3D Motion Estimation of Articulated Objects for Object–Based Analysis–Synthesis Coding (OBASC)

Geovanni Martínez
Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung
Universität Hannover, Appelstraße 9A, D–30167 Hannover , F.R. Germany
Phone: ++49–511–762–5311, email: martinez@tnt.uni–hannover.de

Abstract

An algorithm for motion estimation of articulated 3D objects for object–based analysis–synthesis coding (OBASC) is presented. For motion estimation an articulated object is first decomposed automatically into flexibly connected object–components. Each rigid 3D object–component is assumed to be connected to the object by spherical joints. Then, the 3D motion of the largest object–component is estimated without considering the other components. Finally, the motion analysis is propagated to the rest of the object–components taking into account the spatial constraints enforced by the spherical joints. The developed algorithm has been incorporated in an OBASC which uses articulated 3D model objects. For the standard videophone test sequence Claire (CIF, 10Hz) the transmission rate decreases from 53 kbit/s to 48 kbit/s at a fixed image quality by applying the proposed motion estimation algorithm.

1. Introduction

For coding of moving images at low data rates object–based analysis–synthesis coding (OBASC)[8] is investigated. An OBASC scheme describes each image of a sequence by moving model objects. Each model object is defined by three sets of parameters defining its motion, shape and color. Color parameters denote the luminance as well as the chrominance reflectance on the object surface. The sets of parameters depend on the applied source model and have to be estimated automatically. Image regions where the description by the applied source model fails are called **M**odel **F**ailure regions (MF–regions). Color parameters are transmitted for MF–regions only. Since the transmission of color parameters is expensive in terms of data rate, the total size of all MF–regions should be kept as small as possible.

Ostermann[9] proposes an OBASC scheme based on a source model of "moving rigid 3D objects" (OBASC_{R3D}). According to this source model, model objects are rigid with 3D shape and moving in the 3D space. The motion is defined by a set of 6 parameters which describe the translation and rotation of the object in the 3D space. The 3D shape is represented by a mesh of triangles which is put up by vertices denoted as control points. The color parameters are taken by projection of a real image onto the surface of the mesh of triangles. Objects may be articulated, i.e. may consist of two or more flexibly connected rigid 3D object–components. In computer graphics, object–components are sometimes called links. Each object–component has its own set of motion, shape and color parameters. Since the shape of each object–component is defined by its control points, object–components are connected by those triangles having control points belonging to different object–components. Due to these connecting triangles, object–components are flexibly connected. Connecting triangles may enforce constraints on the spatial location of object–components.

For motion estimation of articulated objects a hierarchical coarse to fine approach is proposed by Koch[4] and used by Ostermann[9], Kampmann[3] and Martínez[5][7]. However, motion estimation fails when the object–components have strongly different motions, because no spatial constraints are considered, i.e. the connecting triangles do not enforce a constraint on the spatial location of the object–components. In a paper by Hsu[2] the 3D motion estimation of a person's arm is calculated, however no spatial constraints are taken into account. Holt[1] uses spatial constraints in order to improve motion estimation of articulated objects. Therefore, the object is first manually decomposed into simple articulated subparts. Each subpart contains a small number of object–components. Components conforming a subpart are confined to motion within a plane (coplanar motion) and connected to each other by revolute joints[10] i.e. the spatial constraints between two components are modeled by one revolute joint. A revolute joint allows only relative angular rotation between components about the revolute joint axis which is perpendicular to the motion plane and does not allow that components may rotate themselves. Motion estimation determines first the motion of the most simple subpart(s) and then propagates the analysis to the remaining subparts of the object. The estimation algorithm evaluates more than two consecutive frames of the image sequence to generate one estimate.

In this contribution an algorithm for motion estimation of articulated objects is presented which considers more sophisticated spatial constraints. For motion estimation, the object is first automatically articulated into flexibly connected object–components using the method for object–articulation proposed in [5][7]. Each subpart contains a single object–component. Object–components are connected together by spherical joints[10] instead of revolute joints. A spherical joint allows non restricted relative angular rotations between two object–components and that object–components may rotate themselves. Motion estimation determines first the motion of the largest object–component without considering spatial constraints. Then, motion analysis is propagated to the remaining object–components taking into account the spatial constraints on the articulated object. For motion estimation of an object–component a gradient method is applied which requires the evaluation of two consecutive frames of the image sequence only in order to estimate the motion parameters.

The performance of the developed algorithm is evaluated in an (OBASC)[9] scheme in terms of coding efficiency. The coding efficiency is measured by the reduction of the bit rate at a fixed image quality measured by PSNR.

The paper is organized as follows. In Section 2, the proposed algorithm for 3D motion estimation of articulated objects is presented. In Section 3, experimental results are given. Final discussions are presented in Section 4.

2. Algorithm for 3D motion estimation of articulated objects

In this section, a new algorithm for motion estimation of articulated 3D objects is presented. Fig. 1 represents the stick model of typical articulated object. For motion estimation, object–components are assumed to be rigid and their shape known. The shape of articulated objects can be estimated using the algorithm proposed by Martínez[5][7].

Let m_a and m_b be two object–components connected by a spherical joint (see Fig. 2). The 3D motion of the object–component m_a is described by the motion parameters $\boldsymbol{A}^{m_a} = (T_x^{m_a}, T_y^{m_a}, T_z^{m_a}, R_x^{m_a}, R_y^{m_a}, R_z^{m_a})^T$ defining its translation and rotation in the 3D space. An arbitrary point $\boldsymbol{P^{(i)}}$ on the surface of an object–component m_a with $\boldsymbol{N^{m_a}}$ control points $\boldsymbol{P}^{(i)}_{Cm_a}$ is moved to its new position $\boldsymbol{P}'^{(i)}$ according to the motion equation:

$$\boldsymbol{P}'^{(i)} = [\boldsymbol{R}_{Cm_a}] \cdot (\boldsymbol{P}^{(i)} - \boldsymbol{C}^{m_a}) + \boldsymbol{C}^{m_a} + \boldsymbol{T}^{m_a} \tag{1}$$

with the translation vector $\boldsymbol{T}^{m_a} = (T_x^{m_a}, T_y^{m_a}, T_z^{m_a})^T$, the rotation angles $\boldsymbol{R}^{m_a} = (R_x^{m_a}, R_y^{m_a}, R_z^{m_a})^T$, the object–component center $\boldsymbol{C}^{m_a} = (C_x^{m_a}, C_y^{m_a}, C_z^{m_a})^T = (1/N^{m_a}) \cdot \sum_{i=0}^{N^{m_a}-1} \boldsymbol{P}^{(i)}_{Cm_a}$ and the 3D rotation matrix:

$$[\boldsymbol{R}_{Cm_a}] = \begin{pmatrix} \cos R_y^{m_a} \cos R_z^{m_a} & \sin R_x^{m_a} \sin R_y^{m_a} \cos R_z^{m_a} - \cos R_x^{m_a} \sin R_z^{m_a} & \cos R_x^{m_a} \sin R_y^{m_a} \cos R_z^{m_a} + \sin R_x^{m_a} \sin R_z^{m_a} \\ \cos R_y^{m_a} \sin R_z^{m_a} & \sin R_x^{m_a} \sin R_y^{m_a} \sin R_z^{m_a} + \cos R_x^{m_a} \cos R_z^{m_a} & \cos R_x^{m_a} \sin R_y^{m_a} \sin R_z^{m_a} - \sin R_x^{m_a} \cos R_z^{m_a} \\ -\sin R_y^{m_a} & \sin R_x^{m_a} \cos R_y^{m_a} & \cos R_x^{m_a} \cos R_y^{m_a} \end{pmatrix} \tag{2}$$

which defines the rotation in the mathematically positive direction around the x–, y– and z–axis with the rotation center $\boldsymbol{C}^{m_a}$. A similar motion equation can be written for the object–component m_b:

$$\boldsymbol{P}'^{(e)} = [\boldsymbol{R}_{Cm_b}] \cdot (\boldsymbol{P}^{(e)} - \boldsymbol{C}^{m_b}) + \boldsymbol{C}^{m_b} + \boldsymbol{T}^{m_b} \tag{3}$$

where $\boldsymbol{P^{(e)}}$ is an arbitrary point on the surface of an object–component m_b.

Because of the constraints imposed on the articulated object, motion parameters are not independent. They are, in general, related by a set of constraint motion equations that represent joints. Each constraint motion equation can be used to eliminate one motion parameter by writing this motion parameter in terms of the others, provided the constraint motion equations are linearly independent. In the case of two object–components m_a and m_b connected by a spherical joint, the constraint motion equations require that the global position of the spherical joint $\boldsymbol{J} = (J_x, J_y, J_z)^T$ defined by the set of coordinates of the object–component m_b moves to the same global position $\boldsymbol{J}'$ defined by the set of coordinates of object–component m_a.This condition gives three constraint motion equations that can be written as:

$$[\boldsymbol{R}_{Cm_b}] \cdot (\boldsymbol{J} - \boldsymbol{C}^{m_b}) + \boldsymbol{C}^{m_b} + \boldsymbol{T}^{m_b} = [\boldsymbol{R}_{Cm_a}] \cdot (\boldsymbol{J} - \boldsymbol{C}^{m_a}) + \boldsymbol{C}^{m_a} + \boldsymbol{T}^{m_a} \tag{4}$$

If $\boldsymbol{T}^{m_a}$ is selected as dependent motion parameters it can be expressed in terms of the other motion parameters as:

$$\boldsymbol{T}^{m_a} = [\boldsymbol{R}_{Cm_b}] \cdot (\boldsymbol{J} - \boldsymbol{C}^{m_b}) + \boldsymbol{C}^{m_b} + \boldsymbol{T}^{m_b} - [\boldsymbol{R}_{Cm_a}] \cdot (\boldsymbol{J} - \boldsymbol{C}^{m_a}) - \boldsymbol{C}^{m_a} \tag{5}$$

Combining Eq. (5) and the Eq. (1) the motion equation of an arbitrary object–component m_a, which considers motion constraints enforced by a spherical joint at position $\boldsymbol{J}$**,** is giving by:

$$\boldsymbol{P}'^{(i)} = [\boldsymbol{R}_{Cm_a}] \cdot (\boldsymbol{P}^{(i)} - \boldsymbol{J}) + [\boldsymbol{R}_{Cm_b}] \cdot (\boldsymbol{J} - \boldsymbol{C}^{m_b}) + \boldsymbol{C}^{m_b} + \boldsymbol{T}^{m_b} \tag{6}$$

Thus, for estimating the motion of an arbitrary object–component m_a (see Fig. 2) connected to an object–component m_b whose motion parameters $\boldsymbol{A}^{m_b} = (T_x^{m_b}, T_y^{m_b}, T_z^{m_b}, R_x^{m_b}, R_y^{m_b}, R_z^{m_b})^T$ are known , only the rotation parameters $\boldsymbol{R}^{m_a}$ have to be estimated because the translation $\boldsymbol{T}^{m_a}$ can be written in terms of the independent motion parameters $\boldsymbol{A}^{m_b}$, $\boldsymbol{R}^{m_a}$ and the position of the spherical joint $\boldsymbol{J}$ according to Eq. (5). For determining the rotation parameters $\boldsymbol{R}^{m_a}$, motion estimation supposes that differences between two consecutive images s_k and s_{k+1} are due to the object motion only and that the shape of the object–component is known. The motion estimation method minimizes the mean square luminance difference between a perspective projection of the object–component's luminance onto the image plane of a model camera s'_k and the corresponding luminance of the current image s_{k+1}. Therefore, a gradient method is applied which uses one set of observation points from the model object–component m_a. Each observation point $\boldsymbol{O}_k^{(j)} = (\boldsymbol{P}_k^{(j)}, \boldsymbol{g}^{(j)}, I^{(j)})$ at time instant k is located on the model object–component surface and is described by its position $\boldsymbol{P}_k^{(j)} = (P_x^{(j)}, P_y^{(j)}, P_z^{(j)})^T$, its luminance value $I^{(j)}$ and its linear gradients $\boldsymbol{g}^{(j)} = (g_x^{(j)}, g_y^{(j)})^T$. The luminance and linear gradients are taken from the same image from which the color parameters of the 3D model object were derived. The criterion for selecting observation points is a high spatial gradient. For each observation point $\boldsymbol{O}_k^{(j)}$, the luminance difference $\Delta I^{(j)}$ between s'_k and s_{k+1} is related to the unknown rotation parameters $\boldsymbol{R}^{m_a} = (R_x^{m_a}, R_y^{m_a}, R_z^{m_a})^T$ by the following linearized equation:

$$
\begin{aligned}
\Delta I - \Delta I^{m_b} = \; & - ([P_x g_x (P_y - J_y) + P_y g_y (P_y - J_y) + P_z g_y (P_z - J_z)] F/P_z^2 + (P_y - J_y) \Delta I / P_z) \;\; R_x^{m_a} \\
& + ([P_y g_y (P_x - J_x) + P_x g_x (P_x - J_x) + P_z g_x (P_z - J_z)] F/P_z^2 + (P_x - J_x) \Delta I / P_z) \;\; R_y^{m_a} \\
& - [g_x (P_y - J_y) - g_y (P_x - J_x)] \cdot F/P_z \cdot R_z^{m_a}
\end{aligned} \tag{7}
$$

with

$$
\begin{aligned}
\Delta I^{m_b} = \; & + F g_x / P_z \;\; T_x^{m_b} \\
& + F g_y / P_z \;\; T_y^{m_b} \\
& - [(P_x g_x + P_y g_y) F / P_z^2 + \Delta I / P_z] \;\; T_z^{m_b} \\
& - ([P_x g_x (J_y - C_y^{m_b}) + P_y g_y (J_y - C_y^{m_b}) + P_z g_y (J_z - C_z^{m_b})] F/P_z^2 + (J_y - C_y^{m_b}) \Delta I / P_z) \;\; R_x^{m_b} \\
& + ([P_y g_y (J_x - C_x^{m_b}) + P_x g_x (J_x - C_x^{m_b}) + P_z g_x (J_z - C_z^{m_b})] F/P_z^2 + (J_x - C_x^{m_b}) \Delta I / P_z) \;\; R_y^{m_b} \\
& - [g_x (J_y - C_y^{m_b}) - g_y (J_x - C_x^{m_b})] F / P_z \;\; R_z^{m_b}
\end{aligned}
$$

where F is the focal length of the model camera.

In order to improve the reliability, Eq. (7) has to be established for several hundred observation points and only observation points should be used for which the following inequation is satisfied:

$$|\Delta I - \Delta I^{m_b}| < \sigma_I \tag{8}$$

where σ_I is the standard deviation of all residuals $\Delta I - \Delta I^{m_b}$ according to Eq. (7). The residuum of this equation system is then minimized by an Gauß method for least squares error:

$$\sum_{O^{(j)}} (\Delta I^{(j)} - \Delta I^{m_b})^2 \rightarrow MIN \tag{9}$$

Due to the linearization, motion parameters have to be estimated iteratively. After every iteration, the model object–component m_a is moved according to Eq. (6) using the estimated rotation parameters. A new set of motion equations is then established, giving new rotation parameters updates. In case of convergency the rotation parameter updates approach zero during the iterations.

For typical articulated objects like human bodies and robot arms, which consist of many object–components, if all the constraint motion equations are written down, a very large system is obtained which is most likely impossible to solve. In order to avoid this difficulty an approach of decomposition and propagation of motion estimation is applied in this contribution. This approach decomposes first the object automatically into flexibly connected object–components[5][7]. Object–components are connected to each other by spherical joints. Then, it estimates the motion parameters of an arbitrary object–component m_r (root object–component) without considering motion constraints. Finally, the motion analysis is propagated to the rest of the object–components taking into account motion constraints between object–components using Eq. (5) and Eq. (7). For estimating the 3D motion parameters of the root object–component m_r the algorithm proposed in [9] is applied. Since the reliability of this algorithm depends on the size of the object component[6], the largest object–component of the articulated object is chosen as the root component. Propagation of motion analysis is explained using the articulated object of Fig. 1 as example. After 3D motion estimation of the root object–component m_r the motion parameters of the object–components m_1, m_4 and m_5 can be estimated applying Eq. (5) and Eq. (7). Knowing the motion parameters of the object–components m_1, m_4 and m_5 the motion parameters of the object–components m_2 and m_6 can be then calculated. Finally, using the motion parameters of the object–component m_2 the motion parameters of the object–component m_3 are estimated. This method can be generalized as follows: if the motion parameters of an arbitrary object–component m_r of an articulated object are known, the motion parameters of the rest of the object–components can be estimated propagating motion estimation from the object–component m_r to the farthest object–component using Eq. (5) and Eq. (7).

3. Experimental results

OBASC according to Ostermann[9] and OBASC with the developed algorithm for motion estimation of articulated objects (OBASC*) are applied to the test sequence ”Claire” (CIF, 10Hz). Fig.3 represents the stick model of Claire. The object was automatically decomposed into object–components using the approach for shape estimation of articulated 3D objects proposed by Martínez[5][7]. The position of the spherical joint is supposed to be the center of gravity of the connecting triangles between object–components. Color parameters of model failures were coded with a PSNR of 36 dB. In the experiment both coders were initialized using the first original image of the sequence. The average size of MF–regions obtained by OBASC and OBASC* is 2.9% and 2.5% of the image area, respectively. Using 1.2 bit/pel for coding of color parameters, the overall bit rate is reduced from 5300 bit/frame to 4800 bit/frame (see Fig. 4).

4. Conclusion

In order to reduce the total size of MF–regions the source model of ”moving articulated 3D objects” is used instead the source model of ”moving rigid 3D objects”. In this contribution an algorithm for 3D motion estimation of articulated objects which considers spatial constraints was described. The spatial constraints enforced by the connecting triangles between two object–components are modeled by one spherical joint. In order to reduce the complexity of the motion estimation algorithm, an approach of propagation of motion estimation is applied. The developed algorithm has been incorporated in the image analysis of OBASC[9]. For shape estimation of articulated objects the method proposed by Martínez[5][7] is used. The

”head and shoulders” videophone test sequence ”Claire” (CIF, 10Hz) has been used. Experimental results show that considering spatial constraints the average size of MF–regions decreases from 2.9% to 2.5%. The reduction was particularly large by those frames with strongly different motion between object–components (see Fig. 4). Maintaining the same picture quality measured by PSNR=36 dB in the image regions of model failures, this reduction of the average size of MF–regions leads to a reduction of the transmission rate from 53 kbit/s to 48 kbit/s.

5. References

[1] R. Holt, A. Netravali, T. Huang and R. Qian, ”Determining Articulated Motion from Perspective Views: A Decomposition Approach”, Proc. IEEE Workshop on Motion of Non–Rigid and Articulated Objects, Austin, Texas, Nov. 11–12, 1994, pp. 126–137.

[2] Robert Hsu, M. Konda, Y. Fujinuma, H. Harashima, ”Robust Tracking of Head and Arm for Model Based Image Coding”, Picture Coding Symposium, Sacramento, USA, 21.–23. September 1994, pp. 288–291.

[3] M. Kampmann, J. Ostermann, ”Automatic adaptation of a face model in a layered coder with an object–based analysis–synthesis layer and a knowledge–based layer”, accepted for publication in Signal Processing: *Image Communication*, Special Issue on Communicating Natural Scenes by Image Synthesis.

[4] R. Koch, ”Dynamic 3–D scene analysis through synthesis feedback control”, *IEEE T–PAMI*, Vol. 15, No. 6, pp. 556–568, June 1993.

[5] G. Martínez, ”Automatic analysis of flexibly connected rigid 3D objects for an OBASC”, Picture Coding Symposium, Sacramento, USA, 21.–23. September 1994, pp. 329–332.

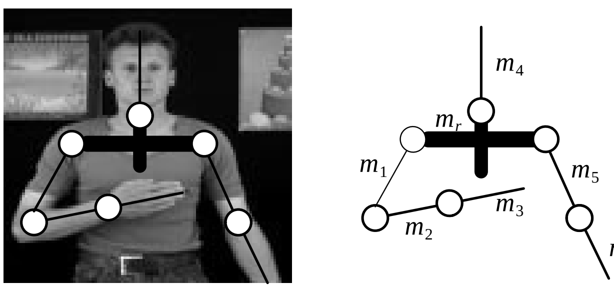
[6] G. Martínez, ”A robust algorithm for 3D–motion estimation of a small surface patch of a 3D object”, Workshop on Image Analysis and Synthesis in Image Coding, Berlin, 18–19 October 1994, pp. D1.1–D1.4.

[7] G. Martínez, ”Shape estimation of moving articulated 3D objects for Object–Based Analysis–Synthesis Coding (OBASC)”, accepted for publication in Signal Processing: *Image Communication*, Special Issue on Communicating Natural Scenes by Image Synthesis.

[8] H.G. Musmann, M. Hötter, J. Ostermann, ”Object–oriented analysis–synthesis coding of moving images”, *Signal Processing: Image Communication*, Vol. 3, No. 2, pp. 117–138, November 1989.

[9] J. Ostermann, ”Object–based analysis–synthesis coding based on the source model of moving rigid 3D objects”, *Image communication*, Vol. 6, No. 2, pp.143–161, May 1994.

[10] A. Shabana, *Dynamics of multibody systems*, John Wiley & Sons, Inc., USA, 1989, Cap. 1 and 2, pp. 1–116.



root object–component
object–component (link)
spherical joint

Fig. 1 Simplified stick model of an articulated object

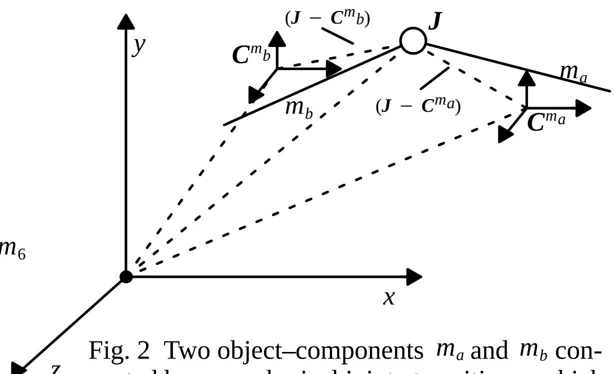


Fig. 2 Two object–components m_a and m_b connected by one spherical joint at position $\boldsymbol{J}$ which allows only relative angular rotations between the two object–components.

root object–component
object–component (link)
spherical joint

Fig. 3 Simplified stick model of the articulaed object Claire

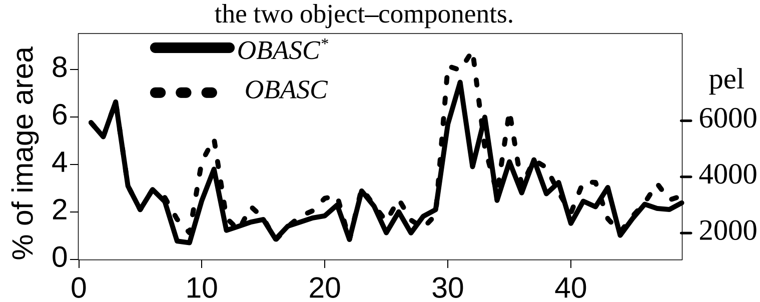


Fig. 4. Size of MF–regions obtained by *OBASC* and *OBASC** in percent of the image area for the test secuence Claire. The total area is 101376 pel. The average size of MF–regions is 2.90% of the image area for *OBASC* and 2.50% for *OBASC**