SIGNAL PROCESSING III: Theories and Applications
I.T. Young et al. (editors)
Elsevier Science Publishers B.V. (North-Holland)

SIMULTANEOUS ESTIMATION OF ROTATION AND TRANSLATION IN IMAGE SEQUENCES*

Hans Burkhardt, Norbert Diehl

Arbeitsbereich Technische Informatik I
Technische Universität Hamburg-Harburg, Postfach 90 14 03
D-2100 Hamburg 90, West-Germany

This paper presents a new, iterative algorithm for the simultaneous estimation of rotation and translation parameters of moving planar objects in grey-scale image sequences. The algorithm combines several advantages such as large stability region, high image-bandwidth-adaptive convergence rate of at least second order near the optimum and a minimum of numeric expense within each iteration step. Furthermore an extension to estimate affine transform parameters and tests of the algorithm using real image data are presented.

1. INTRODUCTION

A well known problem in image processing and scene analysis is the estimation of motion parameters like rotation and translation in image sequences. Applications are found in several areas such as motion compensated image coding, remote sensing by satellites, robotics and biology [1,2,3]. A lot of different algorithms to estimate pure translatoric displacements have been published e.g. [3,4], but only a few papers [5,6,7,8] deal with rotation and translation or give an efficient parameter estimation algorithm based on ordinary grey-scale images without using special features as corresponding points.

In this paper a new fast converging algorithm to estimate rotation and translation simultaneously in grey-scale images is discussed. The estimation problem is formulated as a model adjustment identification problem which is solved by a special minimization algorithm. Also an extension to affine transforms is given.

2. THE ALGORITHM

Two grey-scale images $I_1(x,y)$ and $I_2(x,y)$ are assumed to represent a moving rigid planar object $S(x,y)$ in front of a uniform background where no occlusion effects occur:

$$I_1(\mathbf{x}) = S(\mathbf{x}) = S(x,y)$$
$$I_2(\mathbf{x}) = S(x\cos\phi - y\sin\phi - d_1, x\sin\phi + y\cos\phi - d_2).$$

$I_2(\mathbf{x})$ results from $S(\mathbf{x})$ and therefore from $I_1(\mathbf{x})$ by rotation ϕ and translation d_1, d_2.

To get an appropriate estimate $\hat{\mathbf{T}} = (\hat{\phi}, \hat{d}_1, \hat{d}_2)^T$ of the true motion parameters $\mathbf{T} = (\phi, d_1, d_2)^T$ a *model adaptive identification structure* is used. The motion is modelled by

$$I_m(\mathbf{x}, \hat{\mathbf{T}}) = S(x\cos\hat{\phi} - y\sin\hat{\phi} - \hat{d}_1, x\sin\hat{\phi} + y\cos\hat{\phi} - \hat{d}_2).$$

As model error criterion $J\{e(\hat{\mathbf{T}})\}$ we use the expectation E of the squared model error

$$J\{e(\hat{\mathbf{T}})\} = E\{e^2\} = E\{(I_m(\mathbf{x},\hat{\mathbf{T}}) - I_2(\mathbf{x}))^2\}.$$

For stationary stochastic signals $J\{e(\hat{\mathbf{T}})\}$ equals twice the negative cross-correlation function of $I_m(\mathbf{x},\hat{\mathbf{T}})$ and $I_2(\mathbf{x})$ plus an additive constant. $J\{e(\hat{\mathbf{T}})\}$ is assumed to be at least twice differentiable with respect to $\hat{\mathbf{T}}$.

A global search for an optimal parameter vector $\hat{\mathbf{T}} = \hat{\mathbf{T}}^*$ is unrealistic and results in the calculation of the three-dimensional cross-correlation function $R(\phi, d_1, d_2)$ in conjunction with a tremendeous numerical complexity.

We use instead an *iterative minimizing strategy* changing the model parameter vector $\hat{\mathbf{T}}$ well directed until $J\{e(\hat{\mathbf{T}})\}$ reaches its minimum. The proposed algorithm is an extension of a one-dimensional modified Newton-Raphson algorithm [9] which was developed for time-delay-estimation to estimate parameter vectors based on two-dimensional signals. The main structure of the algorithm is given by the iteration:

$$\hat{\mathbf{T}}^{K+1} = \hat{\mathbf{T}}^K - \mathbf{H}^{-1}(\hat{\mathbf{T}} = \hat{\mathbf{T}}^* = \mathbf{T}) \cdot \mathbf{g}(\hat{\mathbf{T}}^K).$$

The new estimate $\hat{\mathbf{T}}^{K+1}$ of iteration step $K+1$ is given by the estimate $\hat{\mathbf{T}}^K$ of iteration step K and an innovation given by the multiplication of the inverse of the Hessian $\mathbf{H}(\hat{\mathbf{T}} = \hat{\mathbf{T}}^* = \mathbf{T})$ with the gradient vector $\mathbf{g}(\hat{\mathbf{T}}^K)$. Note that the *Hessian* as the second derivatives of the error criterion is always taken *at the optimum* $\hat{\mathbf{T}} = \hat{\mathbf{T}}^* = \mathbf{T}$ and not at the actual iteration point $\hat{\mathbf{T}}^{K+1}$ as in the original Newton-Raphson-algorithm. Using the Hessian at the optimum has several advantages which will be discussed in the sequel.

First we will show how to calculate the Hessian at the optimum before starting the iteration without actually knowing the optimum, using the special motion structure of our model. Therefore we distinguish between pure translation and translation in combination with rotation.

* This work was supported by the Deutsche Forschungsgemeinschaft (DFG)

Expressing the Hessian at the optimum by the derivatives of the images we get

$$\frac{\partial^2 J\{e(\hat{\mathbf{T}}^*)\}}{\partial \hat{T}_i \partial \hat{T}_j} = 2\, E\left\{ \frac{\partial I_m(\mathbf{x}, \hat{\mathbf{T}}^*)}{\partial \hat{T}_i} \frac{\partial I_m(\mathbf{x}, \hat{\mathbf{T}}^*)}{\partial \hat{T}_j} \right\},$$

noting that at the optimum $\hat{\mathbf{T}} = \hat{\mathbf{T}}^* = \mathbf{T}$ the image difference $e(\hat{\mathbf{T}})$ vanishes.

For *pure translation* without rotation the last equation can be simplified furthermore if we express the derivatives with respect to the translation parameters by derivatives with respect to the coordinates noting the fact that at the optimum $I_m(\mathbf{x}, \hat{\mathbf{T}}^*) = I_2(\mathbf{x})$:

$$\frac{\partial^2 J\{e(\hat{\mathbf{T}}^*)\}}{\partial \hat{d}_i \partial \hat{d}_j} = 2\, E\left\{ \frac{\partial I_2(\mathbf{x})}{\partial x_i} \frac{\partial I_2(\mathbf{x})}{\partial x_j} \right\},$$

using the abbreviation $\mathbf{x} = (x = x_1, y = x_2)^T$.

Thus the Hessian does not explicitely depend on the translation parameters d_1, d_2 and can be calculated before starting the iteration. A concrete interpretation of this fact is that at the optimum the cross-correlation function of $I_m(\mathbf{x}, \hat{\mathbf{T}})$ and $I_2(\mathbf{x})$ becomes identical with the autocorrelation function of $I_2(\mathbf{x})$. Therefore we can a priori calculate the curvature of the autocorrelation function of $I_2(\mathbf{x})$ instead of the cross-correlation function at the unknown optimum.

However, if we have *rotation and translation simultaneously* the calculation of the Hessian is not as straightforward as in the case with pure translation. Because of the fact that rotation and translation are not independent and therefore do not commute the true rotation angle has to be known if we now want to express the derivatives with respect to the parameter vector by derivatives with respect to the coordinates. For instance now we get at the optimum with $\hat{\phi} = \hat{\phi}^* = \phi$

$$\frac{\partial I_m(\mathbf{x}, \hat{\mathbf{T}}^*)}{\partial \hat{d}_1} = -\frac{\partial I_2(\mathbf{x})}{\partial x} \cos\phi + \frac{\partial I_2(\mathbf{x})}{\partial y} \sin\phi.$$

Thus if there is rotation and translation the partial derivatives at the optimum and therefore the Hessian explicitely depend on the unknown parameter ϕ. For the motion model under consideration the Hessian can therefore not be calculated before starting the iteration.

Nevertheless as will be shown in the following the Hessian of a *slightly modified structure* can be calculated beforehand and the advantages of the algorithm can be preserved. We introduce a running coordinate system $\{\mathbf{x}^K\} = \{x^K, y^K\}$ and $\tilde{\mathbf{T}}^{K+1} = (\tilde{\phi}^{K+1}, \tilde{d}_1^{K+1}, \tilde{d}_2^{K+1})^T$ as an additional motion vector describing the estimate at the $K+1$-th iteration on the basis of the coordinate system $\{\mathbf{x}^K\}$ of the K-th iteration.

Thus the relation between the model image $I_m(\mathbf{x}, \hat{\mathbf{T}}^K)$ of the K-th iteration step and $I_m(\mathbf{x}, \hat{\mathbf{T}}^{K+1})$ of the $K+1$-th iteration step is described as

$$\begin{aligned} I_m(\mathbf{x}, \hat{\mathbf{T}}^{K+1}) &= I_m(\mathbf{x}, \hat{\mathbf{T}}^K, \tilde{\mathbf{T}}^{K+1}) = S(\mathbf{x}^{K+1}) \\ &= S(x^K \cos\tilde{\phi}^{K+1} - y^K \sin\tilde{\phi}^{K+1} - \tilde{d}_1^{K+1}, \ldots) \\ &= S(x \cos\hat{\phi}^{K+1} - y \sin\hat{\phi}^{K+1} - \hat{d}_1^{K+1}, \ldots). \end{aligned}$$

Thus the innovation $\tilde{\mathbf{T}}^{K+1}$ is given in the transformed coordinates $\{\mathbf{x}^K\}$ and only indirectly in the coordinates $\{\mathbf{x}\}$; i.e. the new model image $I_m(\mathbf{x}, \hat{\mathbf{T}}^{K+1})$ is related to the model image $I_m(\mathbf{x}, \hat{\mathbf{T}}^K)$ by the motion vector $\tilde{\mathbf{T}}^{K+1}$. Therefore instead of differentiating the error criterion with respect to $\hat{\mathbf{T}}$ now the error criterion has to be differentiated with respect to $\tilde{\mathbf{T}}$. The slightly modified algorithm is given by

$$\hat{\mathbf{T}}^{K+1} = \mathbf{A}^{K+1}\hat{\mathbf{T}}^K + \tilde{\mathbf{T}}^{K+1} = \mathbf{A}^{K+1}\hat{\mathbf{T}}^K - \tilde{\mathbf{H}}^{-1} \cdot \tilde{\mathbf{g}}(\hat{\mathbf{T}}^K)$$

with the new Hessian $\tilde{\mathbf{H}}$, the new gradient vector $\tilde{\mathbf{g}}$ and the weighting matrix

$$\mathbf{A}^K = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\tilde{\phi}^K & -\sin\tilde{\phi}^K \\ 0 & \sin\tilde{\phi}^K & \cos\tilde{\phi}^K \end{pmatrix}$$

which describes the connection between $\tilde{\mathbf{T}}^{K+1}$, $\hat{\mathbf{T}}^{K+1}$ and $\hat{\mathbf{T}}^K$.

Using this indirect generation of the model image the Hessian $\mathbf{H}$ at the optimum can be calculated off-line without knowing the true motion vector. Assuming that the optimum is reached at iteration step N; i.e. $I_m(\mathbf{x}, \hat{\mathbf{T}}^N) = I_2(\mathbf{x})$. The estimate $\tilde{\mathbf{T}}^{N+1}$ of iteration step $N+1$ has to be $\tilde{\mathbf{T}}^{N+1} = 0$. This leads to the following derivatives:

$$\left.\frac{\partial I_m(\mathbf{x}, \hat{\mathbf{T}}^N, \tilde{\mathbf{T}}^{N+1})}{\partial \tilde{d}_1^{N+1}}\right|_{\tilde{\mathbf{T}}^{N+1}=0} = -\frac{\partial S(\mathbf{x}^N)}{\partial x^N}$$

$$\left.\frac{\partial I_m(\mathbf{x}, \hat{\mathbf{T}}^N, \tilde{\mathbf{T}}^{N+1})}{\partial \tilde{\phi}^{N+1}}\right|_{\tilde{\mathbf{T}}^{N+1}=0} = -\frac{\partial S(\mathbf{x}^N)}{\partial x^N} y^N + \frac{\partial S(\mathbf{x}^N)}{\partial y^N} x^N$$

If we use these expressions to calculate the Hessian we get for instance

$$\frac{\partial^2 J\{e(\hat{\mathbf{T}}^*)\}}{\partial \tilde{d}_1^{N+1} \partial \tilde{\phi}^{N+1}} = E\left\{ \frac{\partial S(\mathbf{x})}{\partial x} \left(\frac{\partial S(\mathbf{x})}{\partial x} y - \frac{\partial S(\mathbf{x})}{\partial y} x \right) \right\}.$$

Introducing the parameter vector $\tilde{\mathbf{T}}$ and thus relating the new model image to the old one, the Hessian at the optimum can be expressed only by the derivatives of $S(\mathbf{x}) = I_1(\mathbf{x})$ with respect to the original coordinates $\{\mathbf{x}\}$ and is independent of the true motion vector. A concrete interpretation of this fact is that calculating the derivatives at $\tilde{\mathbf{T}} = 0$ can be interpreted as infinitesimal rotation and translation which in this case do commute. The preceeding formula uses the fact that the expectation value is independent of an arbitrary coordinate system. Therefore all derivatives are given in terms of the original image $S(\mathbf{x})$. This is consistent if we have stochastic stationary signals or if we use spatial averaging instead of the expectation operation for isolated objects in front of a uniform background.

Finally with the abbreviation $\partial/\partial\phi = y \cdot \partial/\partial x - x \cdot \partial/\partial y$ and the operator $\partial = (\partial_1, \partial_2, \partial_3)^T = (\partial/\partial\phi, \partial/\partial x, \partial/\partial y)^T$ the elements of the Hessian $\tilde{\mathbf{H}}$ and the vector $\tilde{\mathbf{g}}$ can be written as:

$$\tilde{H}_{ij} = 2\, E\{\partial_i I_1(\mathbf{x}) \partial_j I_1(\mathbf{x})\}$$

$$\tilde{g}_i(\hat{\mathbf{T}}^K) = -2 \sum_{j=1}^{3} B_{ij}^K\, E\{(I_m(\mathbf{x}, \hat{\mathbf{T}}^K) - I_2(\mathbf{x}))\partial_j I_2(\mathbf{x})\}.$$

Because of the fact that under the preceeding assumptions the expectation value is independent of the used coordinate system it is possible to express the derivatives in the gradient vector as derivatives of image $I_2(\mathbf{x})$ and not as derivatives of the model image $I_m(\mathbf{x}, \hat{\mathbf{T}})$. Thus we have to differentiate $I_2(\mathbf{x})$ only once instead of differentiating $I_m(\mathbf{x}, \hat{\mathbf{T}})$ within each iteration step.

The matrix $\mathbf{B}$ describes the connection between the derivatives with respect to $\tilde{\mathbf{T}}$ and those with respect to the original coordinates $\{\mathbf{x}\}$:

$$\mathbf{B}^K = \begin{pmatrix} 1 & \hat{d}_1^K \sin\hat{\phi}^K - \hat{d}_2^K \cos\hat{\phi}^K & \hat{d}_1^K \cos\hat{\phi}^K + \hat{d}_2^K \sin\hat{\phi}^K \\ 0 & \cos\hat{\phi}^K & -\sin\hat{\phi}^K \\ 0 & \sin\hat{\phi}^K & \cos\hat{\phi}^K \end{pmatrix}.$$

3. PROPERTIES OF THE ALGORITHM

The algorithm has *several advantages* such as large stability region, high image-bandwidth-adaptive convergence rate and a minimum of numeric expense within each iteration step.

The algorithm has in general a much *larger stability range* compared to the normal Newton-Raphson-algorithm. One-dimensionally speaking it is like the gradient algorithm stable up to the next optimum of the error criterion $J\{e(\hat{\mathbf{T}})\}$. Like the normal Newton-Raphson-technique the algorithm has good *signal adaptive properties*. The Hessian which can be interpreted as the second derivatives of the autocorrelation function of the images adjusts to the image bandwidth. Therefore the *convergence rate* of the Euclidean error norm $\varepsilon = \| \hat{\mathbf{T}}^K - \mathbf{T} \|$ is at least of *second order*, independently of the chosen signals.

If we have pure translation or pure rotation the convergence rate is even of third order. The same is true if all off-diagonal elements of the Hessian are zero

$$\tilde{H}_{ij} = 0 \qquad \text{for} \qquad i \neq j$$

and additionally

$$E\{\frac{\partial S}{\partial x}\frac{\partial S}{\partial x}\} = E\{\frac{\partial S}{\partial y}\frac{\partial S}{\partial y}\}$$

which means that the error criterion $J\{e(d_1, d_2)\}$ for pure translation is rotationally invariant with respect to d_1, d_2. Because of these properties even large parameter vectors can be identified in a few iterative steps.

Another advantage of the algorithm is the *low numeric complexity* within each iteration step because of the ability to calculate the Hessian once before starting the iteration. This is attractive for near real-time implementations [10].

4. EXTENSION TO AFFINE TRANSFORMS

The given algorithm can be extended to estimate affine parameters $\mathbf{T} = (c_{11}, c_{12}, c_{21}, c_{22}, d_1, d_2)^T$ of the coordinate transform

$$\begin{pmatrix} x\prime \\ y\prime \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}.$$

As before the two model parameter vectors $\hat{\mathbf{T}}$ and $\tilde{\mathbf{T}}$ are introduced with

$$\begin{aligned} \mathbf{x}^{K+1} &= \hat{\mathbf{C}}^{K+1} \cdot \mathbf{x} - \hat{\mathbf{d}}^{K+1} \\ &= \tilde{\mathbf{C}}^{K+1} \cdot (\hat{\mathbf{C}}^K \cdot \mathbf{x} - \hat{\mathbf{d}}^K) - \tilde{\mathbf{d}}^{K+1}. \end{aligned}$$

With the operator $\partial = (x \cdot \partial/\partial x, y \cdot \partial/\partial x, x \cdot \partial/\partial y, y \cdot \partial/\partial y, -\partial/\partial x, -\partial/\partial y)^T$ the Hessian $\tilde{\mathbf{H}}$ at the optimum can be written as

$$\tilde{H}_{ij} = 2\ E\{\partial_i I_1(\mathbf{x}) \partial_j I_1(\mathbf{x})\}$$

and the gradient vector as

$$\tilde{g}_i(\hat{\mathbf{T}}^K) = -2 \sum_{j=1}^{6} B_{ij}^K\ E\left\{(I_m(\mathbf{x}, \hat{\mathbf{T}}^K) - I_2(\mathbf{x}))\partial_j I_2(\mathbf{x})\right\}$$

with a matrix $\mathbf{B}$ which again describes the connection between the derivatives with respect to $\tilde{\mathbf{T}}$ and those with respect to the coordinates $\{\mathbf{x}\}$. The iteration structure is changed only slightly. The innovation is given by

$$\tilde{\mathbf{T}}^{K+1} = -\tilde{\mathbf{H}}^{-1} \cdot \tilde{\mathbf{g}}(\hat{\mathbf{T}}^K)$$

and the new motion vector $\hat{\mathbf{T}}^{K+1}$ is related to $\hat{\mathbf{T}}^K$ and $\tilde{\mathbf{T}}^{K+1}$ by the transforms

$$\hat{\mathbf{C}}^{K+1} = \tilde{\mathbf{C}}^{K+1} \cdot \hat{\mathbf{C}}^K \quad , \quad \hat{\mathbf{d}}^{K+1} = \tilde{\mathbf{C}}^{K+1} \cdot \hat{\mathbf{d}}^K + \tilde{\mathbf{d}}^{K+1}.$$

With this algorithm for instance it is possible to estimate the parameter vector of an object which is inclined, rotated and translated.

5. TESTS WITH REAL IMAGE DATA

The algorithm has been tested with real image data. Therefore several scenes with well defined rotation and translation have been digitized and analysed. Image I shows one typical scene used in the experiments digitized by 512 x 512 pixels with marked regions of interest.

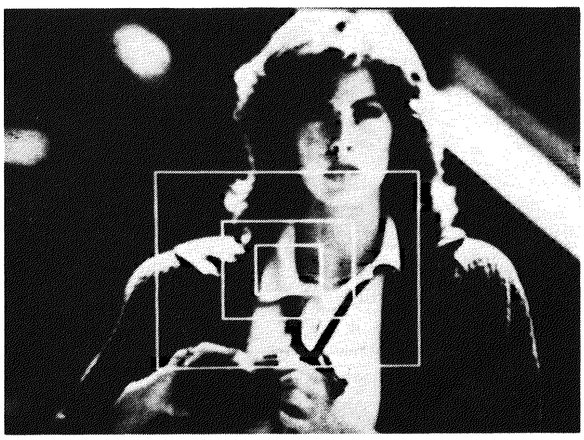

Image I

The first figure gives the joint identification of rotation and translation with the motion vector $\mathbf{T} = (20^\circ, 4, 2)^T$ given in degrees and pixels. The rotation was always around the centre of the marked areas of image I and the smallest of these three areas (51 x 51 pixels) was used as region of interest to calculate the expectation values. The estimated values $\hat{\phi}$ in degrees and $\hat{d}_1, \hat{d}_2$ in pixels are plotted versus the iteration number K.

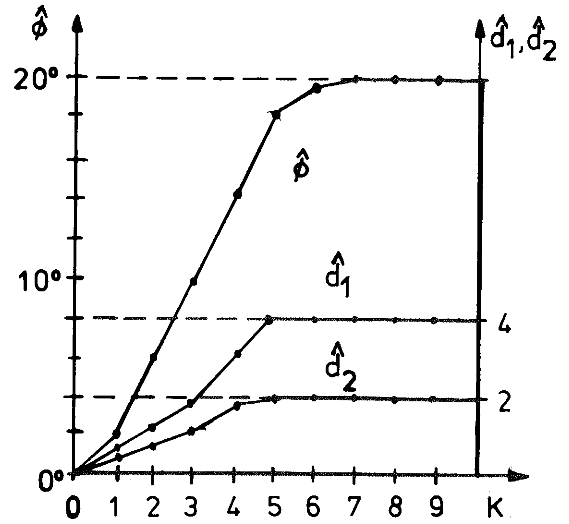

Figure 1

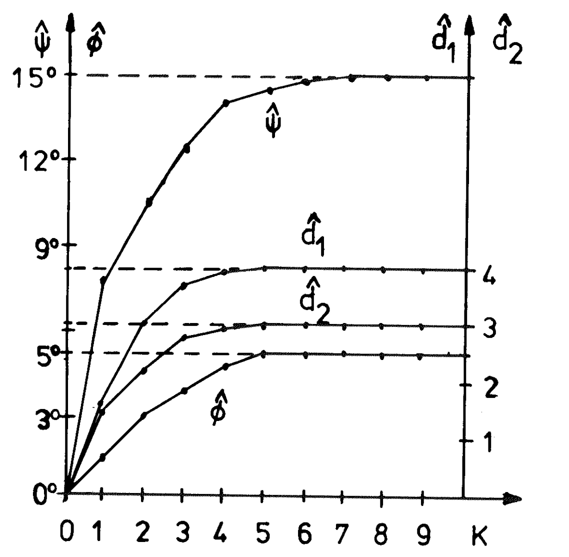

Figure 2

In the second example the parameters of an affine transform are estimated. Therefore the image was inclined by $\psi = 15^\circ$, rotated by $\phi = 5^\circ$, and translated by 4 pixels in x and 3 pixels in y direction. The inclination was around an axis through the centre of rotation. Again the estimated values $\hat{\psi}$, $\hat{\phi}$ and $\hat{d}_1, \hat{d}_2$ are given versus the iteration number K.

6. CONCLUSIONS

The paper describes a fast converging algorithm for the joint estimation of rotation and translation in image sequences. Furthermore an extension of the algorithm to estimate affine transform parameters and tests with real image data are presented.

REFERENCES

[1] Nagel, H.H.: Image sequence analysis: What can we learn from applications. Huang, T.S. (ed.): Image Analysis (Springer 1981)

[2] Nagel, H.H.: Analyse und Interpretation von Bildfolgen. Informatik-Spektrum 8 (1985) pp. 178-200 and pp. 312-327.

[3] Huang, T.S. (ed.): Proc. Nato Advanced Study Inst. on Image Sequence Processing and Dynamic Scene Analysis. Braunlage 1982 (Springer 1983)

[4] Netravalli, A.N.; Robbins, J.O.: Motion-compensated television coding: part I. Bell Syst. Techn. J. 58 (1979) pp. 631-670.

[5] Schalkoff, R.J.; McVey, E.S.: A model and tracking algorithm for a class of video targets. IEEE Trans. PAMI -4, (1982) pp. 2-10.

[6] Legters, G.R.; Young, T.Y.: Mathematical model for computer image tracking. IEEE Trans. PAMI-4, (1982) pp. 583-594.

[7] Axelsson, S.R.J.: On optimum algorithms for imaging tracking systems. Kunt, M.; de Coulon, F. (eds.): Signal Processing: Theories and Applications (North-Holland, 1980) pp. 723-728.

[8] Lenz, R.; Gerhard, A.: Adaptive geometrische Transformationen zur Mustererkennung mit Hilfe eines linearen, lokalen Distanzmaßes. Niemann, H.(ed.): Mustererkennung 1985, 7. DAGM-Symposium, Informatik-Fachberichte 107 (Springer, 1985) pp. 112-177.

[9] Burkhardt, H.; Moll, M.: A modified Newton-Raphson-search for the model-adaptive identification of delays. Isermann, R. (ed.): Identification and System Parameter Estimation (Pergamom, 1979) pp. 1279-1286.

[10] Diehl, N.; Burkhardt, H.: Planar motion estimation with a fast converging algorithm. Submitted for publication.