3D MOTION ESTIMATION OF HEAD AND SHOULDERS IN VIDEOPHONE SEQUENCES

Markus Kampmann

Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung
Universität Hannover, Appelstraße 9A, 30167 Hannover, F.R.Germany
email: kampmann@tnt.uni–hannover.de,WWW: http://www.tnt.uni–hannover.de/~kampmann

ABSTRACT

In this paper, an approach for 3D motion estimation of head and shoulders of persons in videophone sequences is presented. Since head and shoulders are linked together by the neck, constraints for the motion of head and shoulders exist which are exploited to improve the motion estimation. In this paper, the human neck joint is modelled by a spherical joint between head and shoulders and its 3D position is calculated. Then, rotation and translation of the shoulders are estimated and propagated to the head. Finally, a rotation of the head around the neck joint is estimated. The presented approach is applied to the videophone sequences *Akiyo* and *Claire*. Compared with an approach without using a neck joint, anatomically correct positions of head and shoulders are estimated.

1. INTRODUCTION

For coding of moving images at low bit rates, an object–based analysis–synthesis coder (OBASC) has been introduced [1]. In an OBASC, each real object is described by a model object. A model object is defined by three sets of parameters defining its motion, shape and color. In [2], the shape of a model object is represented by a 3D wireframe. The motion is defined by 6 parameters describing the translation and rotation of the model object in 3D space. The color parameters denote luminance and chrominance reflectance on the model object surface. Objects may be articulated, i.e. may consist of two or more flexibly connected 3D object components. Each object component has its own set of 3D motion, 3D shape and color parameters. All these parameters have to be estimated automatically by image analysis.

In the case of typical videophone sequences, the human body in the sequence can be considered as an articulated object and head and shoulders as object components. All three sets of model parameters have to be estimated for these object components. This contribution deals with estimating the 3D motion parameters of head and shoulders.

In [5], a hierarchical approach for 3D motion estimation of object components is proposed. No constraints for the spatial location of the object components are considered. However, for articulated objects like a human body constraints for the spatial location of the object components exist. These constraints can be exploited to improve the motion estimation. By using joints between object components, constraints for the relative motion between the object components are introduced. In [6][7], 2D joint positions in sequent images are estimated and a 3D joint position is calculated. In [8], the 3D joint position is directly estimated using motion parameters from preceding images. Using these joint positions, motion estimation of articulated objects is carried out [6][7][9]. In these algorithms, no a priori knowledge about the object components head and shoulders and the position of the connecting neck joint on a human body is exploited.

In this contribution, an algorithm for 3D motion estimation of head and shoulders in videophone sequences is presented which uses a priori knowledge about the image content of a videophone sequence. Here, the human neck joint is modelled by a spherical joint between head and shoulders. The 3D position of the neck joint is calculated using an automatically generated 3D wireframe of the person in the sequence [4] and knowledge about the position of the neck joint on a human body. This 3D wireframe of the person is splitted into the object components head and shoulders and the motion of head and shoulders is estimated. First, six motion parameters namely three rotation

and three translation parameters are estimated for the shoulders. Since head and shoulders are linked together, the motion of the shoulders is propagated to the head. Finally, three rotation parameters for a rotation of the head around the neck joint are estimated.

2. CALCULATION OF 3D NECK JOINT POSITION

The 3D position of the neck joint is calculated using an automatically generated 3D wireframe of the person [4] and knowledge about the position of the neck joint on a human body. For generation of the 3D wireframe of the person, the pre–defined 3D face model from Fig. 1 is adapted automatically to the person's face in the sequence [3].

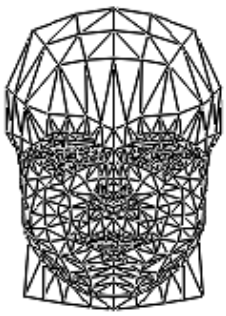


Fig. 1: Pre–defined 3D face model.

Thereby, eyes, mouth, nose, eyebrows and chin and cheek contours are automatically estimated from an image. Using these estimated facial features, the face model is adapted to the person's face. After adaptation, the 3D wireframe of the person is constructed using the person's silhouette and a priori knowledge about the 3D shape of an average human person. Finally, the adapted face model is incorporated into this wireframe (Fig. 2) [4]. Now, using knowledge about the position of the neck joint on a typical human body, the 3D neck joint position $\boldsymbol{J}$ is calculated (Fig. 2).

3. ESTIMATION OF HEAD AND SHOULDER MOTION

For 3D motion estimation, the 3D wireframe of the person is splitted into the object components head and shoulders. First, the 3D motion of the object component shoulders is estimated. A point $\boldsymbol{P}_{sh}$ on the surface of the object component shoulders is moved to its new positon $\boldsymbol{P}'_{sh}$ according to

$$\boldsymbol{P}'_{sh} = [\boldsymbol{R}_{sh}](\boldsymbol{P}_{sh} - \boldsymbol{C}_{sh}) + \boldsymbol{C}_{sh} + \boldsymbol{T}_{sh} \quad (3.1)$$

with the translation vector $\boldsymbol{T}_{sh}$, the center $\boldsymbol{C}_{sh} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{P}_C^{(i)}$ of the N control points $\boldsymbol{P}_C^{(i)}$of the object component shoulder and the rotation matrix $[\boldsymbol{R}_{sh}]$. $[\boldsymbol{R}_{sh}]$ is calculated from the rotation angles $\boldsymbol{R}_{sh} = (R_{sh,x}, R_{sh,y}, R_{sh,z})^T$. Equation (3.1) rotates the object component around its center $\boldsymbol{C}_{sh}$ and then moves it by the translation $\boldsymbol{T}_{sh}$. For motion estimation, the three rotation parameters $\boldsymbol{R}_{sh}$ and the three translation parameters $\boldsymbol{T}_{sh}$ have to be estimated. In order to estimate $\boldsymbol{R}_{sh}$ and $\boldsymbol{T}_{sh}$, the algorithm from [2] is applied.

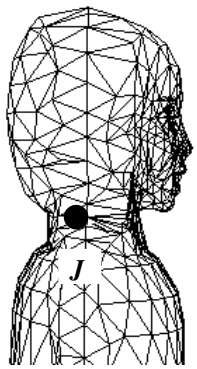


Fig. 2: 3D wireframe of the person with adapted 3D face model and 3D neck joint position $\boldsymbol{J}$ in a side view.

Since head and shoulders are linked together, the motion of the shoulders is propagated to the head. All vertices $\boldsymbol{P}_h$ of the object component head are moved with the estimated motion parameters $\boldsymbol{R}_{sh}$ and $\boldsymbol{T}_{sh}$ of the object component shoulder to its new position $\boldsymbol{P}'_h$ according to

$$\boldsymbol{P}'_h = [\boldsymbol{R}_{sh}](\boldsymbol{P}_h - \boldsymbol{C}_{sh}) + \boldsymbol{C}_{sh} + \boldsymbol{T}_{sh}. \quad (3.2)$$

Similarly, the neck joint $\boldsymbol{J}$ is moved to its new position $\boldsymbol{J}'$ according to

$$\boldsymbol{J}' = [\boldsymbol{R}_{sh}](\boldsymbol{J} - \boldsymbol{C}_{sh}) + \boldsymbol{C}_{sh} + \boldsymbol{T}_{sh}. \quad (3.3)$$

Due to the human anatomy, only a rotation of the head around the neck joint is possible. Therefore, the point $\boldsymbol{P}'_h$ on the surface of the head is moved to its new position $\boldsymbol{P}''_h$ according to

$$\boldsymbol{P}_h^{''} = [\boldsymbol{R}_h](\boldsymbol{P}_h^{'} - \boldsymbol{J}^{'}) + \boldsymbol{J}^{'} \quad (3.4)$$

with the rotation matrix $[\boldsymbol{R}_h]$. $[\boldsymbol{R}_h]$ is calculated from the rotation angles $\boldsymbol{R}_h = (R_{h,x}, R_{h,y}, R_{h,z})^T$. For motion estimation, the three rotation parameters $\boldsymbol{R}_h$ have to be estimated. Here, the same algorithm [2] as for the estimation of the motion parameters of the shoulders is used.

4. EXPERIMENTAL RESULTS

The described algorithm has been tested using the videophone sequences *Akiyo* and *Claire* (CIF, 10Hz). The face model from Fig. 1 is adapted automatically to the individual face, the 3D wireframe of the person is generated and splitted into the object components head and shoulder. Afterwards, the 3D motion of head and shoulders is estimated between two succeeding images of the videophone sequences. Here, the proposed 3D motion estimation algorithm is compared with the hierarchical 3D motion estimation algorithm in [5] which does not use a neck joint for motion estimation. Fig. 3 shows original images of the sequences *Akiyo* and *Claire*. Fig. 4 and Fig. 6 show results of the proposed motion estimation algorithm, Fig. 5 results of the motion estimation algorithm from [5]. Since no constraints for the spatial location of head and shoulders are used in [5], anatomically impossible positions of head and shoulders are estimated with the algorithm in [5] (Fig. 5). Using the proposed motion estimation algorithm, anatomically correct positions of head and shoulders are estimated (Fig. 6).

5. CONCLUSIONS

In this paper, a new approach for 3D motion estimation of head and shoulders in videophone sequences is presented. First, the 3D neck joint position of the person in the sequence is calculated using an automatically generated 3D wireframe of the person and knowledge about the position of the neck joint on a human body. Then, translation and rotation of the shoulders are estimated and propagated to the head. Finally, a rotation of the head around the neck joint is estimated. Compared with an algorithm without using a neck joint, anatomically correct positions of head and shoulders are estimated.

6. REFERENCES

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

[2] J. Ostermann, "Object–based analysis–synthesis Coding based on the source model of moving rigid 3D objects", *Signal Processing: Image Communications*, Vol. 6, pp. 143–161, May 1994.

[3] M. Kampmann, R. Farhoud, "Precise face model adaptation for semantic coding of videophone sequences", *Picture Coding Symposium* (PCS '97), Berlin, Germany, pp. 675–680, September 1997.

[4] M. Kampmann, "Automatic Generation of 3D wireframes of human persons for video coding", *7. Dortmunder Fernsehseminar*, Dortmund, Germany, pp. 105–108, September 1997.

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

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

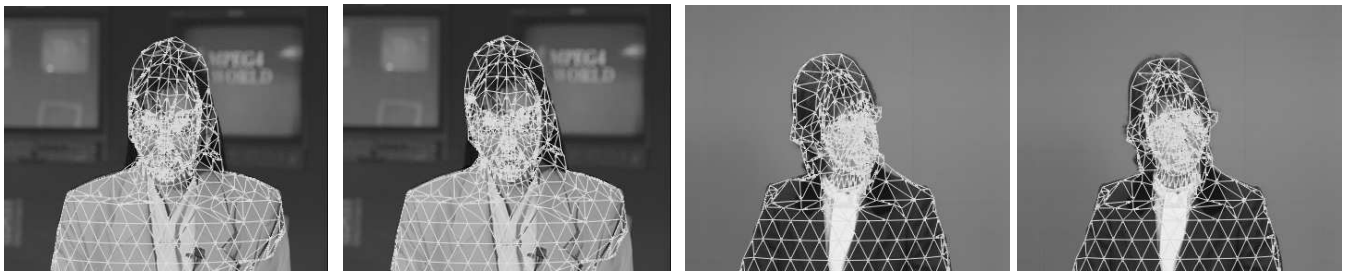
[7] J. Web. and J. Aggarwal, "Structure from motion of rigid and jointed objects", *Seventh International Joint Conference on Artifitial Intelligence*, Vancouver, S. 686–691, August 1981.

[8] G. Martínez, "Analyse–Synthese–Codierung basierend auf dem Modell bewegter dreidimensionaler, gegliederter Objekte", Dissertation, Universität Hannover, 1998.

[9] 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 (VLBV'95)*, Tokyo, Japan, G.1, November 1995.

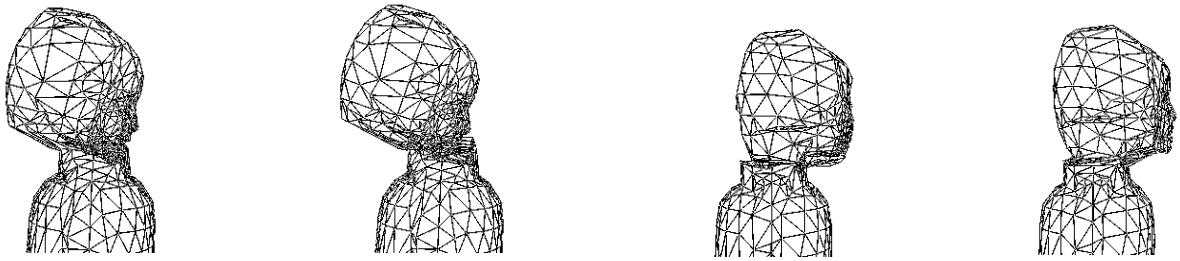
(a) (b) (c) (d)

Fig. 3: Original images (CIF,10Hz): (a) *Akiyo* frame 57, (b) *Akiyo* frame 91, (c) *Claire* frame 32, (d) *Claire* frame 48.



(a) (b) (c) (d)

Fig. 4: 3D wireframe over original image for the proposed algorithm: (a) *Akiyo* frame 57, (b) *Akiyo* frame 91, (c) *Claire* frame 32, (d) *Claire* frame 48.



(a) (b) (c) (d)

Fig. 5: 3D wireframe in a side view for the algorithm in [5]: (a) *Akiyo* frame 57, (b) *Akiyo* frame 91, (c) *Claire* frame 32, (d) *Claire* frame 48.

(a) (b) (c) (d)

Fig. 6: 3D wireframe in a side view for the proposed algorithm: (a) *Akiyo* frame 57, (b) *Akiyo* frame 91, (c) *Claire* frame 32, (d) *Claire* frame 48.