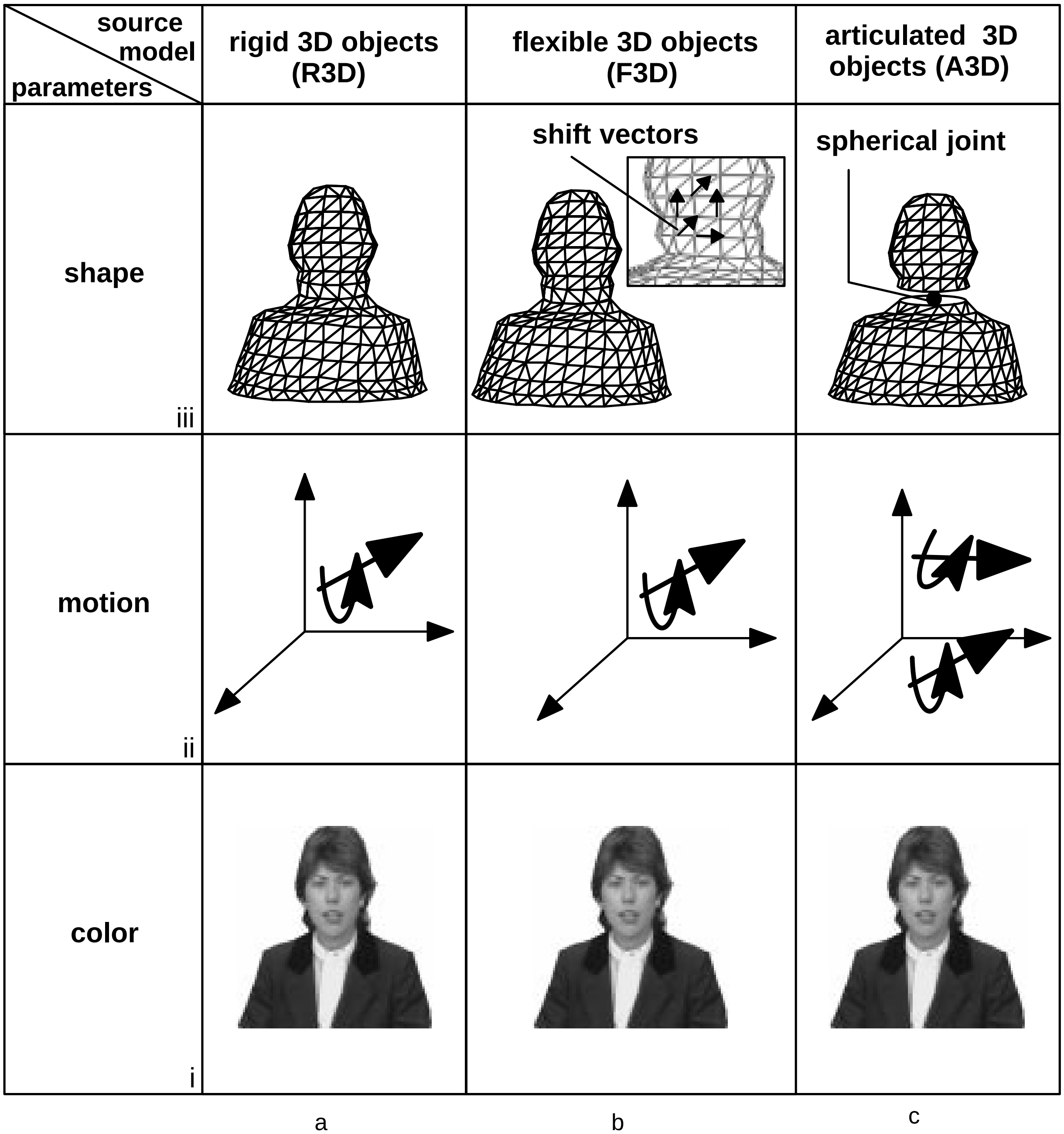

Fig. 1 Model parameters of the source model of rigid 3D objects (a), flexible 3D objects (b) and articulated 3D objects (c)