PARAMETRIC DESCRIPTIONS AND ESTIMATION, A SYNERGETIC APPROACH TO RESOLVING SHAPE FROM SHADING AND MOTION

M.J. Korsten and Z. Houkes

University of Twente, The Netherlands

INTRODUCTION

This paper presents a method combining shape, shading and motion models in order to obtain estimations of 3D shape and motion parameters directly from image grey values. The problem is considered as an application of optimal parameter estimation theory, according to Liebelt (8). This theory has been applied previously, where the emphasis was laid on time–delay, Burkhardt (2), and motion estimation, Diehl (3), Houkes (5), Stuller (9). It is applied here to provide an environment in which somewhat more complicated models can be designed with relative ease and to indicate how the behaviour of the parameters can be investigated. A shading model is added, offering explicit prediction of image grey values. We consider sequences of images, although in a similar way a stereo configuration could have been incorporated. The resulting non–linear estimation problem is linearized around a last parameter guess (8), so that a linear estimator can be applied to compute a new estimate. The various stages of the modeling process are separated by introducing several coordinate systems. Coordinate transformations will show the object from other points of view, and perform an orthographic projection of the 3D scene into the 2D image plane. The explicit grey value prediction yields a template, having a definite extent in the image. This method requires no gradient images, as in the case of estimating shape from motion (6) or stereo (5). The gradients can be computed analytically. To demonstrate the usefulness and the flexibility of our method, we consider a solid cylinder, irradiated with X–rays. The image is a shadow image originating from the absorption of radiation by the cylinder.

PARAMETER ESTIMATION

Linearization

Estimation theory is concerned with the optimal estimation of a vector of parameters $\boldsymbol{\alpha}$ from a vector of measurements $\boldsymbol{\theta}$, related by

$$\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{\alpha}) \tag{1}$$

Linearization of relation (1) offers a systematic way to obtain a solution, independent of the specific form of (1), so that afterwards a model can be put in. In accordance with Liebelt (8) the following quantities are introduced:

$$\delta\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{\alpha}) - \boldsymbol{\theta}(\hat{\boldsymbol{\alpha}})$$
$$\delta\boldsymbol{\alpha} = \boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}$$
$$\mathrm{B} = \left.\frac{\partial\boldsymbol{\theta}(\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}}\right|_{\boldsymbol{\alpha}=\hat{\boldsymbol{\alpha}}} \tag{2}$$

Here $\hat{\boldsymbol{\alpha}}$ is the last estimate of the parameter vector $\boldsymbol{\alpha}$, and $\boldsymbol{\theta}(\hat{\boldsymbol{\alpha}})$ is the 1st order approximation of $\hat{\boldsymbol{\theta}}(\boldsymbol{\alpha})$. The matrix B is the Jacobian of the measurements $\boldsymbol{\theta}$ with respect to the parameters. With these definitions from (1) a linearized matrix equation

$$\delta\boldsymbol{\theta} = \mathrm{B}\delta\boldsymbol{\alpha} + \underline{\mathbf{n}} \tag{3}$$

is obtained, where $\underline{\mathbf{n}}$ is a noisy term taking into account additive measurement noise and the fact, that the higher order terms in the Taylor series expansion are ignored. Equation (3) is the linear form for the observations in a linear parameter estimation problem (8).

Optimal estimation of the parameter vector. The least squares estimator for $\delta\hat{\boldsymbol{\alpha}}$ is:

$$\delta\hat{\boldsymbol{\alpha}} = (\mathrm{B}^{\mathrm{T}}\mathrm{B})^{-1}\mathrm{B}^{\mathrm{T}}\delta\boldsymbol{\theta} \tag{4}$$

and Liebelt (8) discusses the conditions under which this estimator is unbiased. We assume that these conditions are met. An implementation can be performed according to the scheme of fig. 1. Equation (4) is used to update the parameter vector which is used again in the models to compute a new prediction of the image grey values.

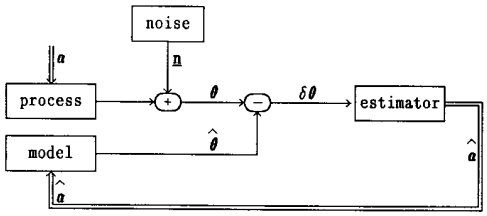

Figure 1 Scheme for the linearized estimator

Estimation theory and shape from shading

The shape from shading problem fits the framework of parameter estimation naturally, if the image grey values are chosen as the measurements from which the body parameters have to be estimated. However, the grey values depend also on properties of the camera system and the radiation source, irradiating the object, which will yield additional parameters. Some of these parameters may be a priori known, others have to be estimated together with the body parameters. The models, required to obtain a prediction of the image grey values, are discussed in detail in the next section. An implementation of an estimation algorithm can be deduced directly from fig. 1, where in this case the measurement vector $\boldsymbol{\theta}$ is composed of a set of grey values belonging to different grey positions $\mathbf{V}_i$:

$$\boldsymbol{\theta}(\boldsymbol{\alpha}) = [\mathrm{I}(\mathbf{V}_1;\boldsymbol{\alpha}),\mathrm{I}(\mathbf{V}_2;\boldsymbol{\alpha}),.....]^{\mathrm{T}} \tag{5}$$

Processing sequences of images. To estimate shape from image sequences, the model in fig. 1 is extended with a body motion model. This model, used in fig. 2, produces the predicted images $\hat{\boldsymbol{\theta}}_1$ at time t_1,and $\hat{\boldsymbol{\theta}}_2$ at time $t_2=t_1+\Delta t$. The differences $\delta\boldsymbol{\theta}_1$ and $\delta\boldsymbol{\theta}_2$ of these images with the real world images, are put together in a new

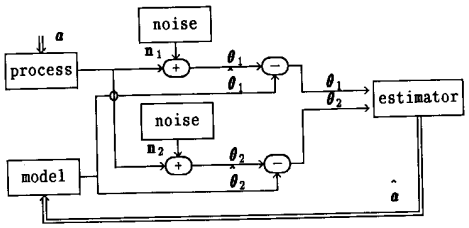

Figure 2 shape and motion estimation

measurement vector $\delta\theta$, which again is used in the estimator (4) to perform estimates of the parameter vector, extended with the motion parameters.

SHAPE, SHADING AND MOTION MODELS

In this section we address the design of the various models used. To obtain information from the image(s), a priori knowledge about the scene is necessary, which is brought into the algorithm with the help of parametric models. With these models grey values are predicted. Several coordinate systems will be introduced to make the modeling transparent and modular.

Coordinate systems

The design of the various models is simplified by the use of several coordinate systems. The cylinder is described in a suitably chosen system of body coordinates (X_1,X_2,X_3). The camera can be described conveniently in the camera coordinates (U_1,U_2,U_3). The image plane is chosen to be the ground plane of the coordinate system, $U_3 = 0$. The position and orientation of the body with respect to the camera appears in the transformation between body and camera coordinates, consisting of successive rotations around the X_1, X_2 and X_3–axes, a translation along a certain vector and a transformation according to Ballard and Brown (1) defined by:

$$U_1 = \frac{fX_1}{f-X_3} \qquad U_2 = \frac{fX_2}{f-X_3} \qquad U_3 = \frac{fX_3}{f-X_3} \tag{6}$$

where f is the perpendicular distance from the radiation source to the image plane. In (6) a projection can be recognized. A 2D image coordinate system (V_1,V_2) is related to the camera coordinates by:

$$V_1 = U_1 \qquad V_2 = U_2 \tag{7}$$

which is simply an orthographic projection. The loss of information lies in the fact, that from the image only U_1 en U_2 can be determined and not U_3.

Homogeneous coordinates. To achieve compact notations, homogeneous body coordinates (x_1,x_2,x_3,x_4) with $x_4=1$, and camera coordinates (u_1,u_2,u_3,u_4) are introduced, analogous to Duda and Hart (4). In common with robotics, the transformations between the coordinate systems are conveniently described by

$$\mathbf{u} = \mathbf{M}\mathbf{x} \tag{8}$$

The matrix M can be written as a product of matrices describing rotations, a translation and a projection respectively as in:

$$M = \begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&-f^{-1}&1\end{bmatrix}\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&l_1\\0&0&0&1\end{bmatrix}\begin{bmatrix}\cos\gamma&-\sin\gamma&0&0\\\sin\gamma&\cos\gamma&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix} *$$

$$*\begin{bmatrix}1&0&0&0\\0&\cos\psi&-\sin\psi&0\\0&\sin\psi&\cos\psi&0\\0&0&0&1\end{bmatrix}\begin{bmatrix}\cos\varphi&0&\sin\varphi&0\\0&1&0&0\\-\sin\varphi&0&\cos\varphi&0\\0&0&0&1\end{bmatrix}\begin{bmatrix}1&0&0&l_2\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}$$

$$= PL_1\Gamma\Psi\Phi L_2 \tag{9}$$

with:

P	the projection matrix,
L_1, L_2	matrices describing translations over distances l_1 and l_2 respectively,
Ψ, ϕ, Γ	matrices describing rotations ψ, φ and γ around the X_1, X_2, and X_3 axes respectively.

Description of the cylinder

Shape of the cylinder. The body coordinates are chosen such that the symmetry axis of the cylinder coincides with the x_2–axis. This simplifies the description of the cylinder considerably. The cylinder has a diameter d, therefore the set of points in space being part of the cylinder obeys the relation

$$x_1^2 + x_3^2 \le (1/2d)^2 \tag{10}$$

and the range of values for x_2 is only constrained by the chosen region of interest (ROI) in the image.

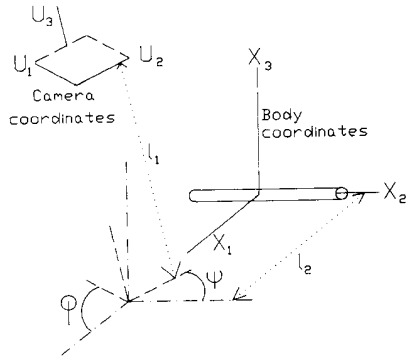

Figure 3 The cylinder in a body coordinate system

Position and orientation of the cylinder. In order to obtain a description of the body from the camera point of view, (10) must be transformed to camera coordinates according to transformation (9). The significance of the parameters contained in the translation matrices L_1, L_2 and the rotation matrices Γ, Ψ and ϕ, is illustrated in fig. 3.
The direction of the translation l_1 is along the vector connecting the centre of the image window and the radiation source (fig. 3). In many cases this centre will be the perpendicular projection of the radiation source in the image plane in which case the translation direction is the U_3 axis. Therefore the parameter l_1 can be considered to be a depth parameter.

Interaction of the cylinder with X–rays. To be able to predict image grey values, the interaction between the X–ray radiation and the cylinder must be known. The absorption and scattering of X–rays diminishes the intensity of the original X–ray according to the well–known Lambert–Beer law:

$$E(\Lambda) = E_0\exp(-\Lambda\mu) \tag{11}$$

Here Λ is the distance traversed by the X–ray through the cylinder, E_0 is the intensity of the entering X–ray radiation and μ is the attenuation coefficient of the

medium. $E(\Lambda)$ will determine the image grey values. Its dependence on the cylinder parameters is found by computing Λ as a function of its parameters. Suppose an X–ray crosses the X_1X_2–plane at a location $\mathbf{x}=(x_1,x_2,0,1)^T$, with direction $\mathbf{y}=(y_1,y_2,y_3,0)^T$ expressed in body coordinates. The total trajectory Λ of the ray through the cylinder is found from (10):

$$\Lambda = \frac{\sqrt{d^2(y_1^2+y_3^2)-4y_3^2x_1^2}}{y_1^2+y_3^2}\|\mathbf{y}\|$$
if the argument of the root is > 0

$$= 0 \quad \text{elsewhere} \tag{12}$$

with $\|\mathbf{y}\|$ the length of the vector $\mathbf{y}$.
The intensity $E(\Lambda)$ is known now as a function of E_0, d, the position $\mathbf{x}$ where it crosses the X_1X_2–plane and the vector $\mathbf{y}$. This vector will be computed later on (23).

Motion of the cylinder. We confine ourselves to rigid body motion which means that all the body points perform the same motion, described as a combination of a translation and a rotation. We assume the shape description to apply at time t_1. If the point $\mathbf{x}'$ at time $t_2=t_1+\Delta t$ corresponds to a point $\mathbf{x}$ at time t_1 we have (small rotation approximation):

$$\mathbf{x}' = D\mathbf{x} \tag{13}$$

$$D = \begin{bmatrix} 1 & -R_3 & R_2 & \Delta X_1 \\ R_3 & 1 & -R_1 & \Delta X_2 \\ -R_2 & R_1 & 1 & \Delta X_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{14}$$

with $\mathbf{R} = (R_1,R_2,R_3)^T$ the rotation vector, $\Delta\mathbf{X} = (\Delta X_1,\Delta X_2,\Delta X_3)^T$ the translation vector.

A different view on motion. The simple description of the body can be preserved if the motion description (13) is regarded as a coordinate transformation, yielding new body coordinates. The motion will be reflected by the fact that the transformation between camera coordinates and body coordinates after motion is not given by (8) but by:

$$\mathbf{u} = MD\mathbf{x}' \tag{15}$$

where $\mathbf{x}' = (x_1',x_2',x_3',x_4')^T$ are now the new body coordinates. In these new coordinates the cylinder is described after the motion by (10). Apart from the projection, both matrices M and D are composed in an analogous way. Therefore the rotation and translation parameters play the same role with respect to the second image as the orientation and position parameters with respect to the first image. The imaging process does not change due to this different view; however, using a second coordinate transformation to describe motion makes the incorporation of motion in the imaging model extremely simple, as we will see below.

Prediction of image grey values

We compute the grey value $I(\mathbf{v};\boldsymbol{\alpha})$ as a function of the parameter vector $\boldsymbol{\alpha}$ and the image position $\mathbf{v}$ as in (17). First we examine which position $\mathbf{x}$ at the cylinder plane is projected onto the image position $\mathbf{v}$. Subsequently the grey value in $\mathbf{v}$ is related to the intensity of the radiation emanating from the cylinder at position $\mathbf{x}$.

First stage, the imaging geometry. To find the position of a body point $\mathbf{x}$ as a function of the corresponding image coordinate $\mathbf{v}$, the relation (9) is inverted. Therefore

$$\mathbf{x} = M^{-1}\mathbf{u} \tag{16}$$

where M^{-1} is the inverse of (9). The transformation from camera to image coordinates is not invertible without the use of a priori knowledge. From Korsten (7) we have

$$\mathbf{x} = M^{-1}(u_4\mathbf{v} + f\mathbf{e}_3) \tag{17}$$

with
$$\mathbf{v} = (V_1,V_2,-f,1)^T$$
and $\mathbf{e}_3$= a homogeneous unity vector

The expression (17) still contains the unknown u_4, reflecting the fact that $\mathbf{u}$ is not determined unambiguously as a function of the image coordinates. The constraint $x_3=0$ is employed to compute u_4, leading to:

$$x_3 = \mathbf{e}_3^T\mathbf{x} = \mathbf{e}_3^TM^{-1}(u_4\mathbf{v} + f\mathbf{e}_3) = 0 \tag{18}$$

Therefore:
$$u_4 = -f\frac{\mathbf{e}_3^TM^{-1}\mathbf{e}_3}{\mathbf{e}_3^TM^{-1}\mathbf{v}} \tag{19}$$

Putting (19) in (17) yields the body position $\mathbf{x}$ as a function of the image position $\mathbf{v}$:

$$\mathbf{x} = fM^{-1}\left\{\mathbf{e}_3 - \left(\frac{\mathbf{e}_3^TM^{-1}\mathbf{e}_3}{\mathbf{e}_3^TM^{-1}\mathbf{v}}\right)\mathbf{v}\right\} \tag{20}$$

where $\mathbf{e}_3^TM^{-1}\mathbf{e}_3$ is the (3,3) element of the matrix M^{-1}.

Second stage, the image grey values. From the interaction between the cylinder and the X–rays at location $\mathbf{x}$, the image grey value $I(\mathbf{v};\boldsymbol{\alpha})$ has to be predicted. The conversion to image grey values is rather complicated; however, the total conversion process is supposed to be linear, therefore we still have an exponential Lambert–Beer relation for the grey values

$$I(\mathbf{v};\boldsymbol{\alpha}) = I_0\exp\{-\mu\Lambda(\mathbf{x})\} \tag{21}$$

where:
- $\mathbf{x}$ is according to (20)
- I_0 is the grey value if no attenuation of the radiation inside the cylinder has taken place.

It remains to find an expression for the vector $\mathbf{y}$ appearing in (12) indicating the direction of the entering radiation at the cylinder. The vector

$$\mathbf{y} = (y_1,y_2,y_3,0)^T \tag{22}$$

is determined by the position of the X–ray source with respect to the position of the body point $\mathbf{x}$. In camera coordinates the radiation source has the position $f\mathbf{e}_3$, and transformed to body coordinates $fM^{-1}\mathbf{e}_3$. In body coordinates $\mathbf{y}$ is

$$\mathbf{y} = fM^{-1}\mathbf{e}_3 - \mathbf{x}$$

$$= f\frac{\mathbf{e}_3^TM^{-1}\mathbf{e}_3}{\mathbf{e}_3^TM^{-1}\mathbf{v}}M^{-1}\mathbf{v} \tag{23}$$

Sequences of images. To obtain explicit grey value predictions, let the two images be registered at time t_1 and $t_2=t_1+\Delta t$ respectively. We suppose the shape model to describe the shape of the body at time t_1, then the problem of obtaining a grey value prediction $I_1(\mathbf{v};\boldsymbol{\alpha})$ at t_1 has been solved above. A prediction of the grey values $I_2(\mathbf{v},\boldsymbol{\alpha})$ at time t_2, as has been pointed out previously, is obtained with the transformation (15) between camera and new body coordinates. So we can immediately write down the position $\mathbf{x}'$ and the direction vector $\mathbf{y}'$ after the motion analogous to (20) and (23):

$$\mathbf{x}' = f(\mathrm{MD})^{-1}\left[\mathbf{e}_3 - \left(\frac{\mathbf{e}_3^T(\mathrm{MD})^{-1}\mathbf{e}_3}{\mathbf{e}_3^T(\mathrm{MD})^{-1}\mathbf{v}}\right)\mathbf{v}\right] \quad (24)$$

$$\mathbf{y}' = f\frac{\mathbf{e}_3^T(\mathrm{MD})^{-1}\mathbf{e}_3}{\mathbf{e}_3^T(\mathrm{MD})^{-1}\mathbf{v}}(\mathrm{MD})^{-1}\mathbf{v} \quad (25)$$

Due to this treatment, parameters of the motion model will be added to the shape and the transformation parameters.
From the combined models of shape and motion, parameters can be estimated. The parameter vector of the total model consists of the following fourteen parameters: I_0, μ, d, l_1, l_2, ψ, φ, γ, ΔX_1, ΔX_2, ΔX_3, R_1, R_2, R_3.

ROBUSTNESS AND ACCURACY OF THE RESULTS

The description of the behaviour with respect to convergence of multi parameter systems is quite complicated, because of the dependence on some parameters. A parameter might be outside its range of convergence because of the values of other parameters.

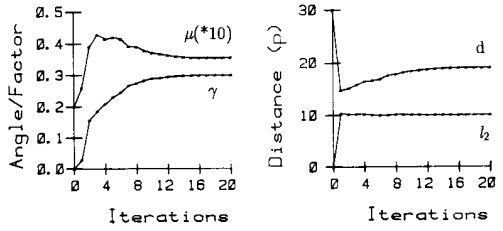

Figure 4a First stage of the algorithm

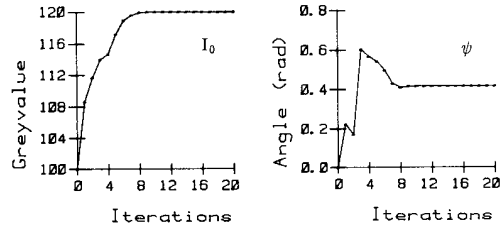

Figure 4b Second stage of the algorithm

This is a known problem considered for instance by Diehl (3). Based on considerations by Korsten (7) with respect to stability and identifiability, we decided to divide the set of parameters into two groups, leading to a two stage algorithm. In the first stage d, l_2, γ and μ are estimated while I_0=100 and ψ=0 are kept fixed at an initial estimate (fig. 4a). In the second stage ψ and I_0 are added to the set of parameters to be estimated, starting from the final result of the first stage (fig. 4b). The experiments were performed on model–generated images without noise. The parameters, used to produce these images were: φ=0, ψ=0.4 rad, l_1/f=0.06, f=8000p, d=20p, l_2=10p, γ=0.3 rad, I_0=120, μ=0.04, where p stands for pixel.

Estimation of motion parameters

Motion influences the second image of the image pair in a similar way as the position and orientation influence the first image. See "a different view on motion". It may be expected, therefore, that motion estimation from the second image has properties resembling very closely the properties of shape estimation from a single image. It may not be expected that simultaneous estimation of shape and motion parameters, using the model of fig. 2, will improve the stability of the algorithm with respect to individual parameters. Shape and orientation parameters can only be estimated from the first image, because of unknown motion. Motion parameters can only be estimated from the second image.
The translation ΔX_2 along the cylinder axis X_2 and the rotation R_2 around this axis are not identifiable (7), therefore they are removed. To prevent ambiguities in the motion parameters R_1,R_3, ΔX_1 and ΔX_3, the translation ΔX_3=30 p is taken fixed.

TABLE 1 Statistics of simultaneous shape, orientation and motion estimation.

	S/N ratio: 11		S/N ratio: 3.7	
Par	Mean	Sigma	Mean	Sigma
ΔX_1	-19.99	0.02	-19.95	0.12
R_1	-0.10	0.02	-0.09	0.10
R_3	0.201	0.006	0.20	0.03
d	20.01	0.02	20.58	0.18
l_2	9.997	0.015	9.98	0.08
γ	0.3001	0.0007	0.301	0.003
μ	0.03993	0.00015	0.0380	0.0008

From the set of shape and orientation parameters yielding stable estimation results d, l_2, γ, μ are selected corresponding to Korsten (7). The total set of parameters is ($d, l_2, \gamma, \mu, R_1, R_3, \Delta X_1$). Table 1 presents some estimation results from a pair of images, corrupted with noise. For each signal to noise ratio 25 "two stage" experiments were performed. During the first stage R_1=0 rad. and R_3=0 rad. were kept fixed. The initial estimate for d=30 p and for μ=0.02, while all the others are zero. The mean and the standard deviations of each parameter are shown in table 1.

Duality. An experiment, with the grey value I_0 incorporated into the estimation, produced dual solutions along with correct solutions. Again 25 two stage

TABLE 2 Statistics of simultaneous shape, orientation and motion estimation.

	S/N ratio: 11 18 measurements correct solution		S/N ratio: 11 7 measurements dual solution	
Par	Mean	Sigma	Mean	Sigma
ΔX_1	-20.00	0.02	-20.03	0.05
R_1	-0.10	0.02	-0.66	0.02
R_3	0.200	0.006	0.356	0.006
d	20.03	0.04	20.02	0.06
l_2	10.00	0.01	10.00	0.02
γ	0.2997	0.0006	0.3004	0.0007
I_0	120.7	0.9	122.4	1.7
μ	0.0402	0.0004	0.0407	0.0006

experiments were performed, where in the first stage I_0=100, R_1=0 rad. and R_3=0 rad. were kept fixed. The mean and standard deviations of each parameter are shown in table 2.
The dual solution is from Korsten (7) R_1=–0.574, R_3=0.296, ΔX_1=–29.2 and ΔX_3=–15.6. Because ΔX_3 was fixed to 30 p during the experiment, the perfect dual solution could not be reached. The influence of ΔX_3 on

the image is only slight, therefore even with ΔX_3=30 p there is a non perfect dual solution close to the perfect dual solution. As can be seen by comparing tables 1 and 2 the duality does not affect the accuracy of the correct solutions.

SUMMARY AND CONCLUSIONS

Shape, shading and motion of a solid cylinder were systematically modeled. An algorithm was derived, which combines these models to predict image grey values. The predicted grey values, together with the corresponding grey values from a real world image, are used in an estimator based on successive linearization around the actual estimate of the parameter vector, to improve the estimate. Experiments, using artificial images instead of real world images, were carried out. The obtained results show that simultaneous shape and motion estimation is conditionally possible, even if the "real world images" are corrupted by noise. When dual solutions arise, the accuracy of the correct solutions is not affected, because of a slight influence of ΔX_3 on the image.

REFERENCES

1. Ballard, D.H., and Brown, C.M., 1982, "Computer Vision", Prentice–Hall Inc., Englewood Cliffs, New Jersey, USA.

2. Burkhardt, H., and Moll, H., 1979, "A modified Newton–Raphson search for the model–adaptive identification of delays", Proceedings of the IFAC–Conference on "Identification and system parameter identification", Pergamon, New York, USA

3. Diehl, N., 1988, "Methoden zur Allgemeinen Bewegungsschatzung in Bildfolgen", VDI Verlag, Dusseldorf, FRG.

4. Duda, R.O., and Hart, P.E., 1973, "Pattern Recognition and Scene Analysis", John Wiley & Sons, New York, USA.

5. Houkes, Z., 1985, "Distance measurement by stereo vision", Proceedings of the Int. Conf. on "Robot Vision and Sensory Controls", Amsterdam, NL.

6. Houkes, Z., 1983, "Motion parameter estimation in TV–pictures", NATO ASI series Vol. F2, "Image processing and dynamic scene analysis", Ed. by Th.S. Huang, Berlin, FRG.

7. Korsten, M.J., 1989, "Three–dimensional body parameter estimation from digital images", University of Twente, Enschede, The Netherlands.

8. Liebelt, P.B., 1967, "An introduction to optimal estimation", Addison Wesley, Reading, MA, USA.

9. Stuller, S., and Krishnamurti, G., 1983, Comp. Vis. Graph. and Im. Proc., 21, 169–204.