Joint Position Estimation
for Object–Based Analysis–Synthesis Coding

Geovanni Martínez*

Escuela de Ingeniería Eléctrica, Universidad de Costa Rica, 2060, San José, Costa Rica

ABSTRACT

An object–based analysis–synthesis coder for coding moving images at low data rates is investigated. 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. In this contribution a new algorithm for joint position estimation is presented. First the motion parameters of the connected object components are estimated at different times. For motion estimation a Maximum–Likelihood estimator is applied. Then the position of the joints is determined by evaluating the estimated motion parameters in the equations representing the constraints imposed by the joints on the relative motion of the connected object components. The algorithm was applied to a synthetic test sequence (CIF, 10 Hz). At a camera noise figure of 40 dB a joint position estimation average error of 0.94 pel can be achieved.

Keywords: object–based analysis–synthesis coding, video coding, video compression, articulated objects, shape estimation

1. INTRODUCTION

For coding of moving images at low data rates object–based analysis–synthesis coding is investigated [2][3][4][6][7][11][12][13][14] (see Fig. 1). An object based coder subdivides each image of an image sequence into uniformly moving objects and describes each object in terms of 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 exact meaning of the shape and motion parameters depends on the source model being applied (see Fig. 2). The parameters are automatically estimated by image analysis.

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 color parameters must be coded and transmitted. Since the transmission of the color parameters is expensive in terms of data rate, the total size of all the MF objects should be kept as small as possible.

In [4] a first complete implementation of an object–based analysis–synthesis coder is presented. The coder is based on the source model of flexible 2D objects (see Fig. 2.a). By this source model the shape of an objet is described by the 2D object silhouette. The object silhouette is estimated by a change detection between the current image and the last transmitted image. The motion of an object is described by a displacement vector field.

In order to reduce the total size of MF objects three–dimensional (3D) source models are used instead of 2D source models [11]. Until now three different 3D source models have been proposed [5]: the source model of rigid 3D objects [13] (see Fig. 2.b), the source model of flexible 3D objects [14] (see Fig. 2.c) and the source model of articulated 3D objects [6][7](see Fig. 2.d).

By the source model of rigid 3D objects the shape of an object is described by a rigid wire–frame. The vertices represent the shape parameters. The wire–frame is completely described by the object silhouette, i.e. there is an algorithm which computes a generalized cylinder from the silhouette [13]. The object silhouette is also 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. The color parameters are defined by projecting a real image onto the wire–frame.

last transmitted image

current image

image analysis

MF Objects
(Model Failures)

shape

motion

color
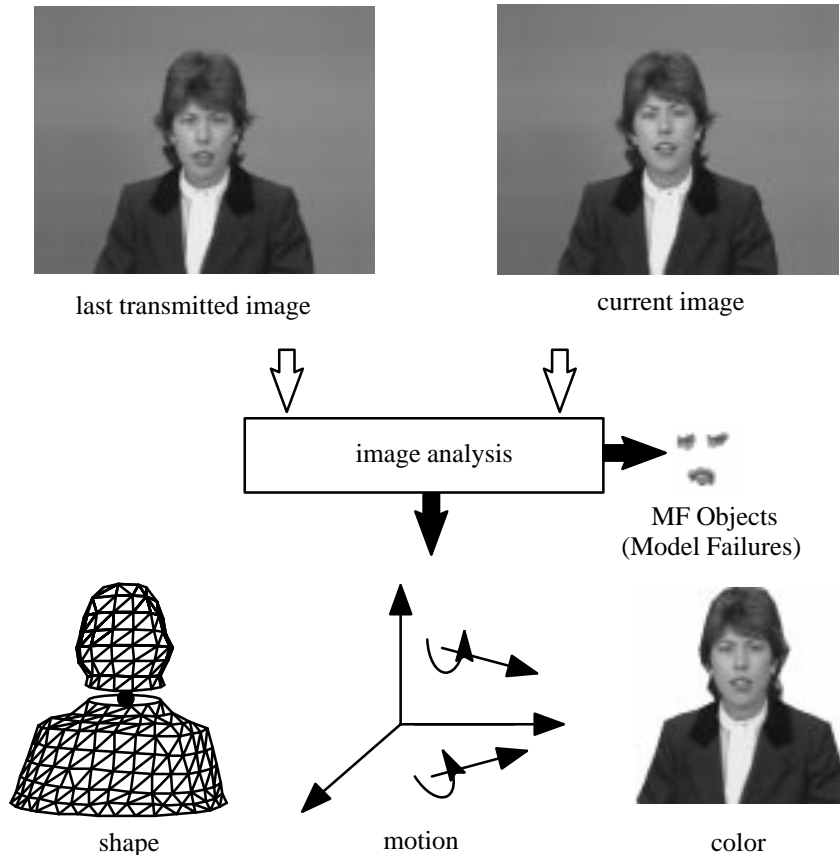
Fig. 1 Object–based analysis–synthesis coding based on the source model of
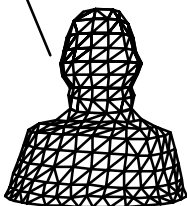articulated three–dimensional objects.

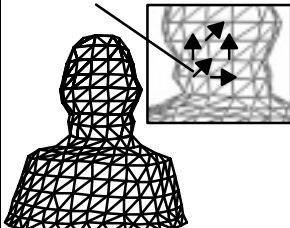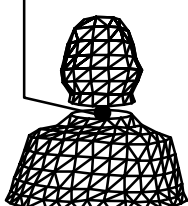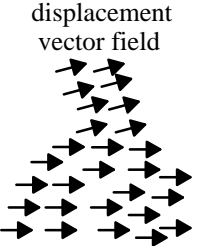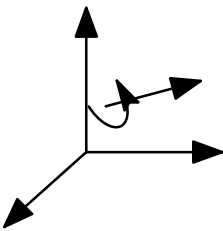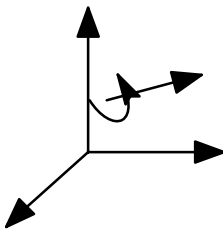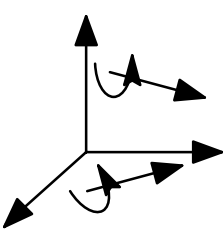While the source model of rigid objects describes only the shape of rigid objects, the source model of moving flexible objects allows also a local deformation of the objects tangential to the object surface. This deformation is carried out by shifting vertices of the wire–frame and is described by shift vectors. The other parameters are defined like they were defined by the source model of rigid 3D objects.

According to the source model of articulated 3D objects an articulated object consists of several connected object components. 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. The object components are connected to each other by spherical joints. A spherical joint is considered 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. The color parameters of an object component are also defined by projecting a real image onto its wire–frame. In this contribution a new algorithm for joint position estimation of articulated objects is proposed.

In order to reduce the data rate for transmision of the shape parameters the shape of the object components is estimated by subdividing a first rigid wire–frame into object components [6][7][8][9] (see Fig. 3). The rigid wire–frame is generated like it is generated in the model of rigid 3D objects. For the subdivision of the wire–frame the parameters of the 3D motion of each triangle of the wire–frame are estimated. For the motion estimation a Maximum–Likelihood estimator is applied[7]. It models statistically the measured values and also detects outliers in them. Then neighboring triangles which exhibit similar 3D motion parameters are clustered into object components. If an object component is detected the corresponding vertices of the rigid wire–frame are assigned to the new object component. The wire–frames of two connected object components remain connected to each other by those triangles having vertices belonging to different object components. Due to these connecting triangles

object–components are flexibly connected. The center of gravity of the connecting triangles is supposed to be the joint position [10]. In this contribution the joint position is estimated from the motion of the object components at different times.

This paper is organized as follows. In Section 2, the algorithm for joint position estimation is described. In Section 3, experimental results for synthetic image sequences are given. The conclusions are presented in Section 4.

	flexible 2D objects	rigid 3D objects	flexible 3D objects	articulated 3D objects
shape		wire–frame	shift vectors	spherical joint
motion	displacement vector field			
color				
	a	b	c	d

Fig. 2 Parameters of the source model of flexible 2D objects (a), rigid 3D objects (b), flexible 3D objects (c) and articulated 3D objects (d).
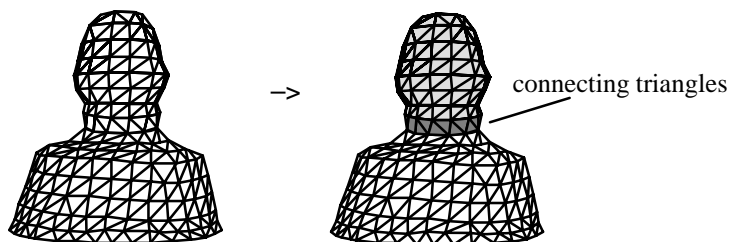
Fig. 3 The shape of the object components is estimated by subdividing a first rigid wire–frame into object components. Until now the center of gravity of the connecting triangles is supposed to be the joint position.

2. JOINT POSITION ESTIMATION

In this section a new algorithm for joint position estimation is presented. Before joint position estimation the shape of the object components is estimated using the algorithm described in [7][8][9].

Let a and b be two object components connected to each other by a spherical joint J. The 3D motion of the object component a is described by a set of six motion parameters $\boldsymbol{B}^{(a)} = (T_x^{(a)}, T_y^{(a)}, T_z^{(a)}, R_x^{(a)}, R_y^{(a)}, R_z^{(a)})$ defining its translation and rotation in the 3D space. An arbitrary point $\boldsymbol{H}_a$ of the object component a is moved to its new position $\boldsymbol{H}_a'$ according to the following motion equation:

$$\boldsymbol{H}_a' = [\boldsymbol{R}^{(a)}] \cdot (\boldsymbol{H}_a - \boldsymbol{G}^{(a)}) + \boldsymbol{G}^{(a)} + \boldsymbol{T}^{(a)} \tag{1}$$

where $\boldsymbol{T}^{(a)} = (T_x^{(a)}, T_y^{(a)}, T_z^{(a)})^T$ is the translation vector, $[\boldsymbol{R}^{(a)}]$ the rotation matrix defined by the three rotation angles $R_x^{(a)}, R_y^{(a)}, R_z^{(a)}$ and $\boldsymbol{G}^{(a)}$ the center of gravity of the object component a. The center of gravity $\boldsymbol{G}^{(a)}$ is calculated as:

$$\boldsymbol{G}^{(a)} = (G_x^{(a)}, G_y^{(a)}, G_z^{(a)})^T = (1/N_a) \cdot \sum_{i=1}^{N_a} \boldsymbol{H}_a^{(i)} \tag{2}$$

where N_a is the number of vertices $\boldsymbol{H}_a^{(i)}$, $i = 1, ..., N_a$, of the object component a.

Similar equations can be written for the object component b:

$$\boldsymbol{H}_b' = [\boldsymbol{R}^{(b)}] \cdot (\boldsymbol{H}_b - \boldsymbol{G}^{(b)}) + \boldsymbol{G}^{(b)} + \boldsymbol{T}^{(b)} \tag{3}$$

$$\boldsymbol{G}^{(b)} = (G_x^{(b)}, G_y^{(b)}, G_z^{(b)})^T = (1/N_b) \cdot \sum_{i=1}^{N_b} \boldsymbol{H}_b^{(i)} \tag{4}$$

The spherical joint J imposes constraints on the relative motion of the connected object components. Due to these constraints the joint can be moved from $\boldsymbol{J}$ to its new position $\boldsymbol{J}'$ applying both Eq. (1) and Eq. (3):

$$\boldsymbol{J}' = [\boldsymbol{R}^{(a)}] \cdot (\boldsymbol{J} - \boldsymbol{G}^{(a)}) + \boldsymbol{G}^{(a)} + \boldsymbol{T}^{(a)} \tag{5}$$

$$\boldsymbol{J}' = [\boldsymbol{R}^{(b)}] \cdot (\boldsymbol{J} - \boldsymbol{G}^{(b)}) + \boldsymbol{G}^{(b)} + \boldsymbol{T}^{(b)} \tag{6}$$

Combining Eq. (5) and Eq. (6) the following system of linear equations results. It represents the constraints imposed by the joint on the relative motion of the connected object components:

$$[\boldsymbol{R}^{(a)}] \cdot (\boldsymbol{J} - \boldsymbol{G}^{(a)}) + \boldsymbol{G}^{(a)} + \boldsymbol{T}^{(a)} = [\boldsymbol{R}^{(b)}] \cdot (\boldsymbol{J} - \boldsymbol{G}^{(b)}) + \boldsymbol{G}^{(b)} + \boldsymbol{T}^{(b)}$$

$$[\boldsymbol{R}^{(b)}] \cdot \boldsymbol{G}^{(b)} - [\boldsymbol{R}^{(a)}] \cdot \boldsymbol{G}^{(a)} + \boldsymbol{G}^{(a)} - \boldsymbol{G}^{(b)} + \boldsymbol{T}^{(a)} - \boldsymbol{T}^{(b)} = [[\boldsymbol{R}^{(b)}] - [\boldsymbol{R}^{(a)}]] \cdot \boldsymbol{J} \tag{7}$$

For joint position estimation two sets of motion parameters for each object component are first estimated (see Fig. 4). $\boldsymbol{B}_{k-2 \to k-1}^{(a)}$ and $\boldsymbol{B}_{k-2 \to k}^{(a)}$ represent the first and the second estimated set of motion parameters of the object component a, respectively. The first set describes its motion from time k–2 to time k–1. The second one describes its motion from time k–2 to time k. $\boldsymbol{B}_{k-2 \to k-1}^{(b)}$

and $\boldsymbol{B}^{(b)}_{k-2\to k}$ correspond to the object component b. For motion estimation a Maximum–Likelihood estimator is applied [7]. By evaluating the estimated sets of motion parameters of the object components a and b in the Eq. (7) at time k–2, the following system of linear equations for the position of the spherical joint $\boldsymbol{J}_{k-2}$ at time k–2 can be established [1]:

$$
\begin{bmatrix} \boldsymbol{E}_{k-2\to k-1} \\ \boldsymbol{E}_{k-2\to k} \end{bmatrix} = \begin{bmatrix} [\boldsymbol{R}^{(b)}_{k-2\to k-1}] - [\boldsymbol{R}^{(a)}_{k-2\to k-1}] \\ [\boldsymbol{R}^{(b)}_{k-2\to k}] - [\boldsymbol{R}^{(a)}_{k-2\to k}] \end{bmatrix} \cdot \boldsymbol{J}_{k-2} + \boldsymbol{r}
$$

$$
\boldsymbol{E} = [\boldsymbol{O}] \cdot \boldsymbol{J}_{k-2} + \boldsymbol{r} \tag{8}
$$

where

$$
\boldsymbol{E}_{k-2\to k-1} = [\boldsymbol{R}^{(b)}_{k-2\to k-1}] \cdot \boldsymbol{G}^{(b)}_{k-2} - [\boldsymbol{R}^{(a)}_{k-2\to k-1}] \cdot \boldsymbol{G}^{(a)}_{k-2} + \boldsymbol{G}^{(a)}_{k-2} - \boldsymbol{G}^{(b)}_{k-2} + \boldsymbol{T}^{(a)}_{k-2\to k-1} - \boldsymbol{T}^{(b)}_{k-2\to k-1}
$$

$$
\boldsymbol{E}_{k-2\to k} = [\boldsymbol{R}^{(b)}_{k-2\to k}] \cdot \boldsymbol{G}^{(b)}_{k-2} - [\boldsymbol{R}^{(a)}_{k-2\to k}] \cdot \boldsymbol{G}^{(a)}_{k-2} + \boldsymbol{G}^{(a)}_{k-2} - \boldsymbol{G}^{(b)}_{k-2} + \boldsymbol{T}^{(a)}_{k-2\to k} - \boldsymbol{T}^{(b)}_{k-2\to k}
$$

$\boldsymbol{G}^{(a)}_{k-2}$ and $\boldsymbol{G}^{(b)}_{k-2}$ represent the center of gravity of the object component a and b at time k–2, respectively and $\boldsymbol{r} = (r_0, r_1, ..., r_5)^T$ is the residuum of the system of linear equations.
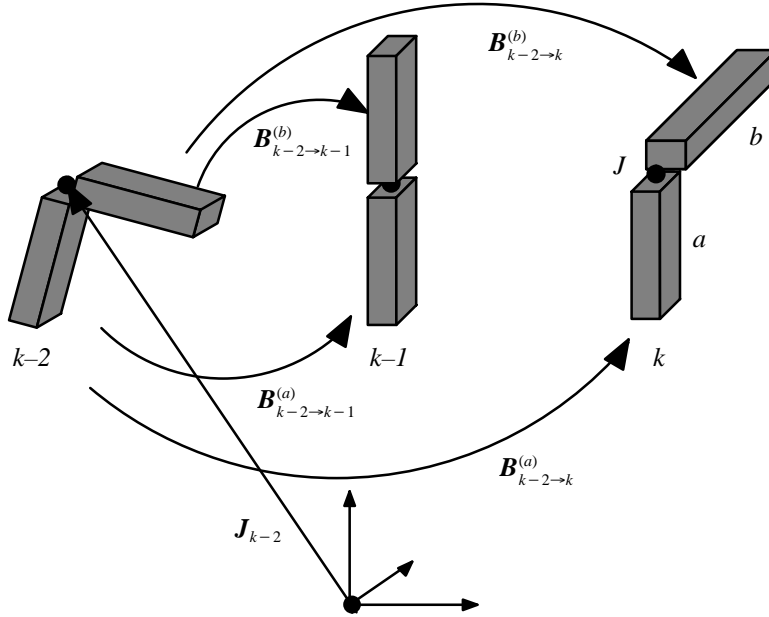
Fig. 4 Two object components a and b connected to each other by one spherical joint J and their position at time k, k–1 and k–2. $\boldsymbol{B}^{(a)}_{k-2\to k-1}$ represents the set of motion parameters of the object component a, which decribes its motion from time k–2 to k–1. $\boldsymbol{B}^{(a)}_{k-2\to k}$ describes its motion from time k–2 to k. $\boldsymbol{B}^{(b)}_{k-2\to k-1}$ and $\boldsymbol{B}^{(b)}_{k-2\to k}$ correspond to the object component b. $\boldsymbol{J}_{k-2}$ represents the joint position at time k–2 referred to the world coordinate system.

The system (8) is solved using linear regression:

$$\hat{\boldsymbol{J}}_{k-2} = ([\boldsymbol{O}]^T \cdot [\boldsymbol{O}])^{-1} \cdot [\boldsymbol{O}]^T \cdot \boldsymbol{E} \qquad (9)$$

The joint position at time k is calculated according to Eq. (1) as:

$$\hat{\boldsymbol{J}}_k = [\boldsymbol{R}^{(a)}_{k-2 \to k}] \cdot (\hat{\boldsymbol{J}}_{k-2} - \boldsymbol{G}^{(a)}_{k-2}) + \boldsymbol{G}^{(a)}_{k-2} + \boldsymbol{T}^{(a)}_{k-2 \to k} \qquad (10)$$

Motion estimation errors decrease the reliability of joint position estimation. In order to improve the reliability, $\eta > 2$ sets of motion parameters for each object component are evaluated:

$$\begin{pmatrix} \boldsymbol{E}_{k-\eta \to k-\eta+1} \\ \boldsymbol{E}_{k-\eta \to k-\eta+2} \\ \boldsymbol{E}_{k-\eta \to k-\eta+3} \\ \vdots \\ \boldsymbol{E}_{k-\eta \to k} \end{pmatrix} = \begin{bmatrix} [\boldsymbol{R}^{(b)}_{k-\eta \to k-\eta+1}] - [\boldsymbol{R}^{(a)}_{k-\eta \to k-\eta+1}] \\ [\boldsymbol{R}^{(b)}_{k-\eta \to k-\eta+2}] - [\boldsymbol{R}^{(a)}_{k-\eta \to k-\eta+2}] \\ [\boldsymbol{R}^{(b)}_{k-\eta \to k-\eta+3}] - [\boldsymbol{R}^{(a)}_{k-\eta \to k-\eta+3}] \\ \vdots \\ [\boldsymbol{R}^{(b)}_{k-\eta \to k}] - [\boldsymbol{R}^{(a)}_{k-\eta \to k}] \end{bmatrix} \cdot \boldsymbol{J}_{k-\eta} + \boldsymbol{r}$$

$$\boldsymbol{E} = [\boldsymbol{O}] \cdot \boldsymbol{J}_{k-\eta} + \boldsymbol{r} \qquad (11)$$

where $\boldsymbol{r} = (r_0, r_1, ..., r_{3\eta-1})^T$ and $\boldsymbol{J}_{k-\eta}$ is the joint position at time $k - \eta$. The system of linear equations (11) is solved using linear regression:

$$\hat{\boldsymbol{J}}_{k-\eta} = ([\boldsymbol{O}]^T \cdot [\boldsymbol{O}])^{-1} \cdot [\boldsymbol{O}]^T \cdot \boldsymbol{E} \qquad (12)$$

and the joint position at time k is calculated according to Eq. (1) as:

$$\hat{\boldsymbol{J}}_k = [\boldsymbol{R}^{(a)}_{k-\eta \to k}] \cdot (\hat{\boldsymbol{J}}_{k-\eta} - \boldsymbol{G}^{(a)}_{k-\eta}) + \boldsymbol{G}^{(a)}_{k-\eta} + \boldsymbol{T}^{(a)}_{k-\eta \to k} \qquad (13)$$

3. EXPERIMENTAL RESULTS

The proposed algorithm for joint position estimation was applied to a synthetic test sequence (CIF, 10 Hz). This sequence shows an articulated object consisting of two object components, which move differently from frame to frame. The object components are connected to each other by a spherical joint. Fig. 5 shows the joint position estimation average error at different η values and camera noise variances. At a camera noise figure of 40 dB measured by PSNR (Peak Signal to Noise Ratio) and $\eta = 5$, the joint position estimation average error was 0.94 pel.

4. CONCLUSIONS

In this contribution an algorithm for joint position estimation is presented. At least two sets of motion parameters for each object component are first estimated. Then the joint position is determined evaluating the estimated set of motion parameters in the equations representing the constraints imposed by the joint on the relative motion of the connected object components. The joint position estimation average error was 0.94 pel when evaluating 5 sets of motion parameters for each object component.

ACKNOWLEDGMENTS
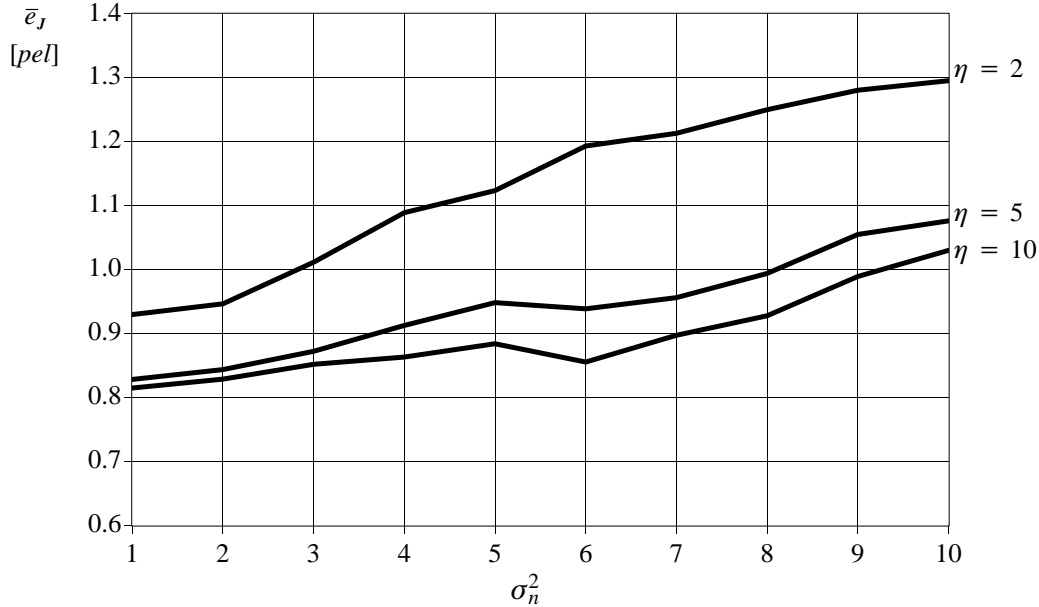
Fig. 5 Joint position estimation average error $\bar{e}_J$ in pel at different camera noise variances. η represents the number of sets of motion parameters for each object component used for joint position estimation.

REFERENCES

[1]   R. Farhoud, *Verfahren zur Schätzung der dreidimensionalen Position der Gelenke gegliederter Objekte für die Bewegtbildcodierung*, Studienarbeit, University of Hanover, Germany, 1996.

[2]   M. Hötter, "Optimization and Efficiency of an Object–Oriented Analysis–Synthesis Coder", IEEE Transactions on Circuits and Systems for Video Technology, **Vol. 4**, No. 2, pp. 181–194, 1994.

[3]   M. Hötter, *Objektorientierte Analyse–Synthese–Codierung basierend auf dem Modell bewegter, zweidimensionaler Objekte*, PhD Thesis, University of Hanover, Fortschritt–Berichte VDI–Verlag, Reihe 10, No. 217, Düsseldorf, Germany, 1992.

[4]   M. Hötter, "Object–Oriented Analysis–Synthesis Coding Based on Moving Two–Dimensional Objects", Signal Processing: Image Communication, **Vol. 2**, No. 4, pp. 409–428, 1990.

[5]   G. Martínez, "Modelos Tridimensionales Utilizados para la Codificación de Análisis–Síntesis Basada en Objetos", *International Conference on Electronics Communications and Computers (CONIELECOMP'2000),* pp. 47–50, Cholula, Puebla, Mexico, 2000.

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

[7]   G. Martínez, *Analyse–Synthese–Codierung basierend auf dem Modell bewegter, dreidimensionaler gegliederter Objekte*, Ph.D Thesis, University of Hanover, Germany, 1998.

[8]   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, 1997.

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

[10]   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 95 (VLBV'95)*, No. G–1, Tokyo, Japan, 1995.

[11]   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, 1989.

[12]   J. Ostermann, *Analyse–Synthese–Codierung basierend auf dem Modell bewegter, dreidimensionaler Objekte*, PhD Thesis, University of Hanover, Fortschritt–Berichte VDI–Verlag, Reihe 10, No. 391, Düsseldorf, Germany, 1995.

[13]   J. Ostermann, "Object–Based Analysis–Synthesis Coding Based on the Source Model of Moving Rigid 3D Objects", Signal Processing: Image Communication, **Vol. 6**, No. 2, pp. 143–161, 1994.

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