# Maximum-Likelihood Motion Estimation of Articulated Objects for Object-Based Analysis-Synthesis Coding

Geovanni Martinez
Visual Computing Lab
University of Houston 515 PGH
Houston, TEXAS 77204-3475
geovanni@uh.edu

## Abstract

*An algorithm for motion estimation of articulated objects from a single video camera is presented. The algorithm is used in an object-based analysis-synthesis coder. The coder is based on the source model of articulated objects. According to this source model, a real object consists of serveral rigid parts connected to each other by spherical joints. The algorithm was applied to a synthetic test sequence (CIF, 10 Hz) with very encouraging results. For example, for camera noise level of PSNR=40 dB, the new algorithm achieves a reduction of the error variance of the estimated translation parameters of up 40% and a reduction of up 35% for the rotation parameters.*

## 1. Introduction

An object-based analysis-synthesis coder for coding video at low data rates based on the source model of three-dimensional (3D) articulated objects is being investigated [6, 7]. According to this source model, an articulated object consists of several parts and each part is described by shape, motion and color parameters. The shape of a part is assumed to be rigid and described by a rigid wire-frame whose vertices represent the shape parameters of the part. The parts are connected to each other by spherical joints. The 3D motion of a part is described by six parameters: three rotation angles and one 3D translation vector. The color parameters of a part denote the luminance as well as the chrominance reflectance on the part's surface and they are defined by projecting a real image onto the part's wire-frame representation. For each part only the 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 cannot be reconstructed with sufficient image quality are called Model Failure (MF) objects. For each MF object its two-dimensional (2D) shape and its color parameters must be also coded and transmitted. In this paper, a Maximum-Likelihood algorithm for estimating the motion parameters of the different parts of an articulated object is presented.

For motion estimation of articulated objects, two methods have been proposed. The first method extracts and tracks features in consecutive frames and then determines the motion from these correspondences [2, 3, 4]. The second method estimates the motion from frame differences at observation points. Those motion parameters, which minimized the frame differences at observation points, are considered to be the estimates of the motion parameters [1, 5, 8, 9].

In this contribution, the motion parameters will be estimated by maximizing the conditional probability of the frame differences at observation points. The conditional probability will be a function of the motion parameters, the frame differences, and the covariance matrix of the luminance error at the observation points. The luminance error will be the result of both shape estimation errors and camera noise. The covariance matrix of the luminance error will be determined by modeling the shape estimation error and the camera noise by stationary zero-mean Gaussian stochastic processes. Additionally, a decomposition approach will be applied to reduce the dimensionality of the parameter space during the estimation [2, 3]. First, the translation and rotation parameters of the root part will be estimated. Then, only the rotation parameters of the rest of the parts are estimated beginng from the root part one after the other.

This paper is organized as follows. In Section 2 the Maximum-Likelihood motion estimation algorithm is described, while in Section 3 experimental results for synthetic image sequences are presented. A brief summary and the conclusions are presented in 4.

## 2. Maximum-Likelihood Motion Estimation Algorithm

Figure 1 depicts an articulated object consisting of two parts A and B connected by a spherical joint J ( at times

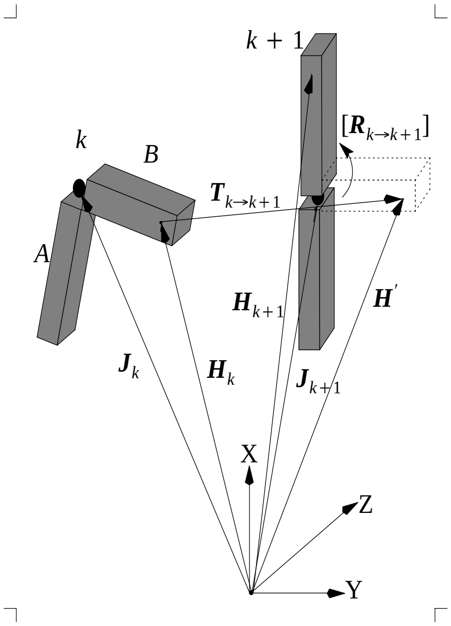

Figure 1: Depiction of an articulated object with two parts A and B connected by a spherical joint J and their positions at times k and k+1.

k and k+1). In the following, without loss of generality, we assume that the part A is the root part of the articulated object and that the shape, position and orientation of both parts as well as the position of the joint $\mathbf{J}_k = (J_X, J_Y, J_Z)^T$ at time k are known or previously estimated. Also, the motion of the part A from time k to k+1 is assumed to be known or previously estimated and it is described by the following set of six motion parameters $\mathbf{B}_{A,k\rightarrow k+1} = (T_{A,X}, T_{A,Y}, T_{A,Z}, R_{A,X}, R_{A,Y}, R_{A,Z})$. An arbitrary point $\mathbf{H}_{A,k}$ of the part A is moved to its new position $\mathbf{H}_{A,k+1}$ according to the following motion equation:

$$\mathbf{H}_{A,k+1} = [\mathbf{R}_{A,k\rightarrow k+1}]\,(\mathbf{H}_{A,k} - \mathbf{G}_{A,k}) + \mathbf{G}_{A,k} + \mathbf{T}_{A,k\rightarrow k+1}$$

where $\mathbf{T}_{A,k\rightarrow k+1} = (T_{A,X}, T_{A,Y}, T_{A,Z})^T$ is the translation vector, $[\mathbf{R}_{A,k\rightarrow k+1}]$ is the rotation matrix defined by the three rotation angles $\mathbf{R}_{A,k\rightarrow k+1} = (R_{A,X}, R_{A,Y}, R_{A,Z})$ and $\mathbf{G}_{A,k}$ is the rotation center of the part A. A similar motion equation can be written for part B. For convenience, the rotation center of the part B will be placed at joint position. In following an algorithm for estimation of the translation vector $\mathbf{T}_{k\rightarrow k+1} = (T_X, T_Y, T_Z)^T$ and the rotation angles $\mathbf{R}_{k\rightarrow k+1} = (R_X, R_Y, R_Z)$ of the part B will be described. Since the spherical joint imposes constraints on the relative motion of the connected parts, the translation vector of the part B can be calculated from the motion parameters of the part A and the joint position as follows:

$$\mathbf{T}_{k\rightarrow k+1} = [\mathbf{R}_{A,k\rightarrow k+1}](\mathbf{J}_k - \mathbf{G}_{A,k}) + \mathbf{G}_{A,k} + \\ + \mathbf{T}_{A,k\rightarrow k+1} - \mathbf{J}_k \qquad (1)$$

Since the translation vector can be calculated applying equation Eq. 1, only the rotation angles of the part B have to be estimated. For estimation of the rotation angles of the part B one set of observation points $\mathbf{W}_k^{(j)}$, $0 \ldots j \ldots \Omega - 1, \Omega > 3$ , $\mathbf{W}_k^{(j)} = (\mathbf{H}_k^{(j)}, \mathbf{g}^{(j)}, L^{(j)})$ are evaluated. Each observation point is located on the model part surface and is described by its position $\mathbf{H}_k^{(j)} = (H_X^{(j)}, H_Y^{(j)}, H_Z^{(j)})^T$, its luminance value $L^{(j)}$ and its linear gradients $\mathbf{g}^{(j)} = (g_x^{(j)}, g_y^{(j)})^T$. The luminance and linear gradients are taken from the same image from which the color parameters of the model part were derived. The criterion for selecting observation points is a high spatial linear luminance gradients. The perspective projection of an observation point $\mathbf{W}_k^{(j)}$ into the image plane is defined as $\mathbf{h}_k^{(j)} = (h_x, h_y)^T$ and the frame difference at this point is $fd(\mathbf{h}_k^{(j)})$.

The estimation of the rotation angles of the part B is carried out in two steps. In the first step the translation vector is calculated using Eq. 1. Then, the part B is motion compensated using the calculated translation vector. Each observation point $\mathbf{W}_k^{(j)}$ is moved from $\mathbf{H}_k^{(j)}$ to $\mathbf{H}^{(j)'}$ according to the following motion equation (see Fig. 1):

$$\mathbf{H}^{(j)'} = \mathbf{H}_k^{(j)} + \mathbf{T}_{k\rightarrow k+1} \qquad (2)$$

The joint is also moved according to Eq. 2 from $\mathbf{J}_k$ to $\mathbf{J}^{'} = \mathbf{J}_{k+1}$. Additionally, the projection of the observation point $\mathbf{W}_k^{(j)}$ moves from $\mathbf{h}_k^{(j)}$ to $\mathbf{h}^{(j)'}$ in the image plane. The displaced frame difference at point $\mathbf{h}^{(j)'}$ is defined as $dfd(\mathbf{h}^{(j)'}, \mathbf{T}_{k\rightarrow k+1})$. In the second step, the rotation angles of the part B are estimated. Each observation point $\mathbf{H}^{(j)'}$ is moved to $\mathbf{H}_{k+1}^{(j)}$ using the following motion equation:

$$\mathbf{H}_{k+1}^{(j)} = [\mathbf{R}_{k\rightarrow k+1}]\left(\mathbf{H}^{(j)'} - \mathbf{J}_{k+1}\right) + \mathbf{J}_{k+1}$$

where $[\mathbf{R}_{k\rightarrow k+1}]$ is the rotation matrix defined by the rotation angles $\mathbf{R}_{k\rightarrow k+1}$. For the estimation a Maximum-Likelihood method is used. Thus, those rotation angles $\widehat{\mathbf{R}}_{k\rightarrow k+1}$ which maximize the conditional probability of the displaced frame difference at the observation points $p_{\mathbf{DFD}/\widehat{\mathbf{R}}_{k\rightarrow k+1}}(\mathbf{DFD}/\widehat{\mathbf{R}}_{k\rightarrow k+1})$ are considered to be the estimates of the rotation angles:

$$\frac{\partial \ln\, p_{\mathbf{DFD}/\widehat{\mathbf{R}}_{k\rightarrow k+1}}(\mathbf{DFD}/\widehat{\mathbf{R}}_{k\rightarrow k+1})}{\partial \widehat{\mathbf{R}}_{k\rightarrow k+1}} = 0 \qquad (3)$$

where

$$p_{\mathbf{DFD}/\widehat{\mathbf{R}}_{k\to k+1}}(\mathbf{DFD}/\widehat{\mathbf{R}}_{k\to k+1}) = \frac{1}{\sqrt[2]{2\pi^{\Omega}|[\mathbf{U}]|}}\cdot$$

$$\cdot e^{(-\frac{1}{2}(\mathbf{DFD}-[\mathbf{O}]\widehat{\mathbf{R}}_{k\to k+1})^T[\mathbf{U}]^{-1}(\mathbf{DFD}-[\mathbf{O}]\widehat{\mathbf{R}}_{k\to k+1}))}$$

$$\mathbf{DFD} = (dfd(\mathbf{h}^{(\Omega-1)'}, \mathbf{T}_{k\to k+1}), dfd(\mathbf{h}^{(\Omega-2)'},$$

$$, \mathbf{T}_{k\to k+1}), ..., dfd(\mathbf{h}^{(0)'}, \mathbf{T}_{k\to k+1}))^T$$

$$[\mathbf{O}] = (\mathbf{o}^{(\Omega-1)T}, \mathbf{o}^{(\Omega-2)T}, ..., \mathbf{o}^{(0)T})^T$$

$$o^{(j)} = \begin{bmatrix} \theta_1^{(j)} \\ \theta_2^{(j)} \\ \theta_3^{(j)} \end{bmatrix}$$

$$\theta_1^{(j)} = [H_X^{(j)'} g_x^{(j)}(H_Y^{(j)'} - J_Y') + H_Y^{(j)'} g_y^{(j)}(H_Y^{(j)'} - J_Y') +$$

$$+ H_Z^{(j)'} g_y^{(j)}(H_Z^{(j)'} - J_Z')] f / H_Z^{(j)'2}$$

$$\theta_2^{(j)} = -[H_Y^{(j)'} g_y^{(j)}(H_X^{(j)'} - J_X') + H_X^{(j)'} g_x^{(j)}(H_X^{(j)'} - J_X') +$$

$$+ H_Z^{(j)'} g_x^{(j)}(H_Z^{(j)'} - J_Z')] f / H_Z^{(j)'2}$$

$$\theta_3^{(j)} = [g_x^{(j)}(H_Y^{(j)'} - J_Y') - g_y^{(j)}(H_X^{(j)'} - J_X')] F / H_Z^{(j)'}$$

$$E[\mathbf{V}\mathbf{V}^T] = [\mathbf{U}] =$$

$$= \begin{bmatrix} \sigma^2_{\Delta s^{(\Omega-1)}} & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2_{\Delta s^{(\Omega-2)}} & 0 & \cdots & 0 \\ 0 & 0 & \sigma^2_{\Delta s^{(\Omega-3)}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_{\Delta s^{(0)}} \end{bmatrix}$$

$$\mathbf{V} = (\Delta s^{(\Omega-1)}, \Delta s^{(\Omega-2)}, \Delta s^{(\Omega-3)}, \cdots, \Delta s^{(0)})^T$$

Eq. 3 is obtained assuming small rotation angles and using a perspective camera model with focal length f as well as the following linear luminance signal model at each observation point:

$$dfd(\mathbf{h}^{(j)'}, \mathbf{T}_{k\to k+1}) = -\mathbf{g}^{(j)T}(h_{k+1}^{(j)} - h^{(j)'}) + \Delta s^{(j)}$$

where $\Delta s^{(j)}$ is the luminance error at the observation point $\mathbf{W}^{(j)}$, which considers both the luminance error $\Delta s_f^{(j)}$ due to the shape error $\Delta\mathbf{H}^{(j)}$ of the model part and the luminance error $\Delta s_n^{(j)}$ due to camera noise $n^{(j)}$. The shape error $\Delta\mathbf{H}^{(j)}$ and the camera noise $n^{(j)}$ are considered to be statistically independent. As a first approach, the shape error $\Delta\mathbf{H}^{(j)}$ is modeled by three uncorrelated stationary zero-mean Gaussian stochastic processes $\Delta X^{(j)}, \Delta Y^{(j)}, \Delta Z^{(j)}$ with variances $\sigma_X^2, \sigma_Y^2, \sigma_Z^2$, respectively, describing the shape error at each observation point in the X, Y and Z directions. The covariance matrix of $\Delta\mathbf{H}^{(j)}$ can be represented as follow:

$$\left[\Delta\mathbf{H}^{(j)}\Delta\mathbf{H}^{(j)T}\right] = [\mathbf{C}_{\Delta\mathbf{H}^{(j)}}] = \begin{bmatrix} \sigma_X^2 & 0 & 0 \\ 0 & \sigma_Y^2 & 0 \\ 0 & 0 & \sigma_Z^2 \end{bmatrix} \quad (4)$$

The luminance errors $\Delta s_f^{(j)}$ and $\Delta s_n^{(j)}$ are also modeled by stationary zero-mean Gaussian stochastic processes with variances $\sigma^2_{\Delta s_f^{(j)}}$ and $\sigma^2_{\Delta s_n^{(j)}}$, respectively. The quantity $\sigma^2_{\Delta s_f^{(j)}}$ is computed in two steps. In the first step $\Delta\mathbf{H}^{(j)}$ is mapped to a vector $\Delta\mathbf{p}^{(j)}$ onto the image plane using the following linear transformation of a perspective camera model:

$$\Delta\mathbf{p}^{(j)} = \begin{bmatrix} \frac{f}{H_Z^{(j)}} & 0 & \frac{-fH_X^{(j)}}{H_Z^{(j)2}} \\ 0 & \frac{f}{H_Z^{(j)}} & \frac{-fH_Y^{(j)}}{H_Z^{(j)2}} \end{bmatrix} \Delta\mathbf{H}^{(j)} \quad (5)$$

$$= [\mathbf{K}_{\Delta\mathbf{H}^{(j)}}]\,\boldsymbol{\Delta}\mathbf{H}^{(j)}$$

The covariance matrix of $\Delta\mathbf{p}^{(j)}$ can be calculated from Eq. 4 and Eq. 5 as follow:

$$\left[\mathbf{C}_{\Delta\mathbf{p}^{(j)}}\right] = [\mathbf{K}_{\Delta\mathbf{H}^{(j)}}]\,[\mathbf{C}_{\Delta\mathbf{H}^{(j)}}]\,[\mathbf{K}_{\Delta\mathbf{H}^{(j)}}]^T \quad (6)$$

In the second step, $\Delta\mathbf{p}^{(j)}$ is mapped to $\Delta s_f^{(j)}$ using the following linear luminance signal model:

$$\Delta s_f^{(j)} = \mathbf{g}^{(j)T}\Delta\mathbf{p}^{(j)} \quad (7)$$

The variance of the luminance error due to the shape error $\sigma^2_{\Delta s_f^{(j)}}$ at the observation point $\mathbf{W}^{(j)}$ is finally calculated from Eq. 6 and Eq. 7 as follow:

$$\sigma^2_{\Delta s_f^{(j)}} = \mathbf{g}^{(j)T}\left[\mathbf{C}_{\Delta\mathbf{p}^{(j)}}\right]\mathbf{g}^{(j)}$$

$$= \frac{\sigma_Z^2}{H_Z^{(j)'2}}((g_x^{(j)}h_x^{(j)'} + g_y^{(j)}h_y^{(j)'})^2 + \frac{F^2}{\sigma_Z^2}(g_x^{(j)2}\sigma_X^2 +$$

$$+ g_y^{(j)2}\sigma_Y^2))$$

The luminance error $\Delta s^{(j)}$ is also modeled by a stationary zero-mean Gaussian stochastic process with variance:

$$\sigma^2_{\Delta s^{(j)}} = \sigma^2_{\Delta s_f^{(j)}} + \sigma^2_{\Delta s_n^{(j)}}$$

Solving Eq. 3 we obtain the Maximum-Likelihood estimates of the rotation angles of the part B as follows:

$$\widehat{\mathbf{R}}_{k\to k+1,ML} = ([\mathbf{O}]^T[\mathbf{U}]^{-1}[\mathbf{O}])^{-1}[\mathbf{O}]^T[\mathbf{U}]^{-1}\mathbf{DFD} \quad (8)$$

Eq. 8 is applied iteratively since Eq. 3 is highly non-linear. For estimation of the translation vector and rotation angles of the part A (the root of the articulated object) a Maximum-Likelihood estimator is also applied.

# 3. Experimental results

The proposed algorithm for motion estimation of articulated objects was applied to a synthetic test sequence (CIF, 10 Hz). This sequence depicts an articulated object consisting of two parts, which move differently from frame to frame. The parts are connected to each other by a spherical joint. Fig. 2 depicts the error variance of the estimated rotation parameters for different levels of camera noise. For camera noise of 40 dB measured by PSNR and estimating the motion parameters by minimizing the frame differences at the observation points the average of the translation parameters estimation error variance and the average of the rotation parameters estimation error variance is 0.7794 $pel^2$ and 0.3078 $degree^2$, respectively. At the same camera noise figure but applying the proposed Maximum-Likelihood method results 0.4687 $pel^2$ and 0.2013 $degree^2$, respectively.

# 4. Summary and Conclusions

In this contribution a Maximum-Likelihood method for motion estimation of an articulated object is presented. First the translation vector and the rotation angles of the root part are estimated. Then only the rotation angles of the rest of the parts are estimated biginning from the root part one after the other. At a camera noise figure of PSNR=40 dB the new method achieves a reduction of the estimation error variance of the translation parameters of up to 40% and a reduction of the estimation error variance of the rotation parameters of up to 35%.

# Acknowledgments

This research was supported by the German Academic Exchange Service (DAAD).

# References

[1] C. Bregler, J. Malik, "Tracking People with Twist and Exponential Maps", Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition",Santa Barbara, CA, pp. 8-15, 1998.

[2] D. M. Gavrila, L. S. Davis, "3-D model-based tracking of humans in action: a multi-view approach", Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, pp. 73-80, 1996.

[3] R. Holt, A. Netravali, T. Huang, R. Qian, "Determining Articulated Motion from Perspective Views: A Decomposition Approach", Proc. of the Workshop on Motion of Non-Rigid and Articulated Objects, Austin, Texas, pp.126-137, 1994.

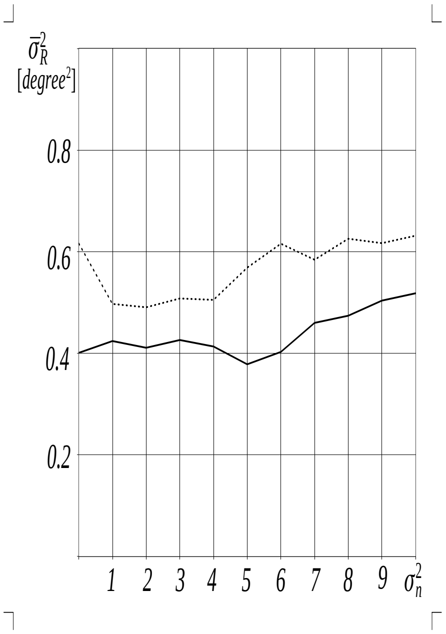

Figure 2: Error variance of the estimated rotation parameters for camera with different variances. Comparison of the results were obtained estimating the motion parameters by minimizing the frame differences at the observation points (dotted line) and by the proposed Maximum-Likelihood method (continuous line).

[4] I. A. Kakadiaris, D. Metaxas, "Model-Based Estimation of 3D Human Motion with Occlusion Based on Active Multi-Viewpoint Selection", Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, pp. 81-87, 1996.

[5] F. Kappei, C. E. Liedtke, "Modelling of a 3-D Scene Consisting of Moving Objects from a Sequence of Monocular TV Images", Proc. SPIE, Vol. 860, pp. 126, 1987.

[6] G. Martinez, "Analysis-Synthesis Coding Based on the Source Model of Articulated Three-Dimensional Objects", Proc. of the Picture Coding Symposium 99 (PCS'99), Portland, Oregon, pp. 213-216, 1999.

[7] H. G. Musmann, M. Hoetter, J. Ostermann, "Object-Oriented Analysis-Synthesis Coding of Moving Images", Signal Processing: Image Communication, Vol. 1, No. 2, pp. 117-138, 1989.

[8] J. Ostermann, "Object-Based Analysis-Synthesis Coding Based on the Source Model moving rigid 3D objects", Signal Processing: Image Communications, Vol. 6, No. 2, pp. 143-161, 1994.

[9] M. Yamamoto, K. Yagishita, "Human Motion Analysis Based on a Robot Arm Model", Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 664-665, 1991.